



**HAL**  
open science

# How does the crowd debate and consent in an open online context? A case study: Reddit – Change My View

Mattias Mano

► **To cite this version:**

Mattias Mano. How does the crowd debate and consent in an open online context? A case study: Reddit – Change My View. Business administration. Institut Polytechnique de Paris, 2020. English. NNT: 2020IPPAX082 . tel-03169853

**HAL Id: tel-03169853**

**<https://theses.hal.science/tel-03169853v1>**

Submitted on 15 Mar 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



INSTITUT  
POLYTECHNIQUE  
DE PARIS

NNT : 2020IPPAX082

Thèse de doctorat



# How does the crowd debate and consent in an open online context? A case study: Reddit – Change My View

Thèse de doctorat de l'Institut Polytechnique de Paris  
préparée à l'École polytechnique

École doctorale n°626 de l'Institut Polytechnique de Paris  
(ED IP Paris)  
Spécialité de doctorat : Sciences de gestion

Thèse présentée et soutenue à Palaiseau, le lundi 14 décembre 2020, par

**Mattias MANO**

Composition du Jury :

Thierry RAYNA Professeur, École polytechnique (i3 – CRG)	Président
Nicolas JULLIEN Professeur, Institut Mines Télécom Atlantique	Rapporteur
Paola TUBARO Directrice de recherches, CNRS (LRI)	Rapporteuse
Matthijs DEN BENSTEN Assistant Professeur, Montpellier Business School	Examineur
Jean-Michel DALLE Professeur, École polytechnique (i3 – CRG)	Directeur de thèse
Joanna TOMASIK Professeur, CentraleSupélec (LRI)	Co-Directrice de thèse

## Résumé français

Le développement d'Internet a révolutionné de multiples aspects du quotidien des individus, et en particulier comment chacun échange les autres. Depuis les années 1980, avec ses premiers *newsgroups*, aux réseaux sociaux d'aujourd'hui, les individus interagissent régulièrement en ligne. Les sites les plus importants, où ces échanges ont lieu, tel que Reddit, regroupent des millions de discussions et d'utilisateurs, mettant en avant leur importance dans nos sociétés. A titre d'exemple, en 2003, 20% des utilisateurs américains d'Internet ont déclaré avoir visité un groupe de discussion en ligne, 17% ont rédigé du contenu en ligne et 10% participé à un groupe de discussion. Ces sites *sociaux* couvrent différents types d'activités, de la recherche d'une solution à un problème technique (par exemple, les services après-vente), à des sites permettant de rester informé des derniers sujets d'information (blogs et fora). Ils sont également le lieu de débats politiques citoyens. En effet, ces sites *sociaux* sont devenus des espaces clés permettant aux citoyens de s'engager. En 2018, plus de deux-tiers des Américains déclarent que les réseaux sociaux aident à donner un espace visible à des *minorités*, et 14% déclarent que ces réseaux sociaux ont changé leur point de vue sur un fait de société.

Pour étudier de tels phénomènes, certains chercheurs s'accordent sur le fait qu'Internet permet à des groupes d'individus, que l'on peut nommer *foule*, de se rassembler *quelque part* afin d'échanger des informations. Les travaux scientifiques ont étudié les comportements de la foule bien avant le développement d'Internet, en particulier dans la sphère politique. Le concept de *tiers lieu* (Oldenburg, 1999) définit un espace public, en-dehors du domicile et de l'espace de travail, où les individus se retrouvent et échangent de façon informelle. Le concept de *tiers espace* (Wright, 2012) transpose le *tiers lieu* dans la sphère numérique. Ces blogs, ces fora en ligne remplacent-ils les lieux d'échanges informelles tels que les restaurants et les bars ? La recherche se doit de mieux comprendre les mécanismes et les procédures qui régissent de ces *tiers espaces* afin de comprendre les nouveaux comportements induits par les nouvelles technologies de l'ère numérique.

L'usage de ces *tiers espaces* s'est intensifié notamment en France, à la fin de l'année 2018. Le mouvement des « Gilets Jaunes » a rassemblé des milliers de personnes, descendant dans les rues pour manifester, chaque dimanche, contre certaines prises de décision du gouvernement. L'un des points soulevés par les manifestants est le manque de compréhension des décideurs politiques par rapport au quotidien du Français *moyen*. En réponse directe, le gouvernement ouvre en janvier 2019 le Grand Débat, qui permet à chaque citoyen de remplir des cahiers de doléances dans les mairies. De plus, un site internet dédié a été ouvert dans ce même but : garantir un espace libre et ouvert à tous les citoyens afin qu'ils puissent s'exprimer, débattre.

En dehors de la sphère politique, Internet a permis l'émergence d'autres *tiers espaces*. Wikipedia, la plus populaire des encyclopédies, intégralement accessible en ligne, est le résultat de productions individuelles. Lancé en 2001, on y trouve fin 2018 plus de 5.764.000 articles et 35.146.000 comptes utilisateurs. Nous pouvons également mentionner le développement de Linux, un système d'exploitation, développé en 1991 par Linus Torvalds, intégralement et gratuitement disponible en ligne. Depuis 2005, 15.600 personnes ont contribué à ce projet. Ces individus ont coopéré gratuitement, permettant la diffusion du savoir partout dans le monde. Ce type de processus a été conceptualisé par la recherche en tant que production d'une intelligence collective.

Le présent travail de recherche étudie l'émergence de l'intelligence collective sur Internet. Plus précisément, nous abordons deux questions de recherches. Premièrement, existe-il différents processus permettant d'avoir un débat constructif ? Deuxièmement, si un consensus est atteint parmi les participants d'un débat, quelles en sont les conditions ?

Pour répondre à ces questions, nous nous sommes appuyés sur un terrain : le forum en ligne Reddit – Change My View. Sur ce forum, un individu expose son avis sur un sujet, et demande à la foule de lui apporter les arguments lui permettant de changer d’avis. La base de données étudiée couvre les discussions depuis janvier 2013, création du forum, jusqu’en novembre 2016. Cela concerne plus de 21.000 discussions, 1.442.000 posts et 72.000 participants. Afin d’étudier les différents processus de débat, nous avons appliqué plusieurs algorithmes de classement, basés sur certaines caractéristiques des discussions, modélisées par des *réseaux* particuliers : les motifs (Milo et al., 2002). Nous observons (Mano, Dalle, and Tomasik, 2017) que les individus qui ouvrent la discussion adaptent leur stratégie en fonction de facteurs externes, tel que le nombre de challengers auxquels il fait face.

Concernant le processus de consensus, nous étudions la causalité de plusieurs variables sur le consensus lui-même. Nous définissons un nouveau concept : le Consentement de la foule (Mano, Dalle, et Tomasik, 2018). Le consensus entre un individu et la foule peut être atteint, si et seulement si la foule a eu l’occasion de participer au débat, et pas seulement de signaler son désaccord au travers des systèmes de récompenses du forum.

Les prochains travaux de recherche devraient s’intéresser à deux aspects. D’abord, l’étude plus approfondi du cycle de vie du forum à l’étude. En effet, tous comme les communautés en ligne, la foule en ligne est un « organisme vivant », avec plusieurs phases d’évolution. Une étude temporelle des réseaux permettrait d’évaluer plus précisément ces évolutions. Ensuite, une étude sémantique des échanges permettraient de mieux appréhender les mécanismes mis en avant dans ce présent travail.

# Contents

<b>1</b>	<b>Introduction</b>	<b>7</b>
<b>2</b>	<b>Wisdom of the Crowd</b>	<b>13</b>
2.1	Who: The crowds . . . . .	15
2.1.1	The crowd before the Internet . . . . .	15
2.1.2	Communities of Practices . . . . .	17
2.1.3	Online Community . . . . .	18
2.2	Why: The motivations . . . . .	20
2.3	How: The framework . . . . .	22
2.3.1	Swarm intelligence, Wisdom of the Crowd, or Collec- tive intelligence . . . . .	22
2.3.1.1	Wisdom of the Crowds . . . . .	23
2.3.1.2	Limits . . . . .	26
2.3.2	Wisdom of the Crowds and Democracy . . . . .	27
2.3.2.1	Epistemic Democracy . . . . .	27
2.3.2.2	Procedural Democracy . . . . .	30
2.3.2.3	Deliberative Democracy . . . . .	34
2.3.2.4	Reconciliation of the approaches . . . . .	38
2.4	What: Adding-values to society . . . . .	41
2.4.1	Knowledge sharing . . . . .	41
2.4.2	Open Innovation . . . . .	42
2.4.3	Crowdsourcing . . . . .	45
2.5	Conclusion . . . . .	47

<b>3</b>	<b>Reddit - Change My View</b>	<b>49</b>
3.1	Who: Usecase . . . . .	50
3.1.1	Reddit . . . . .	51
3.1.2	Change My View . . . . .	51
3.2	Why: The will to change . . . . .	52
3.3	How: Graph models . . . . .	55
3.4	What: CMV descriptive analysis . . . . .	59
3.4.1	Life-cycle of Online Communities . . . . .	61
3.4.2	Forum dynamics . . . . .	64
3.4.3	Threads characteristics . . . . .	66
<b>4</b>	<b>Different types of Debates?</b>	<b>81</b>
4.1	Who: Hypothesis — discussion categories . . . . .	82
4.2	Why: Discussion levels . . . . .	84
4.2.1	Reddit — Change My View (CMV) . . . . .	84
4.3	How: Clustering . . . . .	85
4.3.1	Author Networks and Motifs . . . . .	85
4.3.2	Clustering Algorithm . . . . .	86
4.3.2.1	<i>k</i> -means Clustering . . . . .	86
4.3.2.2	Spectral Clustering . . . . .	86
4.3.2.3	Hierarchical and Ward Clusterings . . . . .	87
4.4	What: Results . . . . .	87
4.4.1	Descriptive Statistics . . . . .	87
4.4.1.1	Motif features . . . . .	88
4.4.1.2	Roles in the motifs . . . . .	89
4.4.2	Clustering analysis . . . . .	96
4.4.2.1	Optimal clustering . . . . .	96
4.4.2.2	Clustering results . . . . .	102
4.5	Conclusion . . . . .	104
<b>5</b>	<b>The Consent of the Crowd</b>	<b>113</b>
5.1	Who: Rewarded discussions . . . . .	114
5.2	Why: Individual vs. crowd . . . . .	115

---

5.3	How: Econometrics . . . . .	117
5.3.1	Reddit – Change My View (CMV) . . . . .	117
5.3.2	First delta context . . . . .	119
5.4	What: The Consent of the Crowd . . . . .	124
5.4.1	Descriptive statistics . . . . .	124
5.4.2	Regression modelling . . . . .	133
5.4.3	Result discussion . . . . .	135
5.4.3.1	Results . . . . .	135
5.4.3.2	Discussion . . . . .	140
5.4.4	Conclusion . . . . .	143
<b>6</b>	<b>Conclusion</b>	<b>153</b>
	<b>Bibliography</b>	<b>159</b>

## Remerciements

A l'issue de ces 5 années de travail de recherche de thèse, je tiens à remercier chacun pour le soutien qu'il m'a apporté.

J'adresse d'abord mes remerciements à mes directeurs de thèse, Joanna Tomasik et Jean-Michel Dalle, qui m'ont épaulé au cours de ces années et ont accepté de suivre le rythme discontinu qu'il m'arrivait d'adopter avec d'inévitables périodes de doutes ; ils m'ont toujours encouragé à persévérer. Je remercie également Thierry Rayna, Paola Tubaro, Nicolas Jullien et Matthijs Den Bensten, qui m'ont fait l'honneur d'évaluer mon travail.

Je tiens également à remercier ceux qui ont rendu possible matériellement et financièrement l'aboutissement de mon travail : Frank Pacard, Dominique Rossin, Hervé Dumez de l'École polytechnique et Yannis Manoussakis du Laboratoire de Recherche en Informatique.

J'ai eu la chance d'être accueilli dans trois laboratoires de recherche, ce qui m'a permis de rencontrer un nombre certain de personnes, toutes différentes, qui ont enrichi ma réflexion et ouvert d'autres horizons. Je remercie donc chaleureusement George M. et Johanne C. du LRI. Michèle B. et Marie-Claude C. du CRG, qui accompagnent sans relâche les doctorants du CRG. Je remercie également les enseignants-chercheurs qui nous y encadrent : Christophe M., Rémi M., Florence C.-D., Sihem B.M.-J., Cécile C., Véronique S., Thomas P., Pierre-Jean B., Nicolas M., Akil A., sans oublier Elodie G. Mes pairs avec lesquels je me suis lancé dans l'aventure du MOOC : Héloïse B., Mathias G., Alexandre V. et Haruki S., sans oublier Julien K., Ariadna A. et Julie F. Je ne peux oublier le laboratoire SES de Télécom, qui m'a autorisé à utiliser ses bureaux. Je remercie donc à ce titre David B., Valérie F., Thomas H., Samuel H., Christophe P. et David M. Et bien sûr, les doctorants avec lesquels je partageais cette même aventure : Raphaël C., Constance G., François A., Elie S., sans oublier Sophie B.

Je ne serai pas arrivé à l'École polytechnique sans la chance que m'a donné Eric Bruillard, alors directeur du STEF de l'ENS Cachan. J'ai eu la

chance d'y être apprenti chercheur avant même d'avoir débuté mon doctorat et d'y rencontrer Mehdi K. et Françoise T. Je remercie chaleureusement Matthieu Cisel, qui m'a ouvert sans réserve les portes du monde des MOOC. Avec Jean C. et Eléonore V., nous étions alors les 4 MOOCquetaires (merci Eléonore pour ce trait d'esprit) qui allions révolutionner l'enseignement en France.

Être chercheur, c'est aussi se confronter aux idées du monde entier. J'ai une pensée émue pour Fred Mulder, alors responsable de la chaire de l'UNESCO Open Education Ressources, décédé depuis, qui, avec Sophie Touzé, m'ont permis d'intégrer le GO-GN et d'assister à l'une de mes premières conférences à Banff, au Canada. J'ai alors rencontré de nombreux chercheurs, expérimentés (Bea DLA., Covadonga R., Martin W., Robert S.) qui ont déployé une énergie impressionnante pour guider ces jeunes chercheurs utopistes (Tina P, Dalila C., Francisco I., Gino F., Eyal R. et tant d'autres) que nous étions. Cependant, tout comme au STEF, nous étions tous convaincus que l'éducation est la seule arme dont nous disposons pour améliorer le monde.

J'ai eu la chance, pendant mes recherches, de continuer à m'investir dans l'enseignement à distance en coordonnant en 2016, la production du MOOC *How to write and publish a scientific paper*. Cela n'aurait pas été possible sans Eric Vantroeyen, chargé de mission elearning de l'École polytechnique. Merci Eric de m'avoir accueilli dans ton équipe pour la suite de mon doctorat. Je remercie également Latifa Berkous, ingénieure pédagogique, pour ces longs échanges passionnants et passionnés.

La fin de mon doctorat n'aurait pu se faire sans le contexte dans lequel Tatiana Defrance et Eric Houël, m'ont embauché à l'École Polytechnique Executive Education. Merci à vous deux d'avoir accepté de recruter un jeune passionné, et de m'avoir fait confiance au cours de ces deux dernières années de doctorat. Je remercie également mes collègues, en particulier Larbi T. et Fabien G., pour leur soutien.

En plus de toutes ces individualités dont j'ai eu la chance de croiser le chemin, je remercie évidemment ma famille qui a également suivi (subit ?) mes sauts d'humeur au cours de toutes ces années. Merci donc à Catherine,

Philippe, Vincent, Audrey, Sarah et José.

Merci évidemment à Amandine, Alexandre, Youcef, Sandra, Olivier, Arthur, Thibaut, Marine, Etienne, Antoine, Charlotte, Flora, Inès, Manon, Manon et Jeanne. Un remerciement tout particulier à Martin et Yann pour votre implication, notamment dans la dernière ligne droite. Merci, enfin aux éternels présents : Florent, Nicolas, Léa, Maëlle, Adrien, William, Christophe, Pauline, David, Juliette et bienvenu à nos futurs (Martin et Sophie et aux suivants).

A très vite pour le prochain chapitre !

# Chapter 1: Introduction

The development of the Internet revolutionized multiple aspects of daily life, and particularly how one communicates with another. From the 1980s and the first *newsgroups* to today's online social networks, people commonly exchange messages online. The biggest sites where online discussions take place, such as Reddit<sup>1</sup>, gather millions of threads and users, underlining the importance of such platforms in a social life. As an example, in 2003, 20% of Internet American users reported having visited online newsgroups and fora, 17% having posted written contents on web sites and 10% participated in an online newsgroup (Lenhart, Horrigan, and Fallows, 2004). This trend increased in time: 15% of Internet users in the U.S. exchange actively on fora, in 2015 (Duggan et al., 2015). Those social sites cover diverse range of activities, from looking for an answer to a technical question (after-sale services site) to being kept informed of the latest news (blogs and fora). More importantly, they have also emerged as important places for political discussions. Latest surveys from Pew Research Center bring evidence of such usages. Social networking sites have become a key space to engage in civic-related activities: more than 50% of Americans have engaged on social media in 2018 and more than two third agree with the fact that it helped give a voice to "underrepresented groups" (Anderson et al., 2018). Not only do they allow the expression of political views, but also contribute to the political debate: 14% of Americans affirmed that they have changed their perception about a social issue because of their interaction with social media (Bialik, 2018).

To study such structure, research consider the Internet as allowing a group

---

<sup>1</sup><https://www.reddit.com/>

of individuals, labelled as a crowd, to gather *somewhere* in order to exchange information, with a wider acceptance of the word. Researchers studied the behavior of a crowd long before the development of the Internet, especially in the political sphere. Oldenburg (1999)'s concept of *third place* defines a public space, beyond the house and the workplace where a group of individuals meets and where people can interact on an informal basis. Wright (2012) extends the concept taking into account the technology of the Internet. What if those blogs, electronic bulletin boards or other online fora are the *third space* (Wright, 2012) where people interact on an informal basis? If they are, research needs to dive into the framework and the process of such third spaces in order to understand new behaviors among individuals.

A recent example of such third space takes place at the end of 2018 in France, where several strikes occurred. The movement of the "Gilets Jaunes"<sup>2</sup> gathered thousands of people in the street every Saturday to protest and became important enough to get government's attention. One of the claims from strikers was related to the difficulty for politicians to understand what the daily life of French citizens are. As a direct response, the government opens the "Grand Débat"<sup>3</sup> in January 2019, which allowed anyone to go to its city hall and to record - on paper-, problems which they think could be solved by government undertakings (such as expensive gasoline price). Such political action reminds us of the "cahiers de doléances" opened by the French king just before the Revolution in 1788. Indeed, the government opened an online platform with the exact same purpose: grant a *space* - open to anyone - in order to allow citizens to declare day-to-day difficulties. This contemporary event highlights the necessity for citizens and governments to have a place, a space, to discuss, exchange, debate about the political sphere.

Besides politics, numerous examples of third space can be found online. Wikipedia, the most popular online encyclopedia, is mainly the result of individual production. Launched in 2001, it counted 5,764,000 articles and 35,146,000 registered users at the end of 2018 <sup>4</sup>. A second example can be

---

<sup>2</sup>Yellow jackets.

<sup>3</sup>The great debate

<sup>4</sup><https://expandedramblings.com/index.php/wikipedia-statistics/>

found in Linux, a computer operating system, designed in 1991 by Linus Torvalds, and fully and freely available on the Internet. Since 2005, 15,600 individuals have contributed to the project (Corbet and Kroah-Hartman, 2017). People are cooperating for free, allowing the spread of knowledge or product all around the world. Such capacity of groups to work together have been denoted as *collective intelligence* and have been researched for more than a century (Galton, 1907b; Malone and Bernstein, 2015).

The study of political debate constitutes an important area of research in many fields (economics, law, computer sciences, management studies, ...). The following thesis studies the emergence of collective intelligence on the Internet. More precisely, we study two main research questions. First, is there different process leading to a constructive debate? Second, if a consensus among participants, or a part of them, is reached, under which circumstances does it occur? Answering to those questions will help fora administrators to manage in a more efficient way their community.

We assume, especially for someone looking for information, that one will have different *approaches* to get what he is looking for from someone else. To solve such problem, numerous online fora include a reward system. Many actions such as to upvote or downvote, like, counter of view, to demonstrate our approval or disapproval, are possible in those *third spaces*. Those indicators, once they are aggregated, allow to better sort out information produced by the crowd.

Throughout this manuscript, we attempt to bring answers to those questions to analyze the emergence of a collective intelligence on the Internet. To do so, the thesis use a case study on an online forum open to anyone: Reddit - Change My View <sup>5</sup>. On this forum, an individual exposes his opinion on a subject, and asks to the community to bring him arguments to change his opinion. The database covers threads from January 2013, date of the subreddit creation, to November 2016. The database includes information about 21,564 discussion threads, 1,441,914 posts and 71,775 unique authors. To test the *different approaches* to debate, we apply several clustering algorithms

---

<sup>5</sup><https://www.reddit.com/r/changemyview/>

on particular characteristics of a discussion, modeled as a particular network: the motifs (Milo et al., 2002). We observed (Mano, Dalle, and Tomasik, 2017) that the individual opening a discussion adapts its strategy in respect with outside characteristics, such as the number of challengers he has in front of him.

Concerning the process of consensus, we study the causality of several variables on the consensus itself. We underline a new concept: *the Consent of the Crowd* (Mano, Dalle, and Tomasik, 2018). The consensus between an individual and a crowd could be attained, but under proper circumstances, highlighted in this work.

The current PhD thesis is composed of 4 chapters and a conclusion. Each chapter are organized the same way and follows the development outline in the seminal work of Malone, Laubacher, and Dellarocas (2009). The authors develop a theoretical framework, analyzing hundreds of online collective intelligence actions. They identify common characteristics and use them to define a collective intelligence problem as a genome composed of four main genes. The first one characterizes *who* participates to the project, distinguishing a hierarchical organization from a "crowd" organization without position of authority. The second gene *why*, defines the *motivations* of the participants, distinguishing intrinsic motivations (such as altruism) and extrinsic motivations (such as monetary compensation). The third gene focuses on *what* is being done, distinguishing between a creation and a decision. Finally, the last gene focuses on *how* it is done, distinguishing between independent actions of members of the group, and dependent ones. Each chapter analyzes through these four genes its subject.

Chapter 2 presents the state of the art of collective intelligence. Study of the existence of a collective intelligence could find its origin to Ancient Greece with Aristotle (1944) (Ober, 2009) who claims that "the many" could make decision, under the right conditions, on certain subjects better than individuals or small groups of elite. He labeled it the "the wisdom of the many". First, *who* are we talking about when we are referring to the crowd? The *why* asks the motivations driving individuals to group and act together.

The *how* gene develop the means at stakes to produce a collective intelligence. Several theories have been developed to put a theoretical framework on this phenomenon. Finally, the *what* gene focuses on the results of such collective intelligence. What comes out when people think and work together?

Chapter 3 presents the case study. The current research tests the collective intelligence framework on a real case: Reddit - Change My View (CMV). CMV is an online forum on which individuals argue on personal opinion. The objective is to change the view of opponents, based upon argumentation. This chapter develops, as well, the modeling framework used to analyze the forum: a network modeling (how). And finally, we present the macrolevel evolution of the forum. The database covers a period of four years, from January 2013, creation of the forum, to November 2016. We study group evolution of CMV in this chapter.

Chapter 4 and Chapter 5 present the main contributions of the thesis. Chapter 4, based on Mano, Dalle, and Tomasik (2017) with updated results, ask several questions. Are the discussions following the same process? If not, is there characteristics to distinguish one discussion to the other? It first presents the sub-data set used for this research (who). Then, it dives into hypothesis about discussion categories (why), subject of the research. In order to drive the analysis, we apply several clustering algorithms (how). Finally, we present the results (what), which highlight that discussions could evolve in different ways. Several discussions stopped on a certain level of consensus, whereas others do not. Nevertheless, we highlight the behavior of the individual who has opened the discussion as a key element in the discussion evolution.

Chapter 5, based on Mano, Dalle, and Tomasik (2018) with updated results, brings up the major contribution of the current Ph.D. thesis. We analyse the emergence of a consensus within a discussion. Indeed, CMV challengers look to bring sufficiently good argument in order to change one point of view. The forum managers have developed a dual-rewarding system to highlight the *best* argument, at least those who have been selected by the participants of the discussion. We thus have to focus on rewarded discussion with an

original pruning process (who). Furthermore, one reward is attributed by one individual, whereas the second one is a collective action opening a tension between the individual reward and the collective one (why). Analyzing this dual system through a statistical modeling (how), we bring evidence on crowd behavior to accept or deny an individual reward. We conceptualize this result under the notion of *Consent of the Crowd*: the consent or dissent of the crowd toward an individual reward has an impact on the evolution of the discussion and on the fact that a discussion reaches a consensus or not. We find that a consensus on rewards system is reached if, and only if, the crowd get involved in the discussion before the individual reward.

Chapter 6 discusses the highlighted results and concludes this PhD thesis.

## Chapter 2: Wisdom of the Crowd

2.1	Who: The crowds . . . . .	15
2.1.1	The crowd before the Internet . . . . .	15
2.1.2	Communities of Practices . . . . .	17
2.1.3	Online Community . . . . .	18
2.2	Why: The motivations . . . . .	20
2.3	How: The framework . . . . .	22
2.3.1	Swarm intelligence, Wisdom of the Crowd, or Collec- tive intelligence . . . . .	22
2.3.1.1	Wisdom of the Crowds . . . . .	23
2.3.1.2	Limits . . . . .	26
2.3.2	Wisdom of the Crowds and Democracy . . . . .	27
2.3.2.1	Epistemic Democracy . . . . .	27
2.3.2.2	Procedural Democracy . . . . .	30
2.3.2.3	Deliberative Democracy . . . . .	34
2.3.2.4	Reconciliation of the approaches . . . . .	38
2.4	What: Adding-values to society . . . . .	41
2.4.1	Knowledge sharing . . . . .	41
2.4.2	Open Innovation . . . . .	42
2.4.3	Crowdsourcing . . . . .	45
2.5	Conclusion . . . . .	47

One of the important achievements of Internet platforms is the creation of a space allowing people from any background to gather in a common "place" to create some *outputs*. We will distinguish in this literature review two kinds of outputs: the economic adding-value outputs (such as real online customer services and community of practices, or some less famous, such as Q&A fora) and the social adding-value outputs, enhancing the political thoughtful citizens.

This research will focus on a concept that has fueled interest recently: the "*Wisdom of the Crowd*". This concept, defined and detailed later on, has become interesting for researchers, with the development of the Internet. Thus, this concept is composed of *wisdom* and *crowd*. It shares characteristics with another one: *Collective Intelligence*. Without going into too much detail now, it seems useful to us, as a prelude, to describe the history of the latter to understand the former.

In order to understand Collective Intelligence (CI), we must define both terms. On the one hand, Leimeister (2010) defines "collective" as a group of individuals. Those individuals may not have the same goal or viewpoints. On the other hand, Wechsler (1939) defines intelligence as "the aggregate or global capacity of the individual to act purposefully, to think rationally and to deal effectively with his environment". The MIT Center for Collective Intelligence<sup>1</sup> defines as a collective intelligence a group of people doing things together that seem intelligent. This definition does not constrain the notion of intelligence. Moreover, it involves several individuals, who are tied by some relationships through their activity. Finally, the intelligence of the behaviors depends on the perspective of the observer (Malone and Bernstein, 2015).

Malone, Laubacher, and Dellarocas (2009) develop a theoretical framework. Analyzing hundreds of online collective intelligence actions, they identify common characteristics. Comparing a collective intelligence problem as a genome, authors distinguish four main genes. The first one characterizes *who* participates to the project, distinguishing a hierarchical organization from a "crowd" organization without position of authority. The second

---

<sup>1</sup><http://cci.mit.edu>

gene *why*, defines the *motivations* of the participants, distinguishing intrinsic motivations (such as altruism) from extrinsic motivations (such as monetary compensation). The third gene focuses on *what* is being done, distinguishing between a creation and a decision. Finally, the last gene focuses on *how* it is done, distinguishing between independent actions of members of the group, and dependent ones. Collective intelligence attracts more and more researchers, even more since the important development of ICT. Indeed, the emergence of the Internet allows individuals to be more connected to one another, favorising exchanges and collaboration. Nevertheless, researchers have observed collective intelligence long before the Internet era. In Economics, the Invisible Hand (Smith et al., 1859) defines a mechanism where the collective action of participants in the market makes it optimum. In Biology, researchers have observed collective intelligence among insect species, from the ant colonies (Gordon, 2010) to beehives (Garnier, Gautrais, and Theraulaz, 2007), known as swarm intelligence (O'Bryan, Beier, and Salas, 2020).

The following sections develop each one of the four genes of collective intelligence.

## 2.1 Who: The crowds

*Man is by nature a social animal [...]. Society is something that precedes the individual. Anyone who either cannot lead the common life or is so self-sufficient as not to need to, and therefore does not partake of society, is either a beast or a god. (Aristotle, 1944)*

### 2.1.1 The crowd before the Internet

At the dawn of the previous century, Le Bon (1895) portrays a negative picture of the crowd:

*Isolé, c'était peut-être un individu cultivé, en foule c'est un barbare, c'est-à-dire un instinctif.* (Le Bon, 1895, p.22)<sup>2</sup>

For Le Bon (1895), the crowd, to be understood here as a unified organism (with a biological meaning), does not have the capacity to reason but only to act, and in particular to destroy. Even if Le Bon does not trust the crowd to reason, he underlines the fact that the characteristics of the crowd is, if not better, at least different from the characteristics of its individual parts. In any case, the final aim of a crowd is to act.

A few years later, Tarde (1901) develops a theory to frame a crowd and the opinion which might emerge from it. First of all, he draws more precisely what must be understood by the crowd. He distinguishes an "audience" from a "crowd":

*Le public, en effet, est une foule dispersée.* (Tarde, 1901, p.7)<sup>3</sup>

The concept of crowd implies an organization of its structure. An audience is a spiritual group, whose members are physically separated and only linked mentally. On the contrary, according to Tarde, a crowd is more primitive, its members acting on the brain of each other, also through a physical contact. Indeed, a crowd acts through a communication method, allowing its members to coordinate themselves, in order to reach the common goal, the crowd goal.

Even if they differ on several points, Le Bon and Tarde agree on the fact that an individual belonging to a crowd loses his reason for a collective one, more primitive. Thus, the vision of political philosophers is pessimistic over the capacity of a crowd to produce positive outputs. Contemporary researchers develop a different conception of the crowd, focusing not anymore on political aspect, but also on an economic one. The following sections focus on smaller crowds, in a particular context: communities. Similar to a crowd, a community is the aggregation of individuals, sharing a common goal. Under the right circumstances, detailed below, such a community is capable of producing an output profitable for each individual inside the community.

<sup>2</sup>Alone, he might be cultivated. Within a crowd, he is a savage, an impulsive individual.

<sup>3</sup>An audience is, indeed, an inattentive crowd.

## 2.1.2 Communities of Practices

A whole part of research on human behaviors analyzes how we interact one(s) to another(s). Section 2.1.1 presents the vision of intellectual class about how individuals merge into an entity: a group, a public or a crowd. This section focuses on a group within a particular organization: firms. A company, seen as an entity, is an organization which produces a product or a service in order to sell it to customers. The classical economics define that a company manages resources, under several constraints — the cost function of the firm — and aims to maximize its profits. Those resources are the workforce, on the one hand, and capital in the other (Smith et al., 1859). Becker (1962) developed his famous theory of *human capital* to better understand the *work* resource. This concept underlines the fact that each employee of a firm has his own capabilities, selling his knowledge and skills to his employer.

With the growth of companies, and then their internationalization, firms have developed new needs. Having subsidiaries all over the world increases the quantity of knowledge accumulated by a firm, but make more and more difficult to disseminate it inside the whole company (Guerineau, 2018). Thus a new management framework arises: knowledge management. We develop the theory later in the literature review (Section 2.4.1). But this framework implies an organization among employees in order to share knowledge: the Community of Practice (CoP).

This concept accepts different definitions (Johnson, 2001). Wenger (1998) sets the basis of conceptualizing a CoP: it is an evolving process for learning inside a group. Such group might exist within defined organizations, but outside and between as well. Moreover, the creation of such community, and its consolidation is a longtime process.

Wenger (1998) develops a definition around groups of professionals, sharing common tasks and responsibilities. For the authors and other (Winsor, 2001; Bielaczyc and Collins, 1999), one key feature of CoP is the dissemination of knowledge through communication.

Thus, those CoPs are composed by employees sharing a same practice.

The development of IT improved the transfer of information inside a team, a firm, an international company. Nevertheless, the exchanges stay among members of a same organization. Internet changes this fact allowing organizations to outsource information: seeking for information outside the organization. We develop the notion later in this chapter (section 2.4.2). The following part addresses the concept of online communities: who are these people interacting one with another without belonging to the same organization?

### 2.1.3 Online Community

Benghozi et al. (2001) and Benghozi (2006) develops a typology of communities, and tries to understand what is an online community especially. The development of the Internet allowed the production of softwares which allows individuals to act collectively without being in the same space. In particular, in firms, *Enterprise Resources Planning* arose and solved a part of the geographical issue. But, they also imply the need for cooperation among coworkers (Benghozi, 2006). Indeed, continuous improvement of softwares removes the issue of technological mastering (software becomes more and more *user-friendly*). Still, they imply important consequences on other levels: management, organizational, work practices (Benghozi, 2006).

Besides, we observe online communities outside companies. As defined by Kraut et al. (2011), an online community (OC) is a virtual space where individuals come together to interact with others (converse, exchange resources, play). Similar definitions are developed in different disciplines (Reingold, 1993; Hagel, 1999; Andrews, 2001; Lee, Vogel, and Limayem, 2003; Iriberry and Leroy, 2009). The creation and development of a new OC faces several challenges. First of all, designers and managers are faced with a critical problem: in order to attract new members, they need an important quantity of content but do not yet have the sufficient number of members. A second challenge is that once the OC is established, it still needs to attract newcomers, in order to replace those who leave. Attracting and socializing those

new members is a challenge because their first interaction with the OC will have an important impact on their commitment, and at the same time, they will disturb for a period, the activity of previous members. Besides, managers need to enhance commitment. Commitment is a feeling of attachment and connection to the community. And if members are committed to the community, they tend to both be more satisfied and perform better (Mathieu and Zajac, 1990).

Finally, as highlighted by Faraj, Jarvenpaa, and Majchrzak (2011), some members can behave in an uncivilized way. Thus, a community needs a framework to regulate behaviors. The difficulty is even higher for an OC due to the anonymity of its members, an ease to enter and exit the community, and the textual communication.

Iriberry and Leroy (2009) develop precise characteristics of OC. The very core of an OC is its activity and the creation of contents. As suggested by the core-periphery model (Borgatti and Everett, 2000), almost all the content is created by a small number of members. But the issue is not the inequality of contributions, but the possible under-contribution. Hagel (1999) defines online communities as “computer-mediated spaces where there is potential for the integration of content and communication with an emphasis on member-generated content.” Lee, Vogel, and Limayem (2003) supports this definition. Furthermore, they ascertain that the content created in online communities brings value to business organizations.

Besides, Millen, Fontaine, and Muller (2002) highlight benefits for organizations which gather such communities. From a customer’s point of view, it increases loyalty. Furthermore, it allows the organization to gather feedback and information on customer needs and requirements, directly from customers, improving organization customer service. Alongside, it increases the organization visibility and reputation. From an employee’s point of view, it increases his trust, increases internal communication allowing everybody to follow all company projects. Then, OC has a direct impact on the productivity of the company, increasing the quality of knowledge, idea creation, product innovation and enhancing problem solving process.

## 2.2 Why: The motivations

Theoretically, the increase in information sharing should improve global efficiency of organizations. But are individuals willing to share their information? How to make employees develop internal mechanisms in order to share their knowledge? Constant, Kiesler, and Sproull (1994) shine lights on those research questions. The theory of interdependence (Kelley and Thibaut, 1978) demonstrates how the (social) environment might put pressure on individuals, producing negative behaviors. An organizational environment might make an employee share his knowledge with another, even if he does not want to but it would have negative impact on his global production. In day-to-day life (cf. Section 2.3.2.3), an individual will easily act, share informations, whereas in an organization, he might not capture this information sharing as a social good (an act, a behavior that may be personally costly but would be beneficial to the organization in the long run). How to understand this fact?

There is a distinction between tangible information (seen as a product - such as a document) and intangible information (considered as expertise). Authors assume that, in response to a coworker who had failed to help in the past, people would be more likely to share expertise than a document. The meaning to people of intangible information such as expertise is different than is the meaning of tangible information such as a computer program. The former reflects on its possessor's identity and inner qualities, and that sharing it can have direct personal benefits.

Chiu et al. (2007) highlight factors that increase or reduce individuals' satisfaction in knowledge sharing in open virtual professional communities. Organizations have understood that they do not have at their disposal all the required knowledge within their formal boundaries., some of them developed professional virtual communities in order to fill that gap. But individuals do not necessarily want to share their knowledge, because of the fear of losing their comparative advantages.

A professional virtual community is defined by three dimensions: its

members, the social network they develop and the knowledge they share. Individuals might be motivated to share knowledge because they expect future rewards, intangible and tangible benefits. Authors base their analysis upon an enhanced model of the expectancy disconfirmation theory developed by Oliver (1980), finding core motivations for the continuance intention in knowledge sharing.

Faraj, Jarvenpaa, and Majchrzak (2011) theorize on how Online Communities (OC) engage in knowledge collaboration. An OC might be seen as a typical organizational structure, but characterized by constant changes (members, contents). Aside from the classical behaviors of knowledge exchanges, an online environment adds possibilities. Members can also recombine, modify, and integrate knowledge that others have contributed to. One can witness knowledge collaboration in OCs despite the lack of direct social relationships. Thus, for Faraj, Jarvenpaa, and Majchrzak (2011), the access to resources cannot explain solely the collaboration. It is due to the unique characteristic of OC, distinguishing it from traditional organization structures: its fluidity. Authors divide this fluidity into five tensions, associated with five resources that have an impact on knowledge collaboration in OCs. The passion of its members (the more passionate will invest more — time, effort — in OC; can be a barrier to collaboration); an important amount of time is required from members (but if few members spend too much time, they may impact the knowledge collaboration process by rejecting newcomers); anonymity (encouraging participation focusing more on the merit rather than the status, but can imply bully behaviors, and even decrease participation if members have the fear to not get any credit for their work); convergence toward a single direction (temporary and incomplete, situated among a subset of actors rather than the entire community). To counter uncivilized behaviors, the OC needs structural mechanisms (such as formal roles and participation rules).

In order to motivate members, OC managers use rewards, and especially rewards for contributions (Andrews, 2001). Member recognition, settled on psychology, is a wide area for researchers (Andrews, Preece, and Turoff, 2001;

Ginsburg and Weisband, 2004; Beenen et al., 2004; Hall and Graham, 2004; Tedjamulia et al., 2005; Butler et al., 2007). Providing rewards for contributions seems to increase the number of messages posted by community members, making it more active and more successful (Iriberry and Leroy, 2009).

## 2.3 How: The framework

### 2.3.1 Swarm intelligence, Wisdom of the Crowd, or Collective intelligence

In companies, projects are nowadays managed by teams and not anymore by one individual (Ilgen, 1999). The assumption behind this evolution lies in the belief that individuals are better, stronger, more effective, when working in teams to solve a problem. Furthermore, several researchers study how companies, and organizations more broadly, could benefit from the *performance and cognitive* advantages teams may provide. This concept has been framed by the term *collective intelligence* (Kurvers et al., 2015). Other research leans on the study of this mechanism but in the non-human animal reign (O'Bryan, Beier, and Salas, 2020). Specifically, in ant or beehives (Beshers and Fewell, 2001). Biologists prefer the concept of *swarm intelligence*. It happens that those two concepts are used one for the other (Krause, Ruxton, and Krause, 2010). But when specific, they are applied to particular forms of group-level intelligence. On the one hand, *swarm intelligence* is applied when the group under study is simple, insects for instance (Garnier, Gautrais, and Theraulaz, 2007). On the other hand, *collective intelligence* is applied to a group of individuals with high capabilities, human beings (Salminen, 2012). Nevertheless, within the human realm, we can be more precise. First, researchers have studied human groups similarly as ant hives (Moussaïd, 2019). Furthermore, when the group is large, such as our crowds or online communities (as defined in Section 2.1), researchers use the concept of *Wisdom of the crowd* (Galton, 1907a; Surowiecki, 2005), as detailed in the following section. And when the group is small, researchers use the concept of *collective intelligence* (Weschler,

	Swarm Intelligence	Wisdom of the Crowd	Collective Intelligence
Representative Groups	Simple animals (e.g., social insects)	Humans (non-interacting individuals)	Human small groups and teams
Inputs	Personal information, behaviors	Individual estimates	Individual traits (e.g., intelligence, knowledge)
Combining Inputs	Interactions between individuals	Combination by centralized entity	Team processes

TABLE 2.1: Differentiating forms of Group-Level Intelligence - O'Bryan, Beier, and Salas (2020)

1971). O'Bryan, Beier, and Salas (2020) offers to focus the distinction on what is members' input and how those inputs are combined. Table 2.1 summarizes it.

As already defined in introduction of this chapter, the current research focuses on the concept of *Wisdom of the Crowd*, developed in the following section.

### 2.3.1.1 Wisdom of the Crowds

In his seminal book, Surowiecki (2005) develops the notion of Wisdom of Crowds (WoC). With several case studies, he observes that under right circumstances, a group brings a better solution to a problem than an expert on the subject. The same idea supports Communities of Practices (cf. Section 2.1.2), for which the total amount of knowledge is higher than the sum of individual knowledge (Gherardi and Nicolini, 2000). Galton (1907b) provided early evidence of the existence of a WoC, by comparing the average estimation of an ox weight from a crowd with the actual weight. The crowd performed surprisingly well, with an estimation error lower than 1%. Way before, in *Politics*, Aristotle affirms:

*For it is possible that the many, though not individually good men, yet when they come together may be better, not individually but collectively, than those who are so.* (Aristotle, 1944, Chapter III, 1281.a - b).

The development of information technologies has considerably renewed this interest in the WoC. four conditions are necessary to observe a wise

crowd: diversity, independence, decentralization, and aggregation (Surowiecki, 2005). The following paragraphs develop each one of these characteristics for the group decision-making model analyzed in this section.

### **Diversity**

How do bees of a hive find flowers? Through a twofold process. First, the hive sends scouts in several directions. Then, the scouts dance for the hive and the intensity of dances describes the best flowers localization. Thus, the first aim is to discover the maximum possibilities, then to choose the best option. Similarly, best innovations were the output of a contest among hundreds of possible innovations (Terwiesch and Ulrich, 2009). But at the end, there are few options left. Thus, there is a need in diversity of options. But what about diversity in members of the decision makers group? Page (2008) highlights the importance of the diversity among members of a group — regarding "intelligence", social background, skills, etc. — through an experiment where a group of heterogeneous skills members outperforms a group with only highly skills members. Numerous research studies draw the same conclusion (Hong and Page, 2004; Aggarwal and Woolley, 2013; Aggarwal et al., 2015; Srba and Bielikova, 2015). For instance, it improves the productivity of an individual in companies, and the bigger the group is, the more diverse it is (Aral, Brynjolfsson, and Brynjolfsson, 2006). Nevertheless, at an individual level, homogeneity of the knowledge is preferable (Adamic et al., 2010). Similarly, FoldIt "gamers" were recently acknowledged for solving the structure of an AIDS-related enzyme after scientists' "failure of a wide range of attempts to solve the crystal structure of M-PMV retroviral protease by molecular replacement" allowing for "new insights for the design of antiretroviral drugs" (Khatib et al., 2011).

### **Social influence/independence**

Nevertheless, the WoC does not always perform better. One of the obstacles comes from social interaction (Lorenz et al., 2011). If and when members of

the crowd have the possibility to exchange about their proposals, a decrease of the diversity can be observed, due to psychosocial mechanisms such as conformism (Asch, 1951), social proof (Milgram, Bickman, and Berkowitz, 1969) or information cascade (Bikhchandani, Hirshleifer, and Welch, 1992). However, under the right circumstances, social influences can have a positive impact on WoC (Madirolas and De Polavieja, 2014; Becker, Brackbill, and Centola, 2017). In particular, based on the dataset from Lorenz et al. (2011), Farrell (2011) demonstrates how the information sharing has increased the confidence of members of the group in their own proposal.

### **Decentralization and aggregation**

Decentralization is a system where decision process is not fully in the hands of one part of the system. Rather, decisions are made by parts of the system from their local perspective and knowledge. Moreover, decentralization fosters specialization, which increases the productivity and efficiency of individuals (Smith et al., 1859). Besides, all knowledge cannot be easily passed on, because of its specificity, to its local application. This kind of knowledge is known as tacit knowledge (Hayek, 1952). Closely related to tacit knowledge is the assumption related to decentralization: the closer an individual is to an issue, the more likely she will have the correct solution. Which also allowed to members of the system to improve their coordination. The main weakness of decentralization is the level of global valuable knowledge exchange among the parts of the system. Such system thus needs aggregation to bring a global value to a local knowledge. The aggregation of knowledge is directly dependent on both, transmission and receipt. Concerning the recipient, it depends on its ability to add new knowledge to its current knowledge. Both transfer and aggregation abilities of knowledge are a major key for an optimal location of the decision-making process within a system (Grant, 2014).

## Summary

Surowiecki (2005) defines a hierarchy between these characteristics. Not in terms of importance, because the four are required, but in terms of ordering. For instance, in Aristotle (1944) scenario of the excellent-judging group, the process of aggregation works because the group is diverse in the right way, diversity becoming a condition for aggregation.

### 2.3.1.2 Limits

Under particular circumstances, the wise crowd might transform itself into a mob, being victim of the "groupthink" effect (Janis, 1972; Janis and Mann, 1977; Janis, 1982). Janis defined *groupthink* as:

*[...] a mode of thinking that people engage in when they are deeply involved in a cohesive in-group, when the members' striving for unanimity override their motivation to realistically appraise alternative courses of action. (Janis, 1972)*

In such a group, the preservation of the collective is the first goal, which should be protected at any cost. It is characterized by three pillars: overestimation of the group, closed-mindedness and pressure toward uniformity. Nevertheless, even if those characteristics seem to prevent a group from producing a positive output, Janis clarifies that all bad calls are not the result of an out-think. Even more, an out-think might succeed. According to Shaw (1964), groups with a higher level of cohesiveness are more effective in achieving their purposes than groups with a low level of cohesiveness. But Janis (1991) highlights the fact that the more the bonds in a group are strong, the more "independent critical thinking will be replaced by groupthink". The harmony of the group becomes the first goal, which inclines members to avoid any contradictory arguments.

### 2.3.2 Wisdom of the Crowds and Democracy

The concept of wisdom within a group can be traced back to Ancient Greece with Aristotle's concept of *wisdom of the many* (Aristotle, 1944). Furthermore, the four characteristics defined by Surowiecki (2005) are similar to research in optimal democratic process. The current section tends to study major research lines which have developed theories and concepts linking the wisdom of a crowd and the democracy. Indeed, the current work, specifically its empirical part, studies how a group makes democratic choices in a context of online argumentative discussion. Understand how a democracy works might bring important highlights on how an online community, focused on debating, succeeds in not to be torn apart by its own members.

We present, in the following, the main definition, characteristics, and limits of three theories of democracy: Epistemic, Procedural and Deliberative.

#### 2.3.2.1 Epistemic Democracy

In Schwartzberg (2015), the author traces the history of Epistemic Democracy, precisising there is "no unequivocal defenses of epistemic democracy in the history of political thoughts." Nevertheless, it develops numerous arguments in favor of the *wisdom of the many* (Aristotle, 1944). For the author, we observe that *Democracy* tends to make right decisions, making this political regime "reliable", in favor of the general will. Schwartzberg (2015) develops four distinct times in Epistemic Democracy history. First, ancient Athens. Aristotle (1944) defines epistemic democracy as a *political decision-making* process where the only purpose is to unveil the *best solution* to of common and societal issues. Furthermore, he supposes that a society is able to properly identify common concerns and, through a proper use of *phronêsis* - practical wisdom, would be able to select the *right* policies, defending those common interests. Recently, Ober (2010) and Ober (2013), who has brought up to date Aristotelician's texts, precisises the promise of epistemic democracy. Under the proper prerequisites, a decision-making process, which expresses and defends democratic values would do better than randomly choosing among

policy options. Ober affirms that success of Athens relies on its ability to gather and aggregate the "dispersed knowledge of its citizens" (Ober, 2010). The epistemic process relies on three steps. First, an *aggregation* step, based upon participation of citizens in the decision-making process, which is possible through two premises. First, for any decision to make, it exists a better option. Second, this better option is identifiable, under the right conditions, by the decision-makers. The second step concerns *alignment*, allowing people with common preferences to coordinate. And finally, the *codification* step, transforming past choices to become "action-guiding rules" for future decisions. Nevertheless, as Ober acknowledges, Athenian democracy reveals a lack of equality and inclusivity, excluding "those deemed inferior in cognitive ability (women, slaves)" (Schwartzberg, 2015). Does it make Athens an aristocracy or an epistocracy (Estlund, 2003; Estlund, 2009)? For Ober (2009), Aristotle (1944)'s "Wisdom of the many" defines clearly that *the many surpass the part or the best*, the several best arguments in favor of democracy.

The second historical period concurs with the writings of Rousseau (1782) and Condorcet (1785). They provide a new argument concerning the *right* choice in politic, meaning choice toward the general will. Indeed, when a law is voted, each citizen chooses independently if the law's outcome goes toward the general will, rather than an outcome toward an individual will. Besides the question of general will, going along with a "right answer to any political question", begins to rise questions. The third historical time relies on the work of Mill (1998). Mill developed an epistemic liberalism (Landemore, 2017), defending free discussion (as a liberal value) and assumption of fallibilism, which ensure an identification and a security of "the truth". Furthermore, Mill affirms that to discover this "truth", we need to develop "a set of institutions protective of the individual liberty of inquiry and exchange" (Schwartzberg, 2015), whereas his predecessors relied only on democratic decision making to lead society to wise choices. Finally, "political discussion" allows one to identify his own and common interests, idea in favor of deliberative democracy, developed later on in Section 2.3.2.3.

The last historical period covers the 20<sup>th</sup> century, with the development

of a pragmatist epistemic democracy, through Dewey (1927)'s work. Defining revisability, experimentalism, deliberation and diversity as core values in pragmatism democracy, he emphasizes the importance of *distributed knowledge* and development of democratic institutional design to test and harness this knowledge. Furthermore, pragmatism framework follows a scientific model of searching for the truth, supporting "the superior knowledge of experts". This goes in contrast with notions of inclusivity and equality, which are more important to epistemic democracy than the research of the truth (MacGilvray, 2014). In the 1980's, a new turn is taken in response to the Social Choice Theory (Arrow, 2012), defined in Section 2.3.2.2, questioning the notion of a popular will (Riker, 1988). Recent epistemic researchers are aware of epistemic theory's limits and update the theoretical framework taking into account previous critics. Coleman and Ferejohn (1986) and Cohen (1986) defend the notion of general will, based on Rousseau's work (Rousseau, 1782). They, along with List and Goodin (2001), develop the idea that democracy is the "best imperfect epistemic procedure" to track the truth (Estlund, 1998; Estlund, 2009). Expressed through vote, judgments of majority thus provides an "imperfect procedure", allowing society to identify the general will.

Finally, researchers detect limits of epistemic democracy theory. First, Ober (2013) emphasizes the risk of transitivity, cycling, already highlighted by Condorcet (1785). Under particular conditions, none aggregative process would be able to extract the general will from individuals' preferences. This puts a warning on the framework of the process. Second, in the same work, Ober (2013) challenges the fact that epistemic democracy theory should concern experts solely. If the goal is to reach the truth, by making the *right* choice, experts of the given domain are more likely to do so than non-experts (e.g. the Callipolis from Platon, 2002). Third, Schwartzberg (2015) reminds us why epistemic democracy is still nowadays controversial. This theory embeds flaws concerning suspicion about the deliberative part, because social exchanges implies a risk of coercion, the appeal to comprehensive doctrines, echoing to Janis (1991)'s *group thinking*. Participation in decision-making allows for a democracy to be epistemic if citizens source their knowledge, to

make choice, on "true expertise and genuine experts" (Ober, 2013).

### 2.3.2.2 Procedural Democracy

This section presents theories focusing more on the question of "how" than on the question of "why", more on the methodology to take decision in democracy, rather than to choose the *right* decision.

Directly developed in opposition to epistemic democracy, procedural democrats define the aim of democracy as a regime to embody "procedural virtues" (List and Goodin, 2001). They postulate that such a thing as right social outcomes does not exist. Instead, "it is the application of the appropriate procedure which is itself constitutive of what the best or right outcome is" (List and Goodin, 2001). Thus, what is the best democratic process to make decisions, if it has on the one hand, to defend democracy's core values (liberty, equality and dignity as defined in Platon (2002)), and on the other, defends citizen interests? Platon (2002), along with other ancient philosophers, developed the idea that defending liberty and equality, democracy leads citizens to make decisions on false opinions rather than on knowledge, in order to pursue individual desires rather than real interests. Dahl and Shapiro (2008) offers an answer: aggregation of preferences by counting individual votes, with an equal weight. Doing so, democratic core values are preserved and citizens may stay focused on their own interests. The risk is ignoring or harming the basic interests of a minority, whom interests would differ from the majority. That is why certain interests are raised to the status of right, legally protected. List and Goodin (2001) sum up the problem:

*Classical debates, recently rejoined, rage over the question of whether we want our political outcomes to be right [epistemic] or whether we want them to be fair [procedural].*

List and Goodin (2001) define a *narrow* framework versus a *broader* one for procedural democracy. On the one hand, the narrow form of procedural democracy defines a framework to transform individual preference into social decisions. First, procedural democrats define a set of minimal rules

(such as the weak Pareto principle, condition of transitivity of social orderings, defined by Arrow, 2012). Then they select, if it exists, an aggregation procedure satisfying the rules. On the other hand, the broader version of procedural democracy is more focused on a set of political and institutional arrangements to reach social decisions. In particular, procedural democrats wonder which political processes should lead to social decisions, what is the role of political communication, who should be a voter, and the time line of elections (rather frequent). Furthermore, any aggregation system should be preceded by a process of political deliberation, allowing anyone who is affected by a decision to be heard. Finally, they remain attentive to risks specific to elections. It should always be "free and fair", without corruption or intimidation. In order to do so, rules of election should be known by everybody, as a common knowledge.

As depicted in Schwartzberg (2015), judgment democracy, based on the *respect for individuals' judgments and need of institution to test those judgments* epistemic criteria, highlights the importance of *deliberation* to perfect individuals' judgments. Furthermore, they emphasize the value of aggregation as the mean to affirm individual's dignity (Waldron, 1999). Even if it is developed in an epistemic framework, we can see the link with procedural democracy. Indeed, judgment democracy "offers a proceduralist or intrinsic justification" (Schwartzberg, 2015) to democracy's legitimacy (Christiano, 1996; Christiano, 2008; Dahl, 1989; Waldron, 1999).

Similarly, Ober (2013) develops the theory of Independent Guess Aggregation — IGA. Canonical forms of IGA assume voter independence – there is no pre-decision information-sharing. Independence is valued as preserving freedom of individual choice, but also because it prevents the informational cascades (group-think) and polarization (extremism) that have been associated by Cass Sunstein, among others, as inherent anti-epistemic features of deliberation (Sunstein, 2000; Sunstein, 2002b; Mendelberg, 2002). Pre-decision communication among decision-makers, in ways that violate the independence of their individual choices may be taken as a source of corruption (List and Pettit, 2004). It has to be noted that recent researchers try to

reconcile epistemic and procedural theories.

In a multiple-choice problem, procedural democrats offer different social decisions rules (Condorcet pairwise comparisons, the Borda count (De Borda, 1781), The Hare or Coombs systems (Grofman and Feld, 2004)). Arrow (2012) developed the Social Choice Theory. Dryzek and List (2003) specifies that it is a mathematical theory of group decision making. On the one hand, it is normative - defining specification that the aggregation process must satisfy. On the other, it is logical - based on the specification, the choice of the aggregation process is logical. By definition, it is not an empirical modeling on the way group could take decisions by aggregating their individual preferences and transforming it into group decisions. In any case, the choice of the procedure implies a choice of the social virtues' priority. In the following we present the Condorcet Jury Theorem Condorcet (1785) as an example of procedural mechanism. He demonstrates, through a mathematical model, that under the right conditions, the majority of a group, with limited information about a pair of alternatives, has a higher probability to choose the "better" alternative than any one member of the group. List and Goodin (2001) phrase it as follow:

*If each member of a jury is more likely to be right than wrong, then the majority of the jury, too, is more likely to be right than wrong; and the probability that the right outcome is supported by a majority of the jury is a (swiftly) increasing function of the size of the jury, converging to 1 as the size of the jury tends to infinity.*

This result relies on three assumptions. First, the existence of a "better" alternative: it is a binary-choice problem. Second, individuals vote independently. And finally, they share a common goal: reach the "better" alternative. An independent-guess aggregation process which match in several points with WoC Surowiecki (2005) framework. Nevertheless, several researchers challenged those assumptions (Nitzan and Paroush, 1982; Grofman, Owen, and Feld, 1983; Miller, 1986; List and Pettit, 2004). Ladha (1992) underlines

the importance of the size of the group. The larger it is, the higher the probability of the "correct" choice being elected will be. And it is easier to release Condorcet's assumptions in such case. Ladha offers a framework with taking into account the limits of Condorcet's assumptions, introducing correlation between votes. Indeed, the independent assumption is the most restrictive. Lindley (1985) explains why this assumption could not be hold: if it is true, we would not observe opinion leaders, communication among voters, no common information. This common knowledge Halpern and Moses (1990) is the main source of correlation. Therefore, one cannot hold this assumption, studying a real-world event. Finally, as List and Goodin (2001) explained, in a binary-choice problem, epistemic and procedural democrats agree on the outcome and converge on the *majority winner*. They differ only on why the outcome is the proper one. However, they diverge in a multiple-choice problem.

Ober (2013) highlights the limits of IGA and procedural democracy. First, the assumption of *the majority is right* is not flawless. Then, IGA comes with an external agenda control built in, with a minority (elites) willing to rule a majority. Furthermore, the voter's independence is not realistic as well. Some opinion leaders influence behaviors, not necessary on factual basis. This makes votes potentially dependent on few schools of thought (Ladha, 1992). Furthermore, Arrow (2012) defining its Social Choice Theory, generalizes Condorcet's paradox of cyclical majority preferences (Condorcet, 1785). The latest demonstrating that under particular conditions, we could not extract a collective decision from individuals' preferences. Arrow proved the non-existence of any aggregation mechanism satisfying a set of seemingly innocuous conditions. Any democratic decision mechanism thus exhibits at least one of the following flaws: a failure to generate a determinate social ordering for certain profiles of personal preference orderings; inefficiency by sometimes ranking Pareto-suboptimal alternatives above Pareto-optimal ones; manipulability by changes of the set of initial alternatives (the 'agenda'); or dictatorship.

Finally, Dryzek and List (2003) remind us that, specifically with Social

Choice Theory, researchers model individuals as *homo economicus*, using maximizer functions. But they highlight the fact that this modeling does not imply necessarily any behavioral assumptions. If we relax one of Arrow's conditions, one could produce a model satisfying all the others. This model could be used for democratic decision making.

### 2.3.2.3 Deliberative Democracy

“Tout le monde a quelque chose en soi pour contribuer à la vérité” (Aristote, 2007)<sup>4</sup>

One major criticism of Procedural Democracy is its independence assumption: one votes along its own will and preferences. According to Riker (1988), any notion of a popular will independent of the mechanism used to aggregate preferences was untenable. Rawls (1997) had rejected its use on the grounds that the “votes of different persons are not independent because of the importance of deliberation”. Cohen (2009) and Dryzek and List (2003) define Deliberative Democracy as the foundation for a legitimacy of democracy, founded on the capacity of those affected by a collective decision to deliberate in the production of that decision. Deliberation allows a change in citizen preferences with light of persuasion (not manipulation or coercion). This theory has not been developed in opposition with Epistemic Democracy but instead generally to re-establish deliberation as instrumental to the aim of correct decision making (Schwartzberg, 2015).

In 1988, Bobbio (2005) makes the following analysis. First, our modern societies are polycentric, and large organizations have appeared. The political power passes from collective actors to collective actors and it is less and less acquired by associated individuals. Second, those collective groups have multiplied, and they have competing interests, which makes impartial will-formation more difficult. Third, the development of state bureaucracies have fostered the development of expertocracy (Grek, 2013). Finally, the mass, to be understood as citizens who are not representative of others, have become

---

<sup>4</sup>Everyone has something in him to contribute to the Truth.

apathetic and alienated from the experts. Our democracy tends more to oligarchy which lead voiceless citizens.

Jürgen Habermas has developed an abundant theory about communication, deliberation and argumentation in our societies. In Habermas (2002), he has developed a Theory of Truth and Knowledge, elaborating thoughts on empirical truth, defining a *proposition* as the truth-bearer.

*the universal-pragmatic meaning of truth... is determined by the demand of reaching a rational consensus*

Being close to Epistemic Democrats (cf. Section 2.3.2.1) on the existence of an overall truth, he settles the deliberation as the foundation to reach this truth, making it closer to Procedural Democrats (cf. Section 2.3.2.2). He affirms again this basis in Habermas (1996) defining a Discourse Theory, which calls for a pragmatic analysis of argumentation as a social practice. In argument-making practices, Habermas distinguishes three aspects: an argument as a product, as a procedure, and as a process. Those aspects are aligned with traditional views on argument evaluation of *logic, dialectic, and rhetoric*. First, at the *logical level*, argument is a product for participants, a set of reasons to support conclusion. An argument is logical if one has taken into account all the relevant information and possible objections. Second, at the *dialectical level*, one has to submit arguments and counterarguments to sufficiently severe procedures of critical discussion. Habermas (1984) and Habermas (2015), in his Theory of Communicative Action - TCA, talks about a "ritualized competition for the better arguments". Discussants have dialectical obligations: offer an argument, answer all relevant challenges and brings proof to his view. Third, the robustness of the dialectical level relies on the *rhetorical quality* of the persuasive process. Habermas (1996) defined the *deliberative process* as follows .

- a Processes of deliberation take place in argumentative form, that is, through the regulated exchange of information and reasons among parties who introduce and critically test proposals.

- b Deliberations are inclusive and public. No one may be excluded in principle; all of those who are possibly affected by the decisions have equal chances to enter and take part.
- c Deliberations are free of any external coercion. The participants are sovereign insofar as they are bound only by the presuppositions of communication and rules of argumentation.
- d Deliberations are free of any internal coercion that could detract from the equality of the participants. Each has an equal opportunity to be heard, to introduce topics, to make contributions, to suggest and criticize proposals. Additional conditions specify the procedure in view of the political character of deliberative processes:
- e Deliberations aim in general at rationally motivated agreement and can in principle be indefinitely continued or resumed at any time. Political deliberations, however, must be concluded by minority decision in view of pressures to decide. Because of its internal connection with a deliberative practice, minority rule justifies the presumption that the fallible minority opinion may be considered a reasonable basis for a common practice until further notice, namely, until the minority convinces the majority that their (the minority's) views are correct.
- f Political deliberations extend to any matter that can be regulated in the equal interest of all. This does not imply, however, that topics and subject matters traditionally considered to be "private" in nature could be a fortiori withdrawn from discussion. In particular, those questions are publicly relevant that concern the unequal distribution of resources on which the actual exercise of rights of communication and participation depends.

g Political deliberations also include the interpretation of needs and wants and the change of prepolitical attitudes and preferences. Here the consensus-generating force of arguments is by no means based only on a value consensus previously developed in shared traditions and forms of life.

Furthermore, he defines the "discourse principle" (D), which specifies that a choice is justified only if all those affected by the rule or choice can accept it in a reasonable discourse. Additionally, Habermas distinguishes different types of validity claim (such as sincerity; truth and rightness; authenticity) with different types of justificatory discourse. Then, one needs to add a further specification of (D) matching with those different types of discourse, from the *decision-oriented deliberations* (political decisions) to *informal processes of opinion-formation in the public sphere*.

The main objective of Habermas with his theories is to understand how to legitimate our western societies. A society is persistent over decades if its citizens perceive it as legitimate, to be understood as the society is organized in accordance with what is good, right and true. In order to do so, modern law has to defend the private autonomy of its citizens, which goes along with individuals with an equal citizenship status, being able to understand the law. The citizens must be able to act like lawmakers, which imply, for Habermas, rights to political participation. He highlights a double autonomy: private (individual rights) and public (citizen as lawmaker). They are "equiprimordial": one can be realized if and only if the other is realized. Citizens shape their individual freedoms through the exercise of public autonomy. Furthermore, this idea of public autonomy goes along with the fact that the lawmaking process must be based upon robust public discourse, influencing legislative organism. As Dewey (1927) depicted it,

*The essential need, in other words, is the improvement of the methods and conditions of debate, discussion and persuasion.*

Nevertheless, Deliberative Democracy presents some limits. Ober (2013) highlights the fact that deliberation brings a risk on individual liberty of

thinking. Independence is valued as preserving freedom of individual choice, but also because it prevents the informational cascades (group-think) and polarization (extremism) that have been associated by Cass Sunstein, among others, as inherent anti-epistemic features of deliberation (Sunstein, 2000; Sunstein, 2002b; Mendelberg, 2002). Pre-decision communication among decision-makers, in ways that violate the independence of their individual choices may be taken as a source of corruption (List and Pettit, 2004). Contrary to IGA, Deliberative Democracy can set its own agenda: what and when would be the issues to take up. Strong forms of Deliberative Democracy require equality of deliberative opportunity and are committed to rules of neutrality. As such, even when it does not require complete consensus, its conventional forms lack a mechanism for closure, and thus cannot offer a practical way forward in time-sensitive decision contexts without violating its own premises.

#### 2.3.2.4 Reconciliation of the approaches

We have presented three schools of thought on democracy: epistemic, procedural and deliberative. Based upon major researchers, we bring on the framework and the limit in each theory. We highlighted some struggle and important divergence points between those theories. Nevertheless, even among researchers referring to themselves as a member of one or another school, we had difficulties to draw a clear frontier between them. In this section we will present works establishing robust bridges among these theories.

First, (Schwartzberg, 2015) reminds us that epistemic democrats usually define deliberation as a major input to produce the best outcomes that drive the agenda. Nevertheless, this school of thought is ambiguous about its approach of deliberation. When Habermas (1996) defines the output of deliberation, which is followed by an aggregation process (majority of votes choose the correct answer), as a "reasonable basis for a common practice", he does not elevate the outcome to necessarily being the correct one, which is what pure epistemic democrats would have done. Indeed, he warns us precisely that majority opinion is fallible, and might be changed if the "minority

convinces the majority that [the minority's] views are correct". Furthermore, Knight and Johnson (1994) argue that deliberation must accompany rather than replace aggregation process. They affirm that deliberation reduce the probability of conflict, improving the stability of democratic decision making. Nonetheless, we must take attention to deliberation, which could bring along group polarization and other biases (Sanders, 1997; Sunstein, 2002a).

Dryzek and List (2003) offer a model to reconcile Procedural (Social Choice Theory) and Deliberative democracies, based on Sen (1986). First we must ask ourselves two questions when studying aggregation processes. First, what is the input of the aggregation? Second, what is its output? Dryzek and List (2003) focus on an input of views of individuals, which are expressed by them, through a voting system for instance, which leads to a decision as an output. They go further on the aggregation process asking two new questions. First, are, prior to the decision, the individuals' view subjected to

- no deliberation - **n**, or
- deliberation - **d** ?

Second, are, when the decision is finally taken, individuals' views expressed

- by anonymous voting - **v**, or
- in a group decision - **d** ?

Dryzek and List (2003) then define four scenarii. When we observe no deliberation and an expression by anonymous voters - **nv**, it matches a *pure aggregative decision making process*. In contrast, when we observe a deliberation and an expression in a group decision - **dd**, it matches a *pure deliberative case*, which lead to a consensus decision. From deliberative democrats view, the second scenario is more desirable than the first one. But they do not reject totally voting-based processes. In the case of deliberation and vote - **dv**, it might work if the number of voters is large or if social coercion is prevented (Fishkin, 1995). The last case, a group decision with no deliberation - **nd**, seems unrealistic.

Dryzek and List (2003) conclude that Social Choice Theory and Deliberative Democracy are mutually supportive. The first one shows to the second what it has to do in order to produce a robust collective decision.

Ober (2013) developed the *Relevant Expertise Aggregation* - REA - model, which addresses and solves limits from both IGA and Deliberative Democracy. REA brings "relevant domain-specific expertise into the process of decision-making", without letting the experts become the decision-makers. Ober settled five conditions in order for the REA to work:

1. Decision-makers addressing an issue seek the best available option.
2. The issue is divisible into parts, each of which has a specifiable (relative to other parts) relevance to the issue in question.
3. Each of those relevant parts is explicable as a domain of knowledge that can be enlightened by identifiable domain-experts willing and able to disclose private information.
4. Conditions 1–3 are common knowledge in a group that exists over time such that its decisions take the form of a repeated game.
5. Its members update beliefs about experts in overall reality-tracking ways.

As a conclusion, Table 2.2 presents the comparison from Ober (2013) between Deliberation democracy, IGA and REA.

	Deliberation	IGA	REA
Issue choice	Exogenous or endogenous	Exogenous	Exogenous or endogenous
Options	Exogenous or endogenous	Exogenous	Exogenous or endogenous
Independence	No	Yes	No
Updating	Yes	No	Yes
Choice among options	By reasons offered	By individuals assessing issues	By expertise, or reputations & arguments of domain-experts
Decision by	Consensus or vote	Independent votes	Relevance-weighted votes

Notes: Endogenous = issues or options are determined by the decision-making group. Exogenous = issues and options are determined by external agency.

TABLE 2.2: Three approaches to epistemic-democratic decision-making

## 2.4 What: Adding-values to society

The purpose of this final section is to highlight different outputs of Wisdom of the Crowd — WoC — in society. Each sub-section addresses an issue and its resolution thanks to the development of the WoC.

### 2.4.1 Knowledge sharing

One of the main issue in economics is the uncertainty in information (Akerlof, 1970). Imperfect information has negative impacts on numerous sectors of activities. Hayek (1937) developed a founding theory against the classical economics theories about knowledge, assumed complete and perfect by any agents. During his 1974 Nobel lecture, he concludes on the fact that any economical agent actually lives in a world of imperfect knowledge. Therefore, they cannot anticipate all the possible market outcomes. Nevertheless, the development of Information and Communication Technologies — ICT — has eased the process of information and knowledge sharing, especially within and between companies.

For a part of the literature, organizational information and knowledge are conceived by organization's members themselves (Gherardi and Nicolini, 2000). Learning is the way to assimilate this knowledge, implying modification in behaviors and actions (Weick, 1991; Gherardi, 1999). Others model firms as organizations with a hierarchy of routines. Therefore, knowledge does not rely on members of the organization but rather in the fact that knowledge is tacit and relies on teams of individuals. Those individuals, grouped in teams, share common events, actions, but experience only part of the routines.

Nevertheless, knowledge within an organization can be seen as a social and collective phenomenon. It is based upon the idea of practice, and therefore on community of practice - CoP (cf. Section 2.1.2). As a reminder, a CoP is defined firstly by its members but also, by the ways they share information and knowledge, and how they perform thanks to the group in their

work (Brown and Duguid, 1991; Lave and Wenger, 1991; Wenger, 1998; Zuchermaglio and Talamo, 2003).

Furthermore, as depicted by Brabham, Sanchez, and Bartholomew (2009), there exists a need for aggregating and sharing all this knowledge inside a framework. Simply put, a moderator should be able to promote individual ideas and submit it to other members of the organism. The concept of *knowledge sharing* might be framed as any activity which involves a dissemination or a transfer of knowledge among individuals. These individuals exchange their tacit and explicit knowledge and create new ones. Through a systemic literature review, Charband and Jafari Navimipour (2016) highlight the positive impact of Internet on the dissemination and exchange of information and knowledge.

A classic risk with collaborative framework is the well-known problem of the *free-rider*, which emphasizes the use of resources (public goods for instance, knowledge in our case) without paying or under-paying for the resources (Baumol, 2004). This might imply an under-production, an overuse and a possible degradation of the good (Rittenberg, Tregarthen, and Institute for the Study of Knowledge Management in Education (ISKME), 2009). Awazu and Desouza (2004) offer solutions to counter the "free-rider" problem. Organization could develop reward mechanisms, meritocratic system, with a "horizontal monitoring" (peer pressure) vs "vertical monitoring" (boss-subordinate level) in companies.

### 2.4.2 Open Innovation

The former section presented the opportunity created by a shared knowledge within a company. The circulation of knowledge within a firm allows all its members to make informed decisions, optimal from a company view, and not only optimal from the individual view. This is particularly useful concerning innovation. Indeed, firms which do not innovate die (Chesbrough, Vanhaverbeke, and West, 2006). Chesbrough observes a *paradigm shift* (Kuhn, 2012) in industrial innovation. The old one is labeled as *Closed innovation*.

A successful innovation requires control, which implies the development of a virtuous circle. First, a company increases its investments in R&D. With some probability, this leads it to fundamental technology innovation and then the release of a new product, a new feature. Through the existing business model, these new products or features increase the sales and then the profits, invested in R&D. The circle can go on, innovation being protected by intellectual property, preventing competitors to access to those innovations, which have been developed exclusively inside the firm.

Chesbrough highlighted five leverages which have broken this virtuous circle. First, we observe an increase in mobility from experienced and skilled people. Second, private Venture Capital has grown. They have created firms which only purpose is to sell external innovation. And it worked. Third, we observe an important diminution of the time-to-market of a product, including the diminution of the time a technology gives an advantage to its firm. Fourth observation concerns the customers and the suppliers who become more and more knowledgeable, and challenge the firm to profit from its knowledge silos. Fifth and final leverage relies on the international competition, which, here again, diminishes the time during which an innovation brings an advantage to a company. Altogether, those leverages have created an outside path which broke the virtuous circle: the company which invest in an innovation does not benefit anymore from it and the company benefiting of it usually does not reinvest its profits for a new round of innovation.

In order to answer to this new worldwide situation, Chesbrough has offered a new paradigm: *Open Innovation*. He suggests that firms should look for new ideas inside themselves, of course, but also from the external world, as much from consumers, suppliers, and from competitors as well which could become partners. Open Innovation combines both external and internal ideas. Furthermore, the innovation could bring profits to other markets than the one aimed by the firm. This Open Innovation framework offers three different processes.

First, the *Inside-Out* process offers two solutions for the firm's innovations. The first one is classical: the sale of the innovation, through technological transfers, patents. The second one is to share part of knowledge, or any input (such as data), which could allow, for instance, the development of a technological standard from its own innovation, giving a real advantage to the parent firm which would stay the only one to fully master the standard.

Second, the *Outside-In* process, on the contrary, defines how a firm could benefit from its network. Here again, two possibilities. Either the firm acquires a new technology, or it could participate to a development through a collaboration process (with competitors or with customers).

Third, the *Coupled* process mixes the first two, with a pure collaborative view. Everything is put in common, with co-creation philosophy from all members of the network. This last process is the foundation of the success of Open Source movement.

Chesbrough's Open Innovation theory opens the field to a whole new research. The current work only offers a glimpse to it. Nevertheless, we emphasize the sub-field of communities in open innovation (West and Dedrick, 2001; West and Gallagher, 2004; West and Lakhani, 2008; West and O'mahony, 2008; Bogers and West, 2012; Perkmann and West, 2014; West and Sims, 2018). Joel West focuses his research on the role of (customers) communities in *creating, shaping and disseminating innovations* (West and Lakhani, 2008). For instance, community, which can be spoken about as "a crowd" in the sense of constituted by unknown members, who are working together because of a shared interest, could become an input to a firm's innovation strategy (Von Hippel and Lakhani, 2000), especially through open online challenges, open to all by definition, through collaborative platforms (such Agorize<sup>5</sup>, Innocentive<sup>6</sup>), which allow any firm to set a real problem to the worldwide Web crowd. This is well known as *crowdsourcing*.

---

<sup>5</sup><https://www.agorize.com>

<sup>6</sup><https://www.innocentive.com/>

### 2.4.3 Crowdsourcing

Development of ICT sheds lights on Collective Intelligence (CI) in Human-Computer Interaction (HCI) and Artificial Intelligence (AI). In HCI, the main subject is crowdsourcing. As defined by Howe (2006), crowdsourcing is when a traditional job performed by an employee in a company is outsourced to an undefined, mostly large group of individuals through an open call. The firm could settle different rules to solve its problem. For instance, the best solution, from the firm perspective, is selected and the winner is awarded: it is a directed crowdsourcing. Brabham (2008) adds the notion of production of the solution: crowdsourcing is a business model. Crowdsourcing, under right circumstances, improves the efficiency of problem solving (Afuah and Tucci, 2012), in different fields, such as in the research value chain (Buecheler et al., 2010). Another development of crowdsourcing is the collaborative one: individuals do not compete against each other but share common interests. But the largest the crowd is, the more it may face several issues. First, the question of leadership and decision making: the group, when is self-organized, risks to take more time on debating about its action than really acting.

CI involves others issues such as coordination (Kittur et al., 2007) of the collective action, conflict (Kriplean et al., 2011) among participants, attractiveness of the platform to enhance participation (Beenen et al., 2004). Artificial Intelligence might bring answers to these difficulties. AI enables optimal design of work-flows and management of task instances, in order to facilitate the use of such platform. Focusing on objective questions — unique correct answer — and assuming that a majority of participants will be correct, researchers have developed algorithms predicting the correct answer and even learning and tracking the skills of the participants in order to weigh participants' responses by their abilities (Dai, Mausam, and Weld, 2011).

In organizational theory, collective intelligence is also an area of interest. Woolley et al. (2010) highlighted the fact that the collective intelligence of a

team is a better predictor of its performance than the intelligence of individual members. Thus, for a manager, it is important to influence the group in order to reach the firms' objectives. Based upon the Star Model of organization design (Galbraith, 2014), Woolley et al. (2010) develop the strategy, the structure, the process, the rewards and the people issues a manager should think about in order to enhance CI in her team.

Development of Internet facilitated the creation of peer production: an online group performs a form of open creation, setting and aiming goals in a decentralized way, members of the group having a wide range of motivations. Moreover, peer-production projects are usually developed outside the classical boundaries of a firm, which is constrained by contract and property rights (Benkler, 2016). The most famous peer-production projects are usually in computer science field, also known as Free/Libre and Open Source Software (FLOSS), such as Linux and the Wikipedia encyclopedia. Researchers focus on organization, motivation and quality of peer-production projects. For the successful ones, researchers observe a performance as well as traditional companies, but without organizational characteristics. They highlight an apparent absence of formal hierarchies and informal norms, such as fairness and reciprocity, rule peer-production projects. Moreover, researchers wonder what types of motivations could incite individuals to be part of such projects. We have already mentioned the extrinsic and intrinsic possible motivations in Section 2.2. Recent researches tend to highlight an interdependence of motivations, leading platform designer to take into account how a particular design trigger individual motivations. Another important question is the quality of the production. For Raymond (2001), open source projects ensure "better quality, higher reliability, more flexibility, lower cost, and an end to predatory vendor lock-in", also formalized by the Linus' Law: "with enough eyeballs, all bugs are shallow".

## 2.5 Conclusion

This literature review brings details on the four genes of collective intelligence actions, as defined by Malone, Laubacher, and Dellarocas (2009): who do the action (Section 2.1), why it is done (Section 2.2), how it is done (Section 2.3) and what is the results of the action (Section 2.4). On the one hand, we have presented major contributions concerning the why and the what. On the other, we focus our research on the who gene, following an historical evolution of the crowds; and on the how gene, presenting numerous researches which try to bring evidences on the processes behind the apparition of a collective intelligence, a wisdom of a crowd, with an important focus on the link between this (not so) new concept of wisdom of the crowd and our democratic process. The following chapters present our study cases (Chapter 3) and why studying an online discussion forum help us to understand the wisdom of the crowd within a democracy-like framework. Chapter 4 focuses on an empirical research about the who gene. Finally, we present the major contribution of our work in Chapter 5, which focuses on the how gene, with an empirical work which highlights the existence of an original concept: the Consent of the Crowd.



## Chapter 3: Reddit - Change My View

3.1	Who: Usecase	50
3.1.1	Reddit	51
3.1.2	Change My View	51
3.2	Why: The will to change	52
3.3	How: Graph models	55
3.4	What: CMV descriptive analysis	59
3.4.1	Life-cycle of Online Communities	61
3.4.2	Forum dynamics	64
3.4.3	Threads characteristics	66

The Internet allows a group of individuals, which could be labelled as a crowd, to gather *somewhere* in order to exchange information, with a wider acceptance of the word. Researchers studied the behavior of a crowd long before the development of the Internet, especially in the political sphere. Oldenburg (1999)'s concept of *third place* defines a public space, beyond the house and the workplace where a group of individuals meets and where people can interact on an informal basis. Wright (2012) extends the concept taking into account the technology of the Internet. What if those blogs, electronic bulletin boards or other online fora are the *third space* (Wright, 2012) where people interact on an informal basis? The current work analyzes how an open online community acts in a particular third space: a debate opinion forum. This chapter presents the case study used to investigate how individuals debate on an open online platform. First section introduces the "who": Reddit - Change My View. Second section explicites the reasons why the study of this particular case helps to better understand the research questions. Third section develops the methodology, the framework — Social Network Analysis — used to conduct the research. We conclude the chapter by presenting the descriptive analysis of our database.

### 3.1 Who: Usecase

Collective discussions are not an emanation of technological evolution. Indeed, Agora in Ancient Greece performed the exact same action: gathering individuals in one place in order to exchange ideas, converse, debate and even produce collective decisions. Social networks have taken an eminent place as debating spaces in political debates. Investigations on US Presidential elections and Brexit vote, both occurring in 2016, highlight the power hold by social networks (Allcott and Gentzkow, 2017; Bossetta, 2018; Del Vicario et al., 2017; Khatua and Khatua, 2016). Nevertheless, social networks such Facebook, LinkedIn or Twitter, are not the only online debating spaces. Indeed, even after the expansion of social networks, still 15% of Internet users

exchange actively on online fora in the United States (Duggan et al., 2015). We analyze this specific kind of debating space.

### 3.1.1 Reddit

Reddit<sup>1</sup>, created in 2005, is a social news aggregation, allowing users to rate web content, and discuss about it. In February 2019, it was the sixth most popular website in the United States and the seventeenth in the world<sup>2</sup>, with 330 million unique users<sup>3</sup>. Their headline - "the front page of the Internet" - highlights the ambition of the platform. Registered users share content (links, text posts or images). Then, members of the community can upvote (assign a +1 vote) or downvote (assign a -1 vote) to any content, and discussion can arise about those contents. Exchanges cover diverse topics, from news, music, fitness to movies, books, video games. Reddit is divided in subspaces called *subreddits*. Those spaces are self-created by communities of users, interested in a common topic. Then, every post is linked to a subreddit. Each subreddit, and Reddit as well, prompts the best posts — posts with the highest *karma*, score after removing the downvotes to the upvotes — on the "top page". Contrary to social networks, where users follow a person (at least, an account), on Reddit, users follow contents. They can *subscribe* to subreddits, but not to other members. The following section focuses on a particular subreddit: *Change My View*.

### 3.1.2 Change My View

*Change My View* - CMV<sup>4</sup> (Pardes, 2019) is a subreddit, as define in Section 3.1.1. It admits discussions and opinions on any subject. Such as *Agora* in Ancient Greece, the initiator — Original Poster (OP) — opens a discussion announcing his idea on a topic. CMV has developed constrained rules to join the debate. To open a topic, one agrees to follow the conversation, at least three

---

<sup>1</sup><https://www.reddit.com/>

<sup>2</sup><https://www.alexa.com/siteinfo/reddit.com>

<sup>3</sup><https://www.socialmediatoday.com/news/reddit-now-has-as-many-users-as-twitter-and-far-higher-engagement-rates/521789/>

<sup>4</sup><https://www.reddit.com/r/changemyview/>

hours after its beginning and one *must personally hold the view and be willing to have it changed*<sup>5</sup>. Alongside with the vote system, CMV created a new reward system. All participants, OP included, can award with a delta  $\Delta$  any argument that allowed him to make a step in the change of his view. The reward attribution must be argued, explaining how the comment has been convincing. The opinion change may be minor as the  $\Delta$  reward does not have to terminate the thread. We thus observe some threads where several  $\Delta$  were assigned. The DeltaBot, a *Reddit bot*, confirms the  $\Delta$  attribution through an automated post.

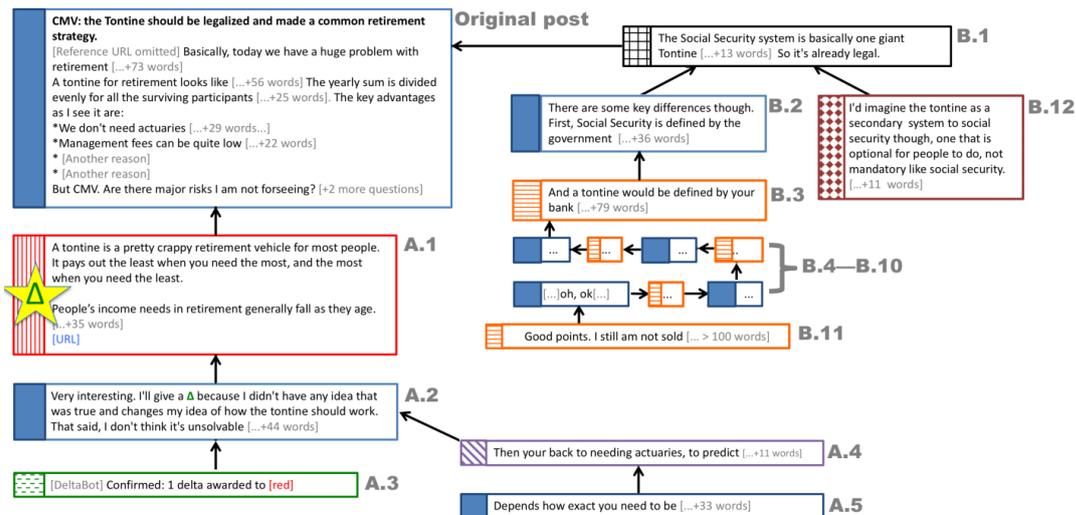
Figure 3.1 presents a thread example in CMV<sup>6</sup>. The boxes with solid blue strip indicate the author who is the OP in this example. This author opens a thread about an investment plan called "tontine". He develops his point of view on the subject. He finishes his post by asking the forum members to change his view (CMV). A first challenger answers (post A.1). Then the OP answers back (post A.2), attributing a  $\Delta$  to the post A.1 and providing explanations about his decision. The DeltaBot confirms the delta award in the post A.3. Similarly, another challenger answers the OP (post B.1), which the OP answers to (post B.2) before the intervention of another orange-horizontal-strip-lines challenger (post B.3). This is followed by a discussion between the OP and this latest challenger (between post B.2 and post B.11). In this example, two other challengers participate (posts A.4 and B.12) with or without an answer from the OP or another challenger.

## 3.2 Why: The will to change

Chapter 2 highlights the shifting from a third place to a third space (Wright, 2012), from a physical to an online forum for citizens to debate. Thus, to improve scholars' comprehension on a theoretical level of collective thinking, it is important to understand the behaviors of people in online fora. It also has practical implications: a better understanding of such behaviors might

<sup>5</sup><http://www.reddit.com/r/changemyview/wiki/rules>

<sup>6</sup>[https://www.reddit.com/r/changemyview/comments/3mzc6u/cm\\_the\\_tontine\\_should\\_be\\_legalized\\_and\\_made\\_a/?sort=top&st=jf2di5d1&sh=27df2763](https://www.reddit.com/r/changemyview/comments/3mzc6u/cm_the_tontine_should_be_legalized_and_made_a/?sort=top&st=jf2di5d1&sh=27df2763)



Note: Colors and shapes in posts indicate distinct users. OP awarded post A.1 with a Δ.

FIGURE 3.1: Real CMV thread, example from Tan et al. (2016)

improve the user experience, the user engagement and even the democratic process of our societies (Medvedev, Lambiotte, and Delvenne, 2018).

How does democracy evolve with the development of Information and Communication Technologies — ICT? Arabic revolutions during the past years embody positive examples on the use of ICT by citizens to claim for more democracy (Aman and Jayroe, 2013; Shirazi, 2008; Shirazi, Ngwenyama, and Morawczynski, 2010), even if these cases might involve armed fights as well. The use of ICT has also been deployed by governments in order to foster citizens engagement, such as to express their ideas, comments on law writing. For the first time in France, in 2015, the government presented a law proposal for a "République Numérique"<sup>7</sup> (the Digital Republic bill<sup>8</sup>). It has given access, online, to the text on a website allowing French citizens to discuss, comment and propose new part of the law. In total, 21,330 participants voted almost 150,000 times and sent 8,500 arguments. Numerous researches link the need for governments to increase a citizen engagement in political subjects and the development of *eDemocracy* and *eParticipation* (Bartlett and Grabbe, 2015; Chadwick, 2003; Dahlberg, 2001; Macintosh, 2008). Others

<sup>7</sup><https://www.republique-numerique.fr/>

<sup>8</sup><https://www.republique-numerique.fr/pages/in-english>

highlight the failure of the Internet to increase the democracy in our societies (Smith, 2008; Grönlund, 2009; Grönlund and Åström, 2009).

Reddit - Change My View (CMV) possesses several characteristics making this third space (Wright, 2012) a proper candidate to study online deliberative behaviors. One difficulty about common space is defined by Chambers (2009):

*Chatting with a neighbor about the deplorable state of the local playground is not the same thing as participating in a citizens' initiative to decide what to do to clean up the playground. One is conversation, the other is deliberation. Central to theories of democratic deliberation is the idea that deliberation precedes and can be directly connected to a course of action collectively decided upon.*

On CMV, the rules constraint an individual to develop a reasoning to support his view. Moreover, the  $\Delta$  is validated by the DeltaBot under the condition that the individual willing to reward a post has argued and detailed how the post has been convincing from his perspective. This is in favor of a *deliberation* framework over the *conversation* one describes by Chambers (2009).

Besides, as mentioned in Schwartzberg (2015), there is a lack of empirical evidence on democracy process. Reddit has gained a central place in the scientific literature thanks to the openness, richness and quality of its data, which allows one to perform longitudinal studies of the whole system and, critically, to ensure reproducible results. Being a part of the forum, CMV incorporates the same qualities.

But the most important argument in favor of CMV is its core characteristics. As many others, individuals debate, exchange ideas, try to convince each other. But the originality of the CMV relies on the person who has opened the debate, and more precisely speaking, on his motivation. The Original Poster opens a thread *in order to be convinced* he is wrong. This motivation relies directly in the intrinsic category, which is the most powerful to stimulate individual participation (Kraut et al., 2011).

Nevertheless, to the best of our knowledge, CMV has been the subject of very few research<sup>9</sup>. Being alike others debate fora (such as Slashdot, Barrapunto, Meneame and even as Wikipedia concerning debates on edit pages) and with one characteristic (the intrinsic motivation from OP to be convinced) distinguishing it from others, makes Reddit — Change My View both an original and interesting case to study.

### 3.3 How: Graph models

Networks is used in a variety of research fields in order to better understand real world phenomena, modeling such phenomenon through a graphical visualization. From analyzing the structure of the brain to have a more precise conceptualization of electronic transmissions in the human brain (Hecht-Nielsen, 1988), to the spreading of a computer virus (Kephart and White, 1992; Kumar and Mahdian, 2010), birds migrations (Shimazaki et al., 2004), or even political network that underlay the birth of Renaissance era in Florence (Padgett and Ansell, 1993)<sup>10</sup>, networks are used to model the behaviors of individuals (neuron, computer, bird, human) in a study. Indeed, a network is a conceptual object where *vertices* (dots) are connected to each other thank to *edges* (links).

The Internet is the contraction for **interconnected network**. It is a global system allowing to link computers all around the world. Thus, using network theory, a part of the well-known graph theory in computer science, to model human behaviors is one of main approach in computational social science (Lazer et al., 2009). This new field of research emerged with the development of Internet. Contrary to self-declarative survey, each action we do with our smartphone, tablet or computer leaves a digital trace, which can be aggregated and analyzed.

Similarly, every discussion on Reddit — CMV is saved on the Reddit servers, allowing researchers to study on a micro-level individual actions

---

<sup>9</sup>CMV website makes an incomplete inventory of the scientific articles: <https://www.reddit.com/r/changemyview/wiki/research>

<sup>10</sup>Cf Figure 4.9.

and interactions. Thus, how to analyze the discussion threads from CMV? The typical approach consists in modeling fora as graphs. A thread is an aggregate of messages linked together. Thus, a discussion takes the form of a reverse *tree* where vertices model the messages and are linked if one answers another directly (cf. Figure 4.12a).

An important part of the literature reports attempts to model the structural evolution of discussions. From the growth of general networks, the random evolution (Erdős and Rényi, 1960) and the small world (Watts and Strogatz, 1998) to the preferential attachment (Barabási and Albert, 1999), researchers demonstrate the relevance of certain patterns in online discussions evolution (Kaltenbrunner et al., 2007). The preferential attachment model can be summarized by the "rich gets richer" slogan. Gómez, Kappen, and Kaltenbrunner (2011) proposed a variant, taking into account the importance in a discussion of the opening message which attracts more answers.

Nevertheless, the above-mentioned studies do not take into account additional information gathered from the fora, such as up- or down-votes for an answer, if the conversation has ended, a level of authors' experience or even the time which passed between consecutive answers. Several studies explore how these variables might explain the discussion evolution. Anderson et al. (2012) highlight results about the first reply to the opening question on the forum StackOverflow<sup>11</sup>, one of the main problem-solving forum in computer sciences. The most experienced users reply before other participants, in terms of order of answers and in terms of speed of posting them, whereas the reputation from the opener does not impact the experience of participants. Those results might be explained by the framework of the forum. Indeed, in StackOverflow, one gets more points if his answer is chosen by the OP as the best answer. Thus, the forum creates a direct incentive to answer quickly. Other studies draw similar conclusions from editing possibilities on StackOverflow. They emphasized the fact that the first best answer is more edited than the second-best, the quality of the answer being a

---

<sup>11</sup><https://stackoverflow.com/>

difference of total up- and down-votes (Dalle, Devillers, and Besten, 2014; Weninger, Zhu, and Han, 2013).

Qualitative approaches have been used as well. Yla R. Tausczik (2014) has defined a taxonomy of "collaborative acts" which describes the collaboration processes involved in MathOverflow<sup>12</sup>, supported by semi-structured interviews to better understand the impact of collaboration, the core feature of such fora. Companies are also interested in those new tools. It is common now to complete after-sales services with a Q&A forum, where customers can both ask questions and answer them, avoiding the repetition of the same topic as the selected solution becomes shared to the whole customer community. Velkovska (2015) studied the evolution of a "brand community" for the department of the customer service of a cell phone company. Bail (2016) studied how advocacy organizations use social networks or fora to shine a light on their cause.

Those researchers improved the comprehension of the discussion evolution and social interactions. There are, however, still difficulties when anticipating this evolution. The anticipation may also consist in predicting the next behavior of a participant. Studies on this subject focus on social roles in online fora. Lumbreras (2016) affirms: *Roles are both descriptive and predictive categories of behavior*. Focusing on conversational behaviors and defining two authors having the same role if they tend to interact in the same type of discussion, he highlights several typical roles. This is allowed by the analysis of the motifs, discussed in the next paragraph.

In addition to studying the variables of the structure or the social variables (experience, for instance), researchers focus on subgraph (mainly on dyads and triads), formalized by Wasserman and Faust (1994). The methodology consists in counting occurrences of a certain subgraph of size  $k$  and make a comparison with a reference model, often built by randomizing the real graphs, preserving some of their properties. Park and Barabási (2007) focus on dyads and develop a model to understand whether a certain property of nodes (characteristic of proteins, for instance) explains the structure

---

<sup>12</sup><https://mathoverflow.net/>

of the graph. Holland and Leinhardt (1970) characterize the sixteen possible triads (cf. Figure 4.3). Milo et al. (2002) defined as *network motifs* subgraphs occurring significantly more than those in randomized networks.

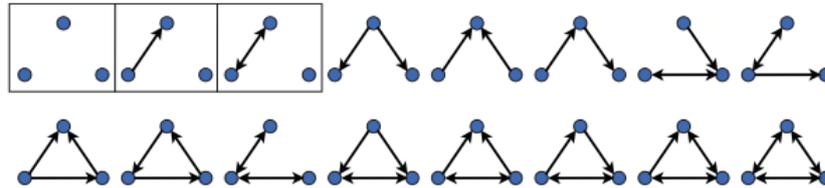


FIGURE 3.2: Triad dictionary from Cunningham et al. (2013)

Motif analysis allows the improvement of network analysis. It leads to the diminution of computational needs, predictions on selection of motifs being not degraded compared to ones made with the whole network (Cunningham et al., 2013). It also highlights specific characteristics of different types of threads (Adamic et al., 2008), in different communities (Conaldi and Lomi, 2013; Lomi et al., 2014).

Furthermore, a discussion can be modeled by a tree, as said above, but we can also build an author network, where nodes are the authors and they are linked if one answers another. Dorat et al. (2007) model a mailing list with the random multi-level approach. Then, either one posts an answer to the OP opening message, either one comments a post. An answer is a suggested solution to the OP problem, whereas comments group requests for precision, correction in case of error. On a structural basis, if we model the discussion through a tree graph, we observe three levels. The first level is composed by a unique vertex: the root. The second level groups the answers *and* comments to the root. Finally, the third level models the comments to answers. One case is not taken into account by the forum design: a comment on a comment. For instance, the OP asks a question, one *comments* it. Then the OP is encouraged to edit his post accordingly to the comment, assuming the comment is correct. But the OP might want to thank the individual, posting an acknowledgement to the comment. From the modeling point of view, analyzing the discussion solely through the tree network, model will connect the *comment* from the OP directly to the root. Thus, the thematic coherence of

the discussion will not hold. As an answer, Web developers often admit the possibility to cite a member writing down his pseudonym in a post, which is the case in Facebook, Twitter, and StackOverflow. Researchers could then, correct the modeling of the discussions adding rules on the edges creations. Nevertheless, the robustness of the modeling relies on the proper comprehension from the users of the design conception. But a common user will just use the service, and not try to understand how it should be used. Numerous researches develop a framework to engage designers to understand the user experience (Forlizzi and Battarbee, 2004).

Reddit in general, and CMV in particular, solves this issue allowing a member to comment a comment at any level. For instance, the discussion modeled in Figure 4.12a is composed by ten levels of answers. On this tree, a node (a post) is properly the child (*i.e.* the answer to) of the connected node above it, with the top node being the root (the opening post of the discussion). Conversely, if a node is linked to nodes under it, the former is the parent to the latest. Thus, the design of Reddit is more robust to hold the thematic coherence of the discussion than fora cited before. A member of the discussion might misuse the forum options, but we assume that the margin for error is lower than in other types of fora. This design is an additional characteristic in favor of the use of Reddit — CMV as a case study.

### 3.4 What: CMV descriptive analysis

Section 3.1.2 presents the operating of the subreddit *Change My View*. The entirety of the empirical study of this PhD thesis is based on a database coming from this subreddit, extracted by Tan et al. (2016)<sup>13</sup>. The database covers threads from January 2013, date of the subreddit creation, to November 2016.

The whole computational work has been realized within Python 3.5.3. Once the dataset uploaded in the software, we have removed a few errors, inherent to any scraping works. The database includes information about

---

<sup>13</sup>An update version of the database, used in the current PhD thesis, can be found here: <https://chenhaot.com/papers/changemyview.html>

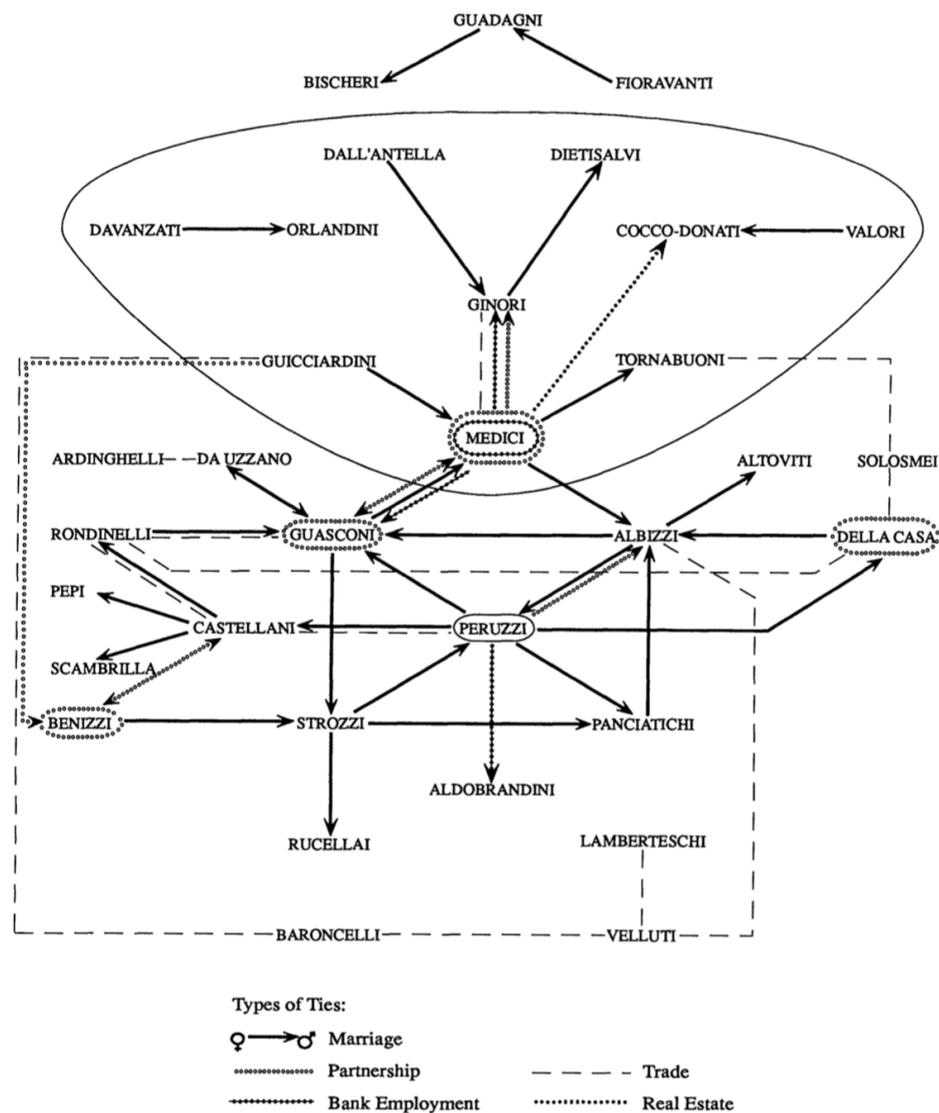


FIGURE 3.3: Marriage and economic network (Padgett and Ansell, 1993)

21,564 discussion threads, 1,441,914 posts, 14,287 OPs and 71,775 unique authors, who did not open a thread, producing in average 16.75 posts each. The penultimate statistic is based on the pseudonym of the participants. For different reasons, an account could be deleted. If so, the pseudonym was replaced by "[deleted]". 156,286 posts - 10.84% - have been written under this pseudonym. We do not observe "deleted" account for the OPs.

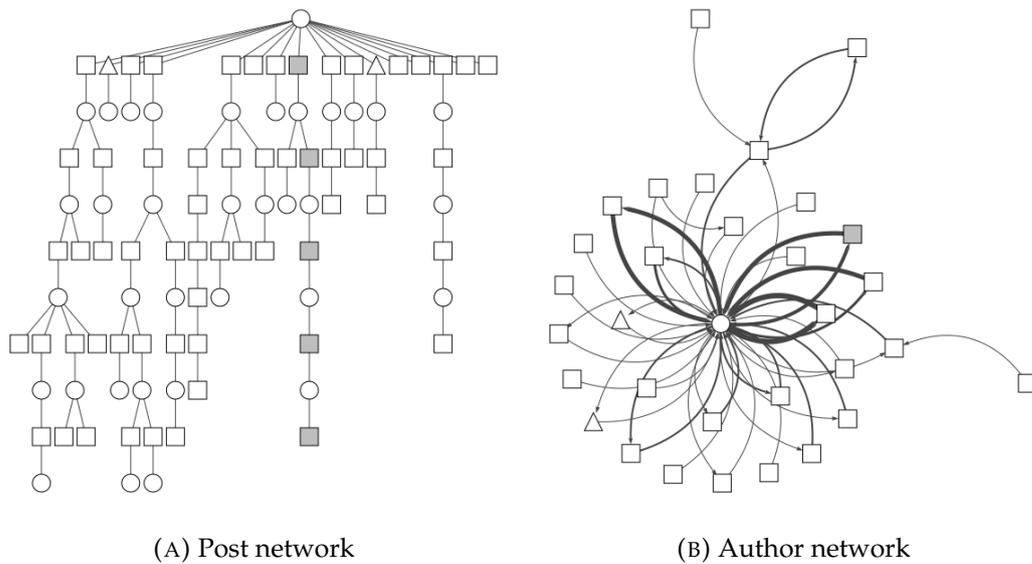


FIGURE 3.4: Network modeling from a CMV thread

### 3.4.1 Life-cycle of Online Communities

The following part explores the activity of the forum on the studied time span (January 2013 — November 2016). We will analyze the life-cycle of CMV, which is a significant research field. Researchers understand the importance of catching at which step of the process a community is. In order to sustain motivation of participants to contribute, fuel for a healthy community (cf. Section 2.2), the management has to develop features according to specific needs, which are different at each step. Authors pinned down different stages of the life-cycle. For Wenger, McDermott, and Snyder (2002), the process of community of practice (CoP) creation is covered by the following steps *potential*, *coalescing*, *maturing*, *stewardship*, and *transformation* (cf. Figure 3.5).

When people begin to network and exchange about a common interest topic, the group enters in the *potential* step. The main problem at this stage is to find enough common ground among participants. If so, they will feel connected to each other and will understand the interest of sharing insights, information, ... Once a community understands what the shared motivation

The jagged line represents the level of energy and visibility that the community typically generates over time.

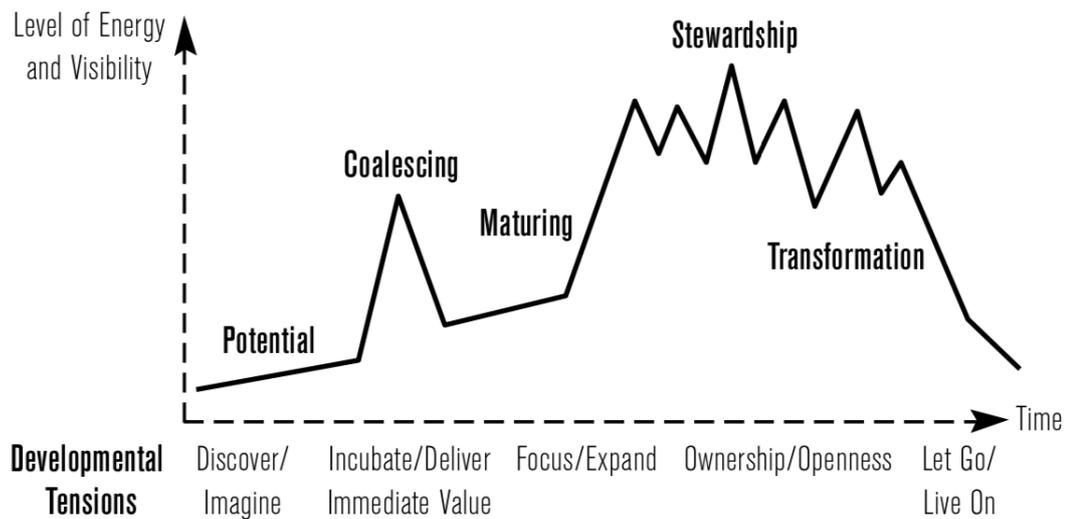


FIGURE 3.5: Stage community development (Wenger, McDermott, and Snyder, 2002)

among its members is and where the community wants to go, then the *coalescing* phase begins. Then, the community is officially created. The main issue at this stage is to generate enough activities to allow members to build relationships. Then, the community enters the *maturing* phase. Focusing on CoP evolution, the challenge during this step is twofold. First, a growth in membership. Second, as a CoP objective is to create an efficient knowledge sharing process, there is a stake to go from tip exchanges to proper knowledge sharing framework. Then, the fourth step begins the *stewardship*. The stake here is to maintain sufficient attractiveness to the members. Indeed, having grown up, the quantity of energy required to manage this new entity is more important. Managers can be moved to other projects. Then, how fewer active members can maintain the community? Usually, they can not until a new "leader" takes the management part, mentoring new members, managing the community's sense of ownership over the community's production, and the tension between those two sides. Finally, the last step is the *transformation*. It sometimes happens that an important event occurs (tremendous number of newcomers, complete loss of activity, ...). Then, the community needs a radical *transformation* or arrives to an end.

Similarly, Malhotra, Gosain, and Hars (1997) defined *inception, beginning of user activity, interactivity, growth and experimentation activities* as the four phases of community evolution. Furthermore, Andrews (2001) offered only three phases: starting, encouraging online interaction, and moving to a self-sustained interactive environment. We will not detail each step. Still, those models present three main steps: birth, growth and death. Indeed, as underlined by Iriberry and Leroy (2009) along with Ahituv and Neumann (1986), those models are linear. Thus, they offer through their own model, specifically applied to online communities, a more iterative process: *inception, creation, growth, maturity, and death* (cf. Figure 3.6).

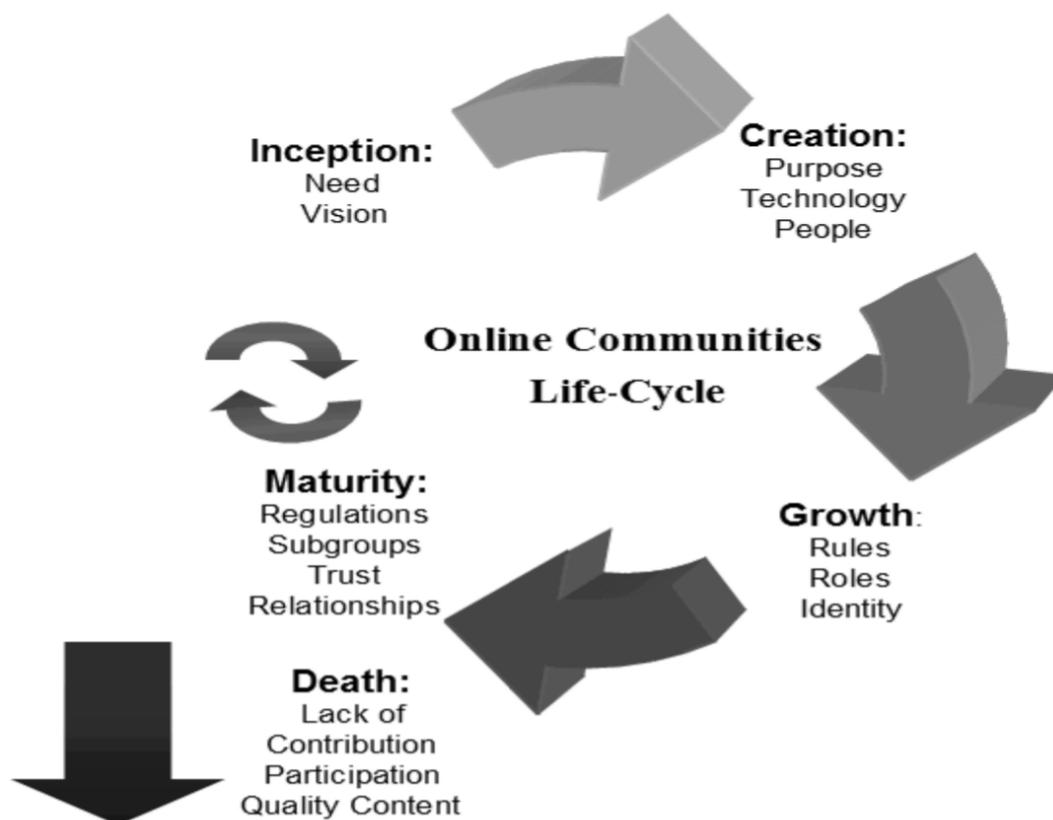


FIGURE 3.6: Stage community development (Iriberry and Leroy, 2009)

Similarly to Wenger, McDermott, and Snyder (2002) *potential step, inception* occurs when a group of people needs information, relationships or support. Specific to Online Communities, the *creation* phase happens once the goal of the community is clear for each member, and when the technological

part is settled: either forum, emailing list or chat, in respect with the needs and preferences of the initiator members. The creation starts when interactions and calls for memberships begin (Malhotra, Gosain, and Hars, 1997). Members of the OC begin to develop an identity and a culture, they select roles to play (from leaders to lurkers), a set of rules arises. Those evolution opens the *growth* phase. Then appears a need for a more formal organization: the OC moves to the *maturity* phase. New features are developed such as the reward system, subgroups. Newcomers arrive and old members leave. The community has to adapt to the evolution of its membership. Either it iterates to a new cycle, in order to integrate newcomers, or the OC does not survive and *dies* due to a lack of quantitative or qualitative participation (Jarvenpaa, Knoll, and Leidner, 1998), for instance.

### 3.4.2 Forum dynamics

Understanding at which step of the life circle an online community is, allows managers and researchers to answer the proper needs of its members. The following analysis spots which steps the CMV community has gone through. Following Iriberry and Leroy (2009) model, the analysis begins once the technological tools have been chosen by the community. Therefore, at the beginning of the time span, the community is at the *creation* step. Figure 3.7 presents different types of monthly activities: threads opening (and percent of threads having at least one  $\Delta$ ) in Figure 3.7a, number of participants in Figure 3.7b, age of participants in Figure 3.7e, posts creation (and corresponding percent of "[deleted]" author posts) in Figure 3.7c, and delta  $\Delta$  awards in Figure 3.7d.

The forum knows its most productive period over the first six months of existence, with an important increase each month. The number of new threads created is multiply by two every month to reach more than 850 new threads in May 2013. We observe this important increase as well for the number of participants (7,500 in August), posts creation (50,000 in August) and

the amount of  $\Delta$  attributed (400 in May). Those six months cover both *creation* and *growth* phases.

The 2013 summer opened a new period of activity: *maturity* cycles. The second semester of 2013 observes a drop in activities. The number of new threads decreases and stabilizes to 550 new threads per month over the last quarter. Similarly, the number of active participants drops to 5,700, producing approximately 35,000 posts, and awarding around 240  $\Delta$ , per month over the end of the year. Moreover, from September onwards the number of newcomers, defined as such during their first month of activity, becomes lower than the one of elders, defined as such if they have been active in past months.

The new year 2014 is characterized by a renewal of activity, holding over the first semester (with a drop during the month of February and another just before summertime). The community grows to reach 6,500 users, producing more than 42,000 posts and more than 250  $\Delta$  (in average) per month over the first six months of 2014. Similarly to 2013, second semester is characterized by a drop in activities. Number of participants, posts and  $\Delta$  awarding reached 4,700, 23,000 and 200, respectively, in January 2015. Besides, the number of opening threads deeply decreases through the whole year to reach 300 new threads in January 2015.

We observe a similar trend in 2015 and 2016: an increase of activity during the first semester, and a decrease after the summer, which is in favor of an iteration of the *maturity* phase (Iriberry and Leroy, 2009). Nevertheless, the global threads opening stays stable around 400 per month, 5,500 participants per month, 30,000 posts per month. The activity does not reach back its initial level, except for the number of  $\Delta$  awarded, which is developed below. Drop of November 2016 is due to the collection date, on the 10<sup>th</sup>.

Furthermore, from the opening to the end 2015, the average slowly but constantly increases, which stabilized after 15 months, even if the variance increases as well. The participants stay longer as the forum is getting older. Forum developers succeeded in keeping the attention of participants, but have difficulty to attract more and more newcomers (cf. Figure 3.7b).

Percentage of threads with  $\Delta$ , "[Deleted]" author posts and number of  $\Delta$  awarding add additional information to the analysis. First, we observe a continuous decrease of "[Deleted]" author posts percent (near to 15% in 2013, 11% in January 2015, 7% one year later, dropping to less than 4% in October 2016). We assume that the number of people leaving the community decreases alongside. Assumption strengthened by a global stabilization of the *older* participants, evolving between 3,000 and 4,000 per month from January 2014, following semester trend described above. Moreover, the number of newcomers decreases slowly from mid-2013 to stabilize from January 2015, evolving between 1,000 and 2,000 newcomers per month, following semester trend described as well. This is in favor of the regulation of the community membership. Second, the number of  $\Delta$  awarded is the only metrics increasing from summer 2016. With the decrease of the number of new threads, the percent of threads observing at least one  $\Delta$  reaches 57% in October 2016, against 40% in January 2015 and 25% one year before. This is a demonstration of a better use of the rewarding system. Both arguments are in favor of the *maturity* step in the life-cycle model. Finally, as suggested by Iriberry and Leroy (2009), a rewarding system makes the community members more active, posting more messages. Indeed, even if certain stability in posts activity comes along the increase of  $\Delta$  activity, the continuous decrease of "[Deleted]" author posts tends to confirm the assumption.

### 3.4.3 Threads characteristics

As detailed in Section 3.2, Reddit in general and CMV in particular, is an interesting case study. Among other reasons, this forum offers an open, rich and quality database. From the export structure, we are able to create proper network to model the threads. The following section presents descriptive statistics, but at a thread level. We study different types of variables. First, *network characteristics*:

- #Post — number of posts;

- #Author — number of unique author;
- #Root Answers — number of direct answers received by the OP. It is a proxy for the number of challengers to get a reward;
- #Leaf — number of posts which did not receive any answer. It is a proxy for the number of sub-discussion within the thread;
- Width — maximum number of posts being at the same level of height of the tree<sup>14</sup>;
- Height — maximum level of height.

Second, we calculate the following network measures:

- Modularity, defined as  $p = \frac{2 \times \#Edge}{\#Nodes \times (\#Nodes - 1)}$ , indicates the cohesiveness among potential communities<sup>15</sup> within a graph. The higher the modularity is, the denser communities will be.
- Average Sackin index, defined as  $S_{av} = \frac{S}{\#Leaf}$ . Sackin index, defined as  $S = \sum_i dist_{root}(Leaf_i)$ , is the sum of the leaf distance from root. It summarizes the shape of a tree (Sackin, 1972).

Finally, we have at our disposal *complementary information*:

- Age (Month): as defined in Section 3.4.2, we express the age of a participant with the number of months since his first post. The *average age* is the average age of authors within a thread;
- Experience: we have the number of  $\Delta$  received by an author from its *birth*. We use this variable as a proxy for the experience of the author in debating. Here again, we present below the statistics about the average experience within threads;

<sup>14</sup>A network distance is the length of the shortest path between two vertices of the network. For instance, the distance between a direct answer to the root and the root is 1. Furthermore, due to the "tree" modeling, we consider the *level of the height* of the tree. Then, the level of height of a post is equal to its distance to the root.

<sup>15</sup>Several algorithms allow one to detect for *communities*, sub-graphs where the connections among members are denser than between sub-graphs.

- OP age and experience: we present a focus on OP age and experience;
- Duration (Hour): how long last a discussion? We present the duration in hours.

Table 3.1 presents the descriptive statistics of those variables. An average discussion attracts 25 participants, producing 67 posts. 13 posts are a direct answer to the root, and we count 26 sub-discussions (cf. Leaf statistics). Thus, each direct answer implies 2 sub-discussions. Its corresponding tree is 15 large, reaching it solely at the second level, whereas it has a height of 8. The modularity is really low, indicating that there is no communities in those networks, in average. Authors are 8.5 months old, OP is younger with an age of 4.5 months. Authors are experienced, being rewarded with 11  $\Delta$ . The OP is much less experienced, with almost 3  $\Delta$ . Furthermore, among threads where  $\Delta$  being attributed, almost 2 were rewarded and at least one by the OP. Finally, it lasts more than 10 days.

*Root Answers*, *Leaf* and *Width* measure at different steps of the discussion the number of attempts to convince the OP. *Root Answers* is the rough number of challengers. We could assume that if there is too much of those possibilities, OP will explore only some of them. If the number of leaves is higher than the number of root answers, some tracks have attracted other participants, creating new sub-discussions. But not all those attempts attract the same level of debate (*i.e.* the same amount of exchange). This is confirmed by the Average Sackin Index. Whereas in average, a thread has a height of 8, the average Sackin index is equal to 3.51. This indicates that some sub-discussions are less dense than others. Finally, *Width* indicates the number of attempts which have reached the same steps of development, *Level Width* precising which one. Being between levels 1 and 2, and being close to the average number of root answers, we could assume that *Width* and *Root Answers* reflect the same information.

We presented above the picture of the average discussion. Nevertheless, analyzing standard deviations brings important information. Indeed, the majority of those standard deviations are really high. Taking into account

the difference between the quartiles, the minimum and maximum, we affirm that the trees are highly heterogeneous, in terms of characteristics. The picture of the maximum values is really different. The most important thread<sup>16</sup> attracted 774 participants, who have produced 1,776 posts, with 18  $\Delta$  attributed, reaching a width of 263 and a height of 10, lasting over a year. This is confirmed by the study of histograms and density of each variable (cf. Figure 3.8). Indeed, the density of *#Post*, *#Author*, *#Root Answers*, *#Leaf Width*, *OP Age (Month)*, *Av. Experience*, *OP Experience* and *Duration* are close to a power-law one. *Width Level* and *Height* densities are constrained by the number of values taken by those variables: from 1 to 10 maximum. A similar effect occurs for *#Delta* and *#Delta awarded by OP*. Finally, *Av. Sackin Index* and *Av. Age Author (Month)* density are closer to Gaussian density function.

The *power law* effect occurring on the majority of the variables is a classical characteristic for online network studies. It refers to the *scale-free* framework (Barabási and Albert, 1999) and to the preferential attachment model (Gómez, Kappen, and Kaltenbrunner, 2011): at some point, the more a discussion attracts participants, the more authors will participate to it.

Finally, we analyzed links between variables. We apply the Pearson test to every combination among the set of variables. Table 3.2 presents the coefficients of correlation and the result of the test. First of all, all the coefficients are highly significant for *OP Age* and *#Root Answers*. Similarly, the variable *OP Experience* is significantly correlated only with *Av. Experience*. To make the reading easier, we produced the correlation heatmap on Figure 3.9.

Firstly, we observe a significant highly positive correlation between variables characterizing the form of the tree: *#Post*, *#Author*, *#Root Answers*, *#Leaf Width*. The denser the discussion is, the more it involves participants, and so challengers answering the OP, opening more sub-discussions, feeding the discussion which becomes wider. *Height*, and *Level Width* to a lesser extent, are positively, but weakly, correlated to previous variables. Figure 3.10 presents the distribution of *Width* and *Height* per thread. On the one hand,

<sup>16</sup>Assuming that the maximum values are describing the same thread, which is not the case.

	count	mean	std	min	25%	50%	75%	max
#Post	21564	66.87	88.65	1.0	22.00	41.00	74.00	1776.00
#Author	21564	25.34	31.86	1.0	10.00	16.00	27.00	774.00
#Root Answers	21564	12.56	11.83	0.0	6.00	10.00	15.00	216.00
#Leaf	21564	26.45	37.00	1.0	9.00	16.00	28.00	874.00
Width	21564	14.82	14.70	1.0	7.00	11.00	17.00	263.00
Level Width	21564	1.67	1.14	0.0	1.00	1.00	2.00	10.00
Height	21564	7.60	2.74	0.0	5.00	9.00	10.00	10.00
Modularity	21436	0.08	0.11	0.0	0.03	0.05	0.09	1.00
Av. Sackin	21564	3.51	1.35	0.0	2.55	3.43	4.41	10.00
Av. Age Author (Month)	21564	8.53	5.90	0.0	3.73	7.75	12.38	45.00
OP age (Month)	21564	4.48	7.88	0.0	0.00	0.00	6.00	45.00
Av. Experience	21487	11.24	12.64	0.0	3.93	7.69	14.02	186.00
OP experience	21454	2.67	17.48	0.0	0.00	0.00	0.00	257.00
#Delta*	7650	1.61	1.12	1.0	1.0	1.0	2.0	18.0
#Delta from OP*	7650	1.30	1.15	0.0	1.00	1.00	2.00	18.00
Duration (Hour)	21564	255.04	688.72	0.0	24.78	47.83	105.04	9578.17

Note: \* - count on threads with at least one  $\Delta$ , 35.50% of the threads.

TABLE 3.1: Descriptive statistics

the higher the thread is, the wider it might be. But, even at the higher levels, the distribution of *Width* cover the whole range: it happens that discussion takes the shape of a unique branch. Besides, network measures (*Modularity* and *Av. Sackin*) are, by construction, elaborate on tree characteristics.

Concerning other variables, we do not observe any large correlation, except when one is a subset of the other: *OP Age* with *Av. Age Author*; *OP Experience* and *Av. Experience*; *#Delta from OP* and *#Delta*. Finally, we observe that *#Delta* and *Duration* are lightly and positively correlated with structural tree variables.

	#Post	#Author	#Root Answers	#Leaf	Width	Level Width	Height	Modularity	Av. Sackin
#Post	0.94***	0.97***	0.93***	0.41***	0.44***	-0.36***	0.41***	0.41***	
#Author	0.89***	0.98***	0.95***	0.28***	0.37***	-0.33***	0.25***	0.25***	
#Root Answers	0.87***	0.87***	0.92***	0.09***	0.35***	-0.37***	0.11***	0.11***	
#Leaf	0.97***	0.95***	0.96***	0.32***	0.37***	-0.33***	0.28***	0.28***	
Width	0.41***	0.28***	0.28***	0.37***	0.39***	-0.39***	0.23***	0.23***	
Level Width	0.44***	0.37***	0.39***	0.37***	0.37***	-0.29***	0.58***	0.58***	
Height	-0.36***	-0.33***	-0.39***	-0.33***	-0.68***	-0.48***	0.76***	0.76***	
Modularity	0.41***	0.25***	0.23***	0.58***	0.76***	-0.48***	0.15***	0.15***	
Av. Sackin	0.13***	0.14***	0.12***	0.09***	0.15***	-0.19***	0.12***	0.12***	
Av. Age Author (Month)	0.03***	0.03***	0.01***	0.04***	0.05***	-0.06***	0.07***	0.07***	
OP age (Month)	-0.11***	-0.12***	-0.13***	-0.03***	-0.01**	0.03***	0.02***	0.02***	
Av. Experience	-0.0	-0.02***	-0.01*	0.01	0.02***	-0.0	0.05***	0.05***	
OP experience	0.25***	0.26***	0.26***	0.09***	0.14***	-0.17***	0.1***	0.1***	
#Delta	0.1***	0.11***	0.12***	0.05***	0.09***	-0.13***	0.07***	0.07***	
#Delta from OP	0.37***	0.37***	0.36***	0.14***	0.17***	-0.15***	0.13***	0.13***	
Duration (Hour)									

	Av. Age Author (Month)	OP age (Month)	Av. Experience	OP experience	#Delta	#Delta from OP	Duration (Hour)
#Post	0.13***	0.03***	-0.11***	-0.0	0.25***	0.1***	0.37***
#Author	0.14***	0.03***	-0.12***	-0.01	0.26***	0.11***	0.37***
#Root Answers	0.1***	-0.0	-0.12***	-0.02***	0.24***	0.12***	0.33***
#Leaf	0.13***	0.02***	-0.12***	-0.01	0.26***	0.1***	0.37***
Width	0.12***	0.01***	-0.13***	-0.01*	0.26***	0.12***	0.36***
Level Width	0.09***	0.04***	-0.03***	0.01	0.09***	0.05***	0.14***
Height	0.15***	0.05***	-0.01**	0.02***	0.14***	0.09***	0.17***
Modularity	-0.19***	-0.06***	0.03***	-0.0	-0.17***	-0.13***	-0.15***
Av. Sackin	0.12***	0.07***	0.02***	0.05***	0.1***	0.07***	0.13***
Av. Age Author (Month)	0.62***	0.62***	0.3***	0.1***	0.09***	0.11***	-0.02**
OP age (Month)	0.62***	0.18***	0.18***	0.15***	0.01**	0.02***	-0.02***
Av. Experience	0.3***	0.18***	0.51***	0.51***	0.01**	0.05***	-0.07***
OP experience	0.1***	0.15***	0.51***	0.03***	0.03***	0.04***	-0.0
#Delta	0.09***	0.01**	0.01**	0.03***	0.91***	0.91***	0.1***
#Delta from OP	0.11***	0.02***	0.05***	0.04***	0.91***	0.91***	0.02***
Duration (Hour)	-0.02**	-0.02***	-0.07***	-0.0	0.1***	0.02***	0.02***

Note: Significance of correlation test of Pearson: "\*\*\*" - pvalue < 0.01, "\*\*" - pvalue < 0.05, "\*" - pvalue < 0.1, "" - non significant.

TABLE 3.2: Correlation matrix

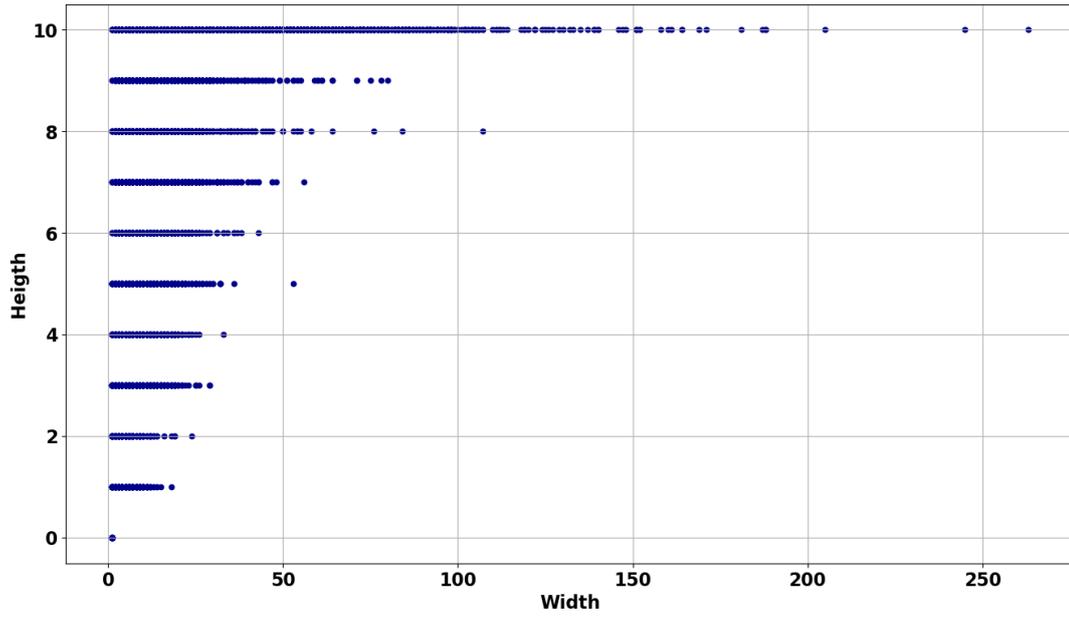
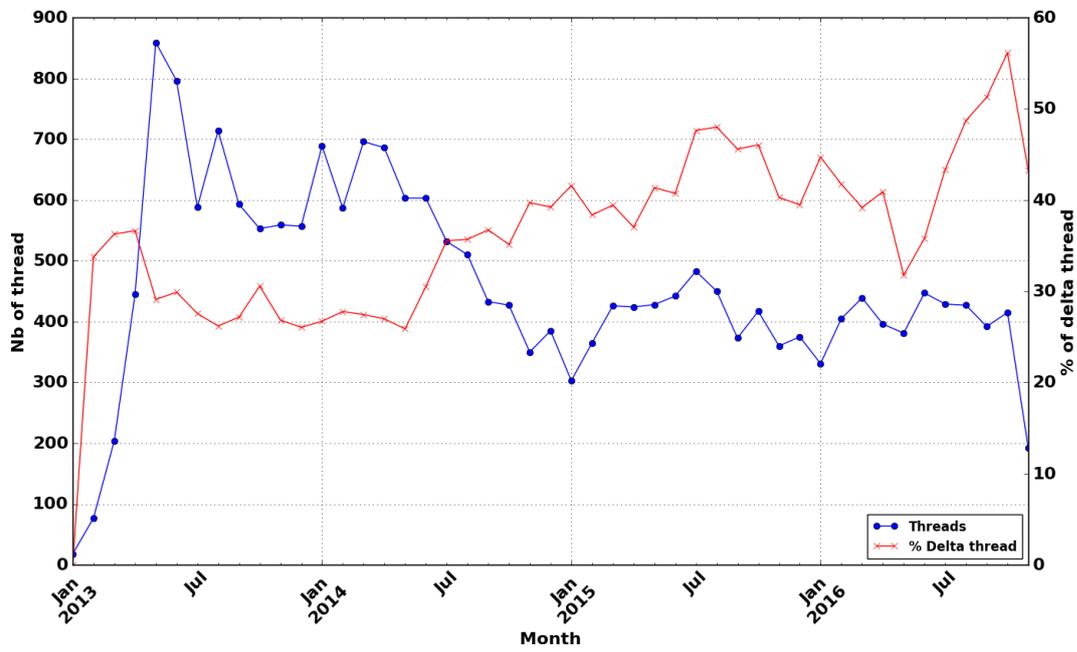
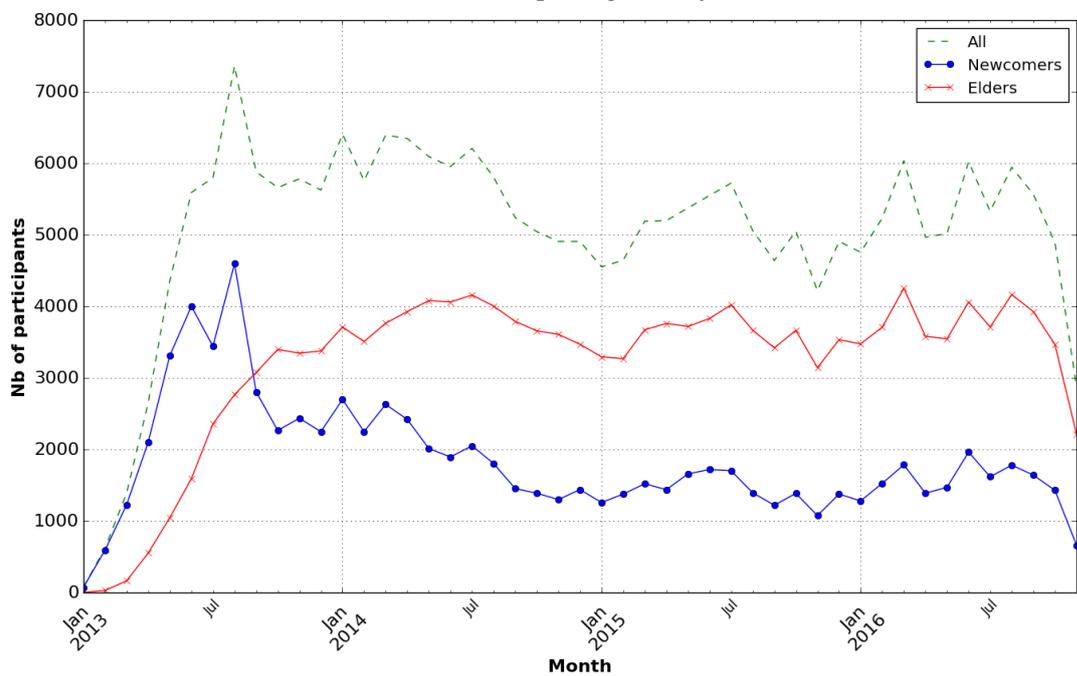


FIGURE 3.10: Width versus Height

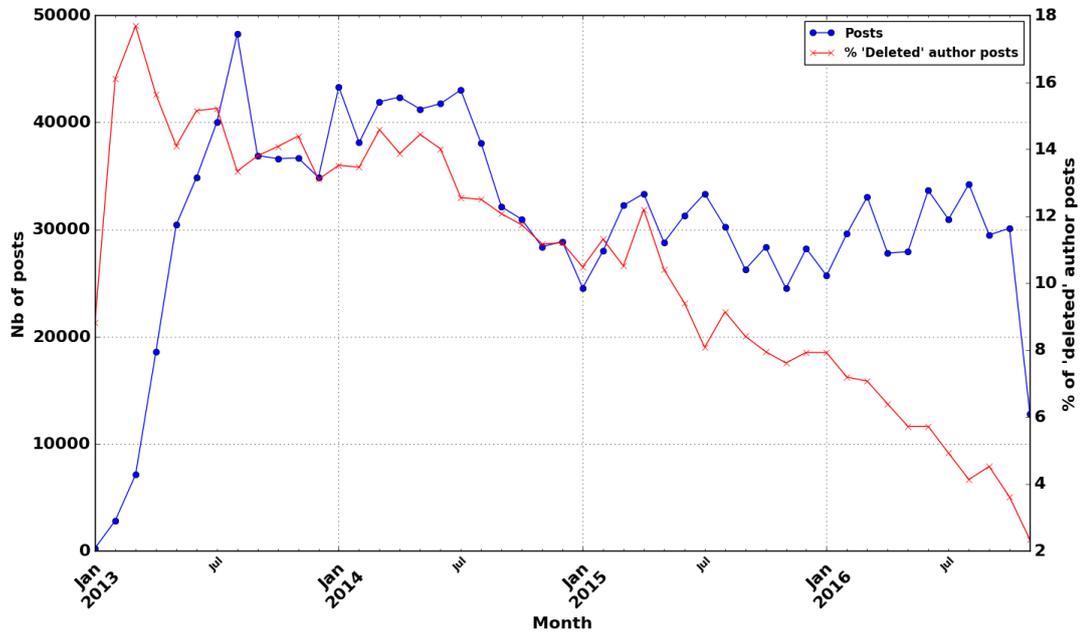


(A) Threads opening activity



(B) Unique active participants

FIGURE 3.7: Monthly activities



(c) Posts activity

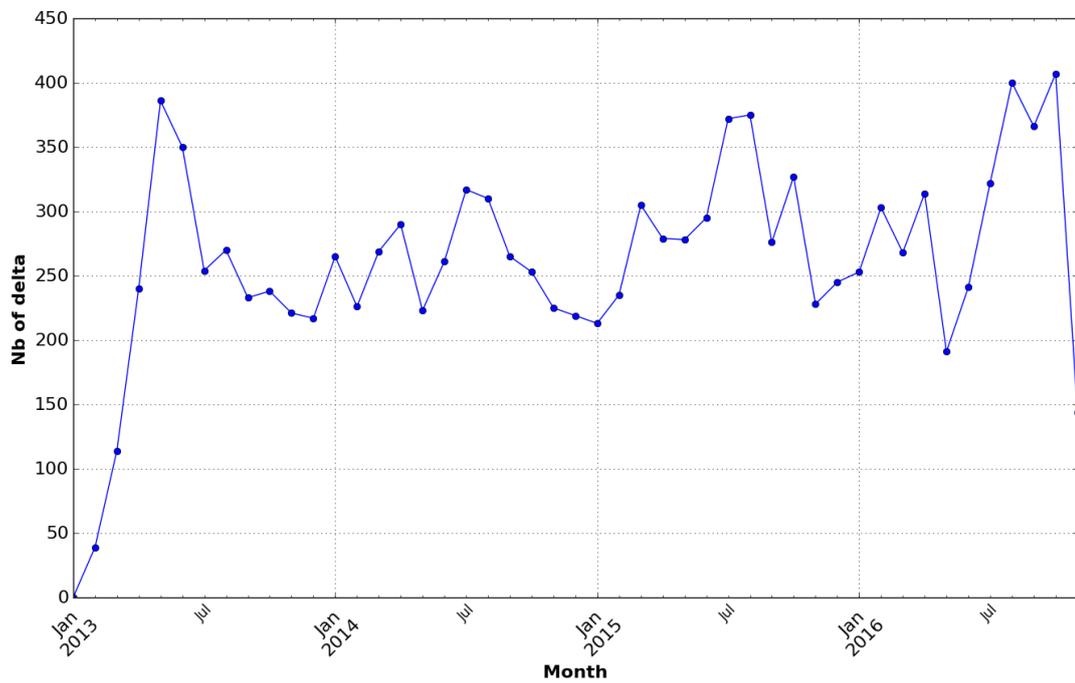
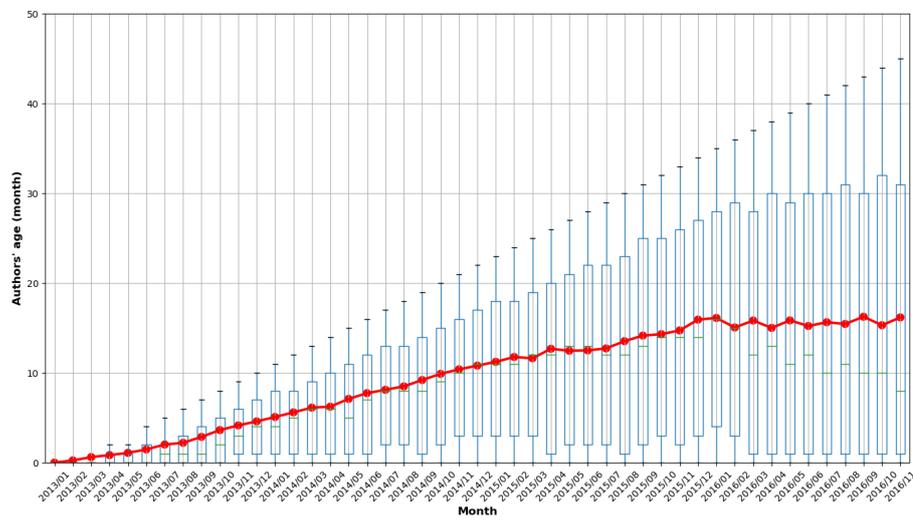
(d)  $\Delta$  activity

FIGURE 3.7: Monthly activities (cont.)



Note: Mean is plotted through the red dotted line.

(E) Distribution of participants age by month

FIGURE 3.7: Monthly activities (cont.)

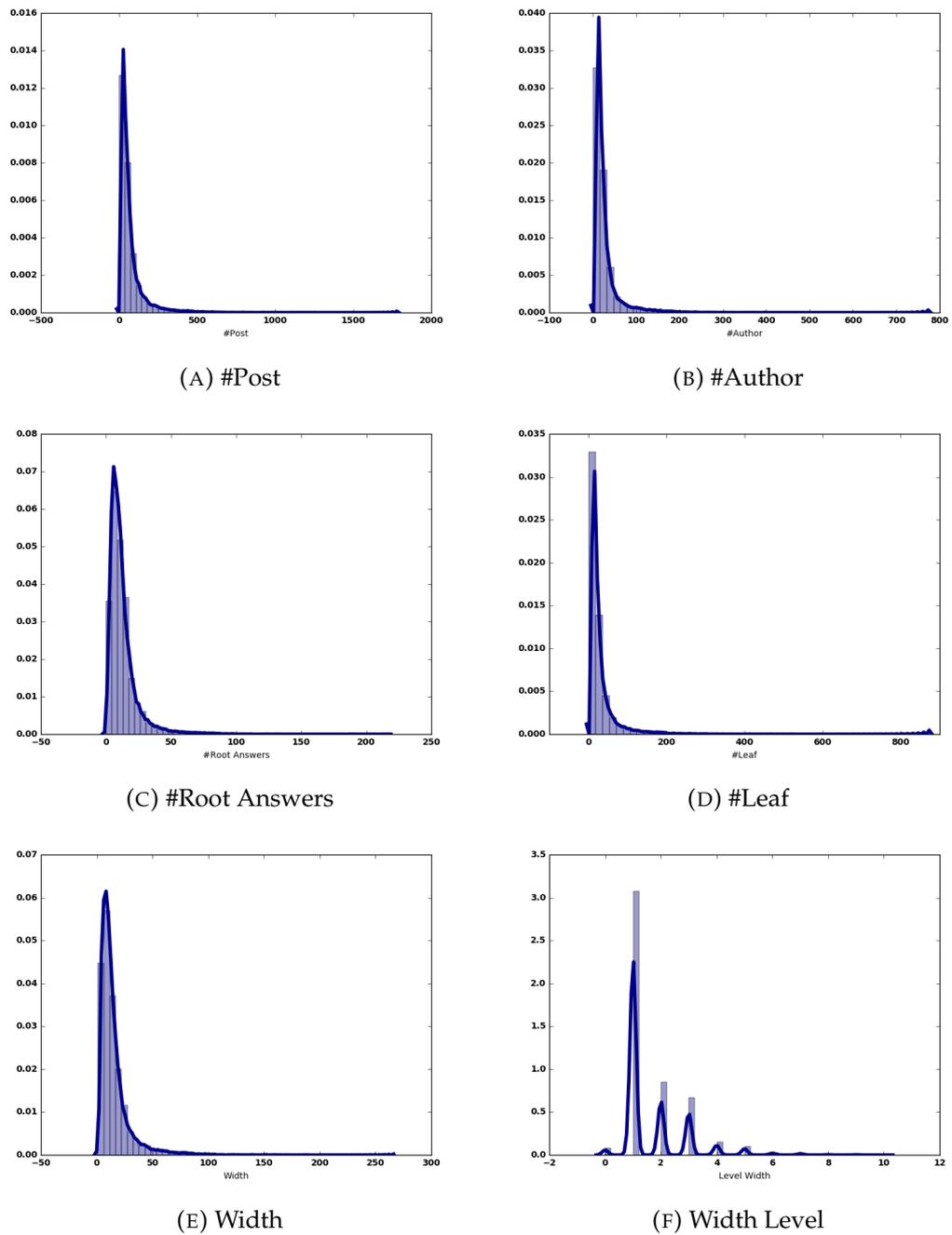
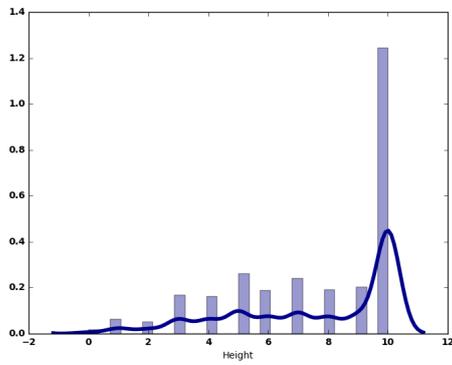
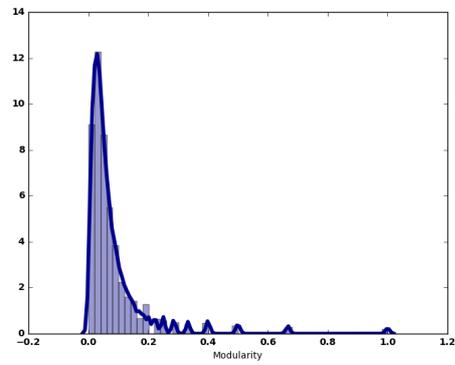


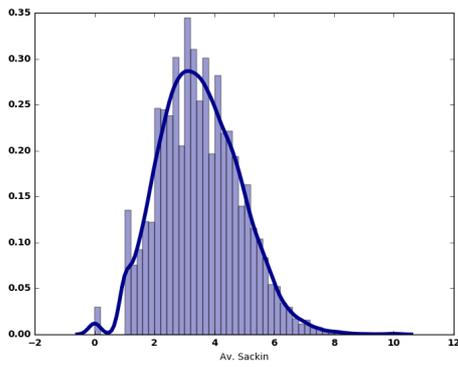
FIGURE 3.8: Normalized histogram and corresponding density



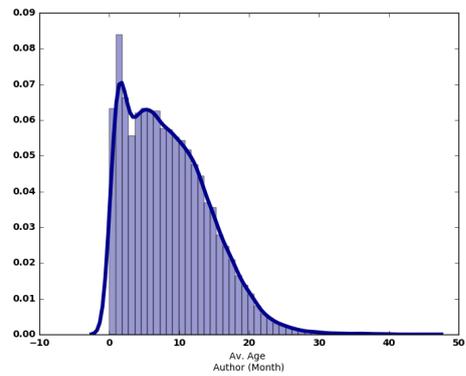
(G) Height



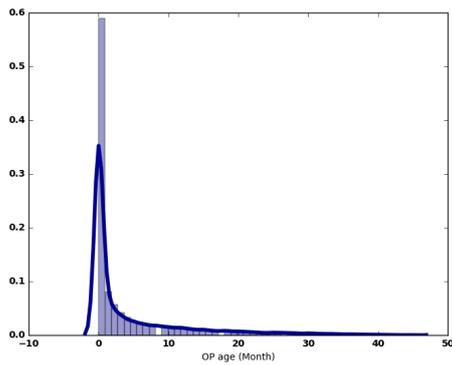
(H) Modularity



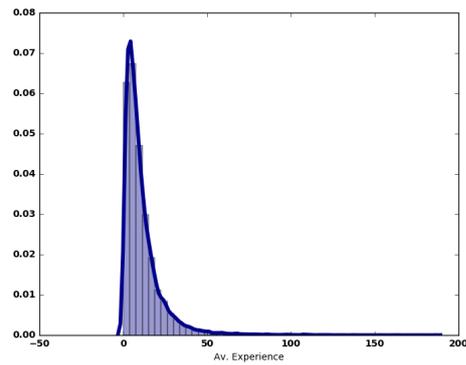
(I) Av. Sackin index



(J) Av. Age Author (Month)

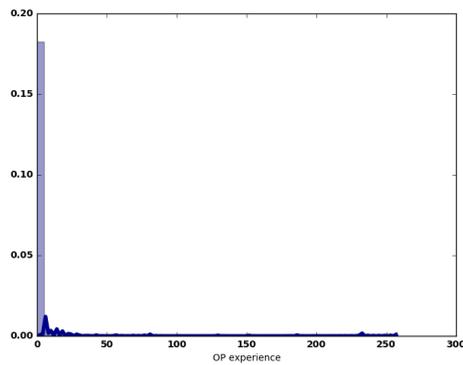


(K) OP Age (Month)

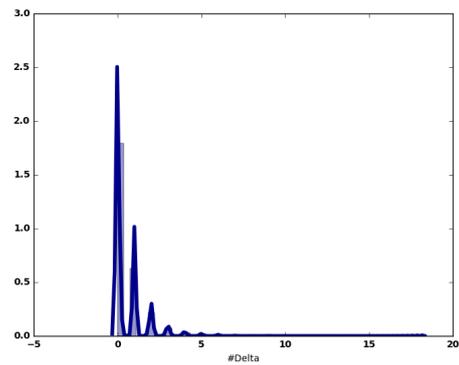


(L) Av. Experience

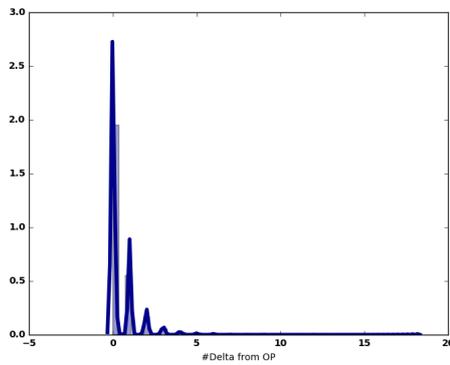
FIGURE 3.8: Normalized histogram and corresponding density (con't)



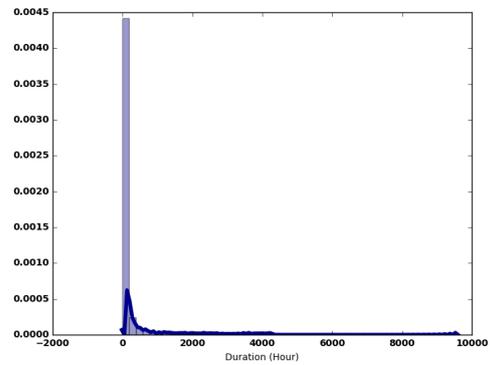
(M) OP Experience



(N) #Delta

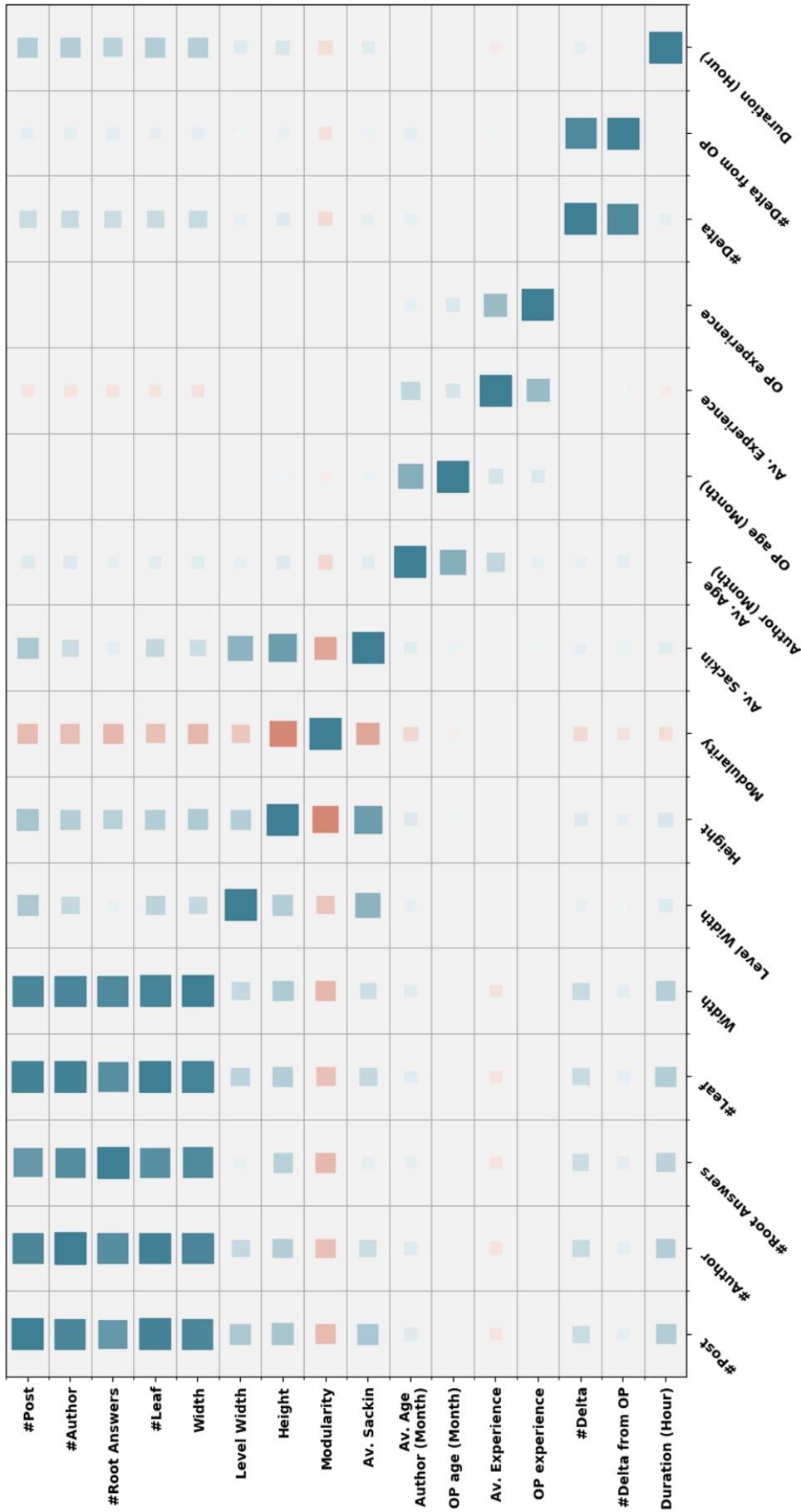


(O) #Delta awarded by OP



(P) Duration (Hour)

FIGURE 3.8: Normalized histogram and corresponding density  
(con't)



Note: The more the coefficient is close to one, the more the color tends toward a deep blue; the more the coefficient is close to -1, the more the color tends toward a deep red.

FIGURE 3.9: Correlation heatmap



# Chapter 4: Different types of Debates?

This chapter is an update of Mano, Dalle, and Tomasik (2017).

4.1	Who: Hypothesis — discussion categories . . . . .	82
4.2	Why: Discussion levels . . . . .	84
4.2.1	Reddit — Change My View (CMV) . . . . .	84
4.3	How: Clustering . . . . .	85
4.3.1	Author Networks and Motifs . . . . .	85
4.3.2	Clustering Algorithm . . . . .	86
4.3.2.1	<i>k</i> -means Clustering . . . . .	86
4.3.2.2	Spectral Clustering . . . . .	86
4.3.2.3	Hierarchical and Ward Clusterings . . . . .	87
4.4	What: Results . . . . .	87
4.4.1	Descriptive Statistics . . . . .	87
4.4.1.1	Motif features . . . . .	88
4.4.1.2	Roles in the motifs . . . . .	89
4.4.2	Clustering analysis . . . . .	96
4.4.2.1	Optimal clustering . . . . .	96
4.4.2.2	Clustering results . . . . .	102
4.5	Conclusion . . . . .	104

## 4.1 Who: Hypothesis — discussion categories

Since the development of the Internet people commonly exchange messages online — from the 1980s and the first *newsgroups* to today's online forums, social networks, chats, etc. Even after the expansion of social networks, in the U.S., still 15% of Internet users exchange actively on forums (Duggan et al., 2015). Online discussions are active on any kind of subject, from politics to videogames, to finding a technical answer to a problem through a Q&A site, or even to seek the help of fellow students in Massive Open Online Courses (MOOC) forums. The biggest sites where online discussions take place, like StackOverflow, gather millions of threads and users, highlighting the importance of such platforms in a social life.

Open collaboration research is a major field where online collaboration has emerged and developed on the Internet. The understanding of coordination and collaboration has interested numerous researchers over the past 15 years. However, and even if online discussions have regularly been addressed in the literature, they have only rarely been the main focus and almost always second to other issues. Furthermore, evidence for self-organizational properties has started to be gathered (Besten, Dalle, and Galia, 2008), which has pleaded for the existence of stigmergic phenomena driving the allocation of online efforts and thus orienting coordination within Free/Libre Open Source Software (FLOSS) projects (Bolici, Howison, and Crowston, 2009; Dalle and David, 2003), for instance. These coordination models assume that the code and/or other existing online artifacts actually mediate coordination, making it mostly indirect and/or implicit, in such a way that reduces transaction costs (Dalle and David, 2003).

The latter approach has recently received further support through the emergence of theoretical notions such as transparency (Dabbish et al., 2012) or superposition (Howison and Crowston, 2014), which in both cases also tend to emphasize indirect and/or implicit coordination mechanisms, certainly not as *opposed* to online discussions, but *rather than* direct dialectic coordination, up to the point that both of these approaches curiously neglect

the importance of semantic signals even though they stand at the core of stigmergic approaches (Besten et al., 2010).

Such a theoretical framework contrasts with a number of other researches works that have more directly focused on the role of online discussions in other types of online communities. Altogether, these inquiries tend to cast some doubts on the relative neglect from which online discussions have suffered in online communities, perhaps also due to the fact that some discussion channels such as instant messaging are not easy to access for researchers. Notwithstanding the relevance of the stigmergic-transparent-superposed (STS) model with respect to an emerging theory of the 'FLOSS way', we believe that further inquiries about the nature and characteristics of online discussions to be a relevant issue. Furthermore, online discussions within online communities might also endorse specific aspects to the 'FLOSS way' and might be affected by or conversely influence the STS model in ways that have yet to be understood. Finding ways to appropriately cluster online discussions thus appears as a specially relevant method with respect both to the understanding of the determinants of discussion outcomes *à la* Alan Cox (1998), and to the identification of potentially relevant features of online discussions *vis-à-vis* the STS model.

The current work underlines the heterogeneity in coordination modes for discussions inside a forum, where rules are defined to frame one unique way to coordinate. We first present our level of analysis: the discussions themselves. Then we present the methodology used. Before concluding, we present our results.

## 4.2 Why: Discussion levels

### 4.2.1 Reddit — Change My View (CMV)

As in Tan et al. (2016), we use an open database<sup>1</sup> from a sub-reddit called Change My View, a forum where an Original Poster (OP) opens a discussion about a topic and ask the community to bring him evidences to change his view on this topic (cf. Chapter 3 for more detail). We select threads with at least 10 posts and 2 authors. Using Reddit API, we extract all discussions from January 2013, creation of CMV, to November 2016. Thus, we analyzed 19,956 discussions, composed of 1,432,588 posts, 13,371 OP and 71,865 unique authors<sup>2</sup>, who did not opened a thread, producing in average 71.79 posts each. Beside, an account could be deleted. If so, the pseudonym was replaced by "[deleted]". 155,371 posts - 10.85% - have been written under this pseudonym. Within a discussion, we define all posts from a "[deleted]" account coming from the same author. Finally, we do not observe "deleted" account for the OPs.

Contrary to mailing list, in this forum, as a user can answer directly to a post, the "thread" framework preserves the thematic coherence of a discussion. Moreover, we access an important quantity of information, such as user past activity. Among such information, we recover data on the reward system implemented by reddit. Such system consists in award a delta  $\Delta$  to a post if it has change my view. The  $\Delta$  system allows to ascertain on the *quality* of a conversation, being aware of its bias. Even if the  $\Delta$  does not guarantee a full change of view, it gives some hints on the number of relevant arguments and thus, if a discussion progresses or stagnates.

---

<sup>1</sup><https://chenhaot.com/pages/changemyview.html>

<sup>2</sup>identified by their pseudonyms.

## 4.3 How: Clustering

### 4.3.1 Author Networks and Motifs

We model a discussion as a rooted tree. We extract the corresponding graph whose vertices model authors. When author  $A$  answers to author  $B$ 's post, we draw an arc from  $A$  to  $B$ . Figure 4.1 presents discussions with, on the left, the tree graph and on the right the corresponding author network.

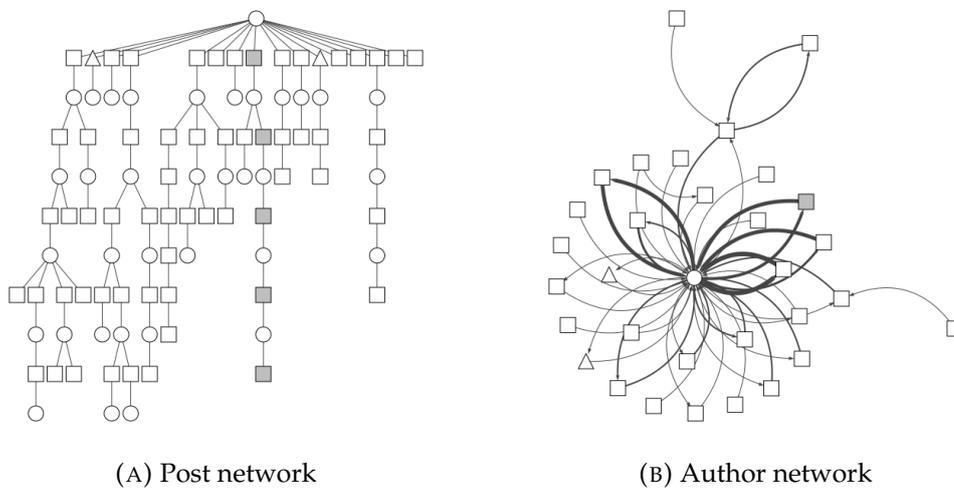


FIGURE 4.1: Different modeling of the same discussion

In order to cluster our discussions we used the last thirteen motifs, following Milo et al. (2002), presented on Figure 4.3.

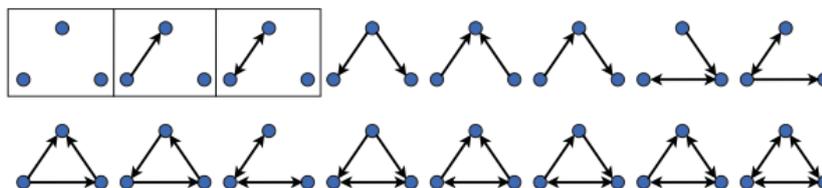


FIGURE 4.2: Triad dictionary from Cunningham et al. (2013)

### 4.3.2 Clustering Algorithm

In order to offer robust results, we compared different clustering algorithms. The following section presents the four algorithms used to test the heterogeneity of the discussion types in CMV. The four algorithms are  $k$ -means, Spectral, Hierarchical Clustering and Ward.

#### 4.3.2.1 $k$ -means Clustering

Firstly, we apply a  $k$ -means algorithm to cluster authors into different subgroups. This algorithm clusters feature vectors (or authors), based upon their similarity. With mathematical modeling tools, researchers are able to highlight contexts or behaviors which are similar. More precisely, the  $k$ -means algorithm clusters data into  $k$  groups, where the number of clusters to be formed is given. Then, the algorithm sets  $k$  points randomly (with the dimension of coordinates being equal to the size of the vector defined in input). The coordinates are updated during each step, with respect to the average of elements which are the closest to each cluster center. This algorithm is very efficient, but really dependent on the initial position of the centers, that are set randomly. We used the  $k$ -means++ methodology, defined by Arthur and Vassilvitskii (2007), to define the initial centers, the initial *means*.

The  $k$  means methodology also implies to choose the number of cluster  $k$  in which we want to group our users. Several methodologies help researchers to select the proper  $k$ . We used the Silhouette method, developed by Rousseeuw (1987). In practice, we compute the  $k$ -means algorithm for different values of  $k$  - from 2 to 10. Then, we calculate the average Silhouette of observations, and doing so for each  $k$ . Finally, we select  $k$  such that it maximizes the the average Silhouette.

#### 4.3.2.2 Spectral Clustering

We also use Ng, Jordan, and Weiss (2001) that allows us to apply spectral clustering. Here again, the number of clusters has to be specified in advance. We apply the Silhouette method to select the optimal number of clusters.

Before running a proper clustering algorithm, it performs a low-dimension embedding of the affinity matrix between samples. Then it performs the clustering (for instance, a  $k$ -means clustering), of the components of the eigenvectors in the low-dimensional space.

#### 4.3.2.3 Hierarchical and Ward Clusterings

Hierarchical clustering algorithm — HCA, presented in Rokach and Maimon (2005), is a family of algorithms that build nested clusters by merging or splitting them successively. This hierarchy of clusters is represented as a *dendrogram*. At the top of it, we find the unique cluster that gathers all the samples, the leaves being the clusters with only one sample. We could create a hierarchy either with an *agglomerative* approach (from clusters with one observation, to the cluster with all observation) or the *divisive* approach (from the cluster with all observations to clusters with one observation each). We need to choose, as well, the *linkage criteria*, which determines the metric used for the merge strategy.

Alternatively, Ward clustering (Ward, 1963) minimizes the sum of squared differences within all clusters. It is a variance-minimizing approach and in this sense is similar to the  $k$ -means objective function but tackled with an agglomerative hierarchical approach.

## 4.4 What: Results

### 4.4.1 Descriptive Statistics

Following the presentation from Chapter 3, Table 4.1, Table 4.2 and Figure 5.3 present the different descriptive statistics of this sub-dataset. Comparing the *tree features* between the two datasets (the full one presented in Chapter 3 and the dataset of the current chapter), we can see that the differences are not really high. Still, the database presented in this chapter attract more users on average than the average of the full database. The following section presents the new result: the distribution of motifs in these discussions.

#### 4.4.1.1 Motif features

Chapter 3 presented the descriptive statistics for the *non-motif* characteristics of the discussions. This section will focus on the motifs ones. Researchers focus on subgraph (mainly on dyads and triads, respectively combination of two and three nodes) (Wasserman and Faust, 1994). The methodology consists in counting occurrences of a certain subgraph of size  $k$  and make a comparison with a reference model, often built by randomizing the real graphs, preserving some of their properties. Holland and Leinhardt (1970) characterize the sixteen possible triads (cf. Figure 4.3). Milo et al. (2002) defined as *network motifs* subgraphs occurring significantly more than those in randomized networks.

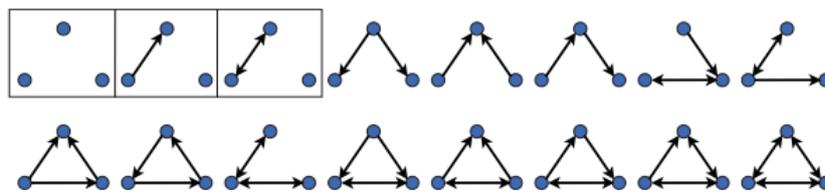


FIGURE 4.3: Triad dictionary from Cunningham et al. (2013)

Motif analysis allows the improvement of network analysis. It leads to the diminution of computational needs, predictions on selection of motifs being not degraded compared to ones made with the whole network (Cunningham et al., 2013). It also highlights specific characteristics of different types of threads (Adamic et al., 2008), in different communities (Conaldi and Lomi, 2013; Lomi et al., 2014).

Thus, we selected networks with at least 10 posts and at least 3 authors involved in the discussion, following Tan et al. (2016). Table 4.1 presents the motif percentage in the 19,956 networks.

Motifs  $A \rightarrow B \leftarrow C$ ,  $A \leftrightarrow B \leftarrow C$  and  $A \leftrightarrow B \leftrightarrow C$  embody almost 90% of motifs distribution.

Motifs are not correlated with each other (Table 4.2 or Figure 5.3) except for two, which are significantly highly and negatively correlated:  $A \rightarrow B \leftarrow C$  and  $A \leftrightarrow B \leftrightarrow C$ . Moreover, they are both correlated to several "tree

features", and always in an opposite direction: number of leaves, the OP degree, width, on the one hand, and Sackin index, to a lesser extent, on the other.

We assume that, if a discussion has an important share of motifs  $A \rightarrow B \leftarrow C$ , it characterizes the initiation of several sub-discussion between authors. This implies that OP has several challengers but does not answer to all of them. On the contrary, if a discussion presents an important share of  $A \leftrightarrow B \leftrightarrow C$ , OP answers to a majority of his challengers. We assume in this case that the number of challengers is lower than in the previous case. Moreover, the latter motif is, even weakly, negatively correlated to the percentage of experts. We thus assume that discussion characterized by  $A \leftrightarrow B \leftrightarrow C$  attract more experts. Besides, in 98% of discussions, the opening post will have the highest degree. Thus, we assume that the central node in those particular triads ( $B$ ) is mainly the OP. With this assumption, the first motif characterizes a discussion where the OP answers few comments (which we label "wide" tree), where the second one is characterized by the situation in which the OP discusses several times with his or her challengers (which we label "deep" tree). The next section tests these assumptions.

#### 4.4.1.2 Roles in the motifs

Figure 4.5 presents the triads<sup>3</sup> we used to the algorithm, and the *role* in each triad. A role is attributed to a node in respect with its position within a triad. Among the 13 triads, we count 30 roles. Knowing more about who plays which role allows us to develop a robust analysis of the clustering results.

Therefore, for each author in each thread, we calculate the distribution of the role this author had<sup>4</sup>. Table 4.3 and Figure 4.6 present the correlation between the possible roles and the following *discussion characteristics*: the author is or is not the OP ( $OP$ , equals to 1 if author is OP), the experience of the author ( $Experience$ ), the  $\Delta$  is given by the OP or not ( $\Delta$  received by, equal to 1 if

<sup>3</sup>Combination of three nodes.

<sup>4</sup>I would like to greatly thank Raphaël Charbey who developed the code to calculate the role distribution.



	count	mean	std	min	25%	50%	75%	max
#Post	19956	71.79	90.37	10.0	25.00	45.00	79.00	1776.00
#Author	19956	27.04	32.52	3.0	11.00	18.00	29.00	774.00
#Root Answers	19956	13.34	11.94	1.0	7.00	10.00	16.00	216.00
#Leaf	19956	28.33	37.83	1.0	10.00	17.00	30.00	874.00
Width	19956	15.77	14.87	1.0	8.00	12.00	18.00	263.00
Level Width	19956	1.73	1.15	0.0	1.00	1.00	2.00	10.00
Height	19956	8.03	2.35	1.0	6.00	9.00	10.00	10.00
Modularity	19956	0.06	0.04	0.0	0.03	0.04	0.08	0.20
Av. Sackin	19956	3.65	1.26	1.0	2.71	3.55	4.50	10.00
Av. Age Author (Month)	19956	8.70	5.71	0.0	4.06	8.05	12.54	35.92
OP Age (Month)	19956	4.50	7.82	0.0	0.00	0.00	6.00	45.00
Av. Experience	19956	11.22	11.98	0.0	4.16	7.83	14.07	175.40
OP Experience	19945	2.63	17.35	0.0	0.00	0.00	0.00	257.00
#Delta*	7478	1.62	1.13	1.0	1.0	1.0	2.0	18.0
#Delta from OP*	7478	1.31	1.16	0.0	1.0	1.0	2.0	18.0
Duration (Hour)	19956	271.78	709.61	0.8	27.41	50.90	113.79	9578.17
$A \leftarrow B \rightarrow C$	19956	0.69	1.44	0.0	0.00	0.00	0.82	20.83
$A \rightarrow B \leftarrow C$	19956	30.40	24.02	0.0	10.00	26.44	46.56	100.00
$A \rightarrow B \rightarrow C$	19956	3.58	5.01	0.0	0.00	1.74	5.26	69.23
$A \leftrightarrow B \leftarrow C$	19956	34.90	15.70	0.0	25.30	38.08	47.04	100.00
$A \leftrightarrow B \rightarrow C$	19956	4.06	5.43	0.0	0.00	2.08	6.23	100.00
$A \rightarrow B \leftarrow C, A \rightarrow C$	19956	0.25	0.85	0.0	0.00	0.00	0.00	16.67
$A \leftarrow B \leftarrow C, A \rightarrow C$	19956	0.04	0.34	0.0	0.00	0.00	0.00	16.67
$A \leftrightarrow B \leftrightarrow C$	19956	24.28	23.91	0.0	6.67	16.38	34.09	100.00
$A \leftarrow B \rightarrow C, A \leftrightarrow C$	19956	0.20	0.74	0.0	0.00	0.00	0.00	20.00
$A \rightarrow B \leftarrow C, A \leftrightarrow C$	19956	0.23	0.92	0.0	0.00	0.00	0.00	33.33
$A \rightarrow B \rightarrow C, A \leftrightarrow C$	19956	0.28	1.21	0.0	0.00	0.00	0.05	100.00
$A \rightarrow B \leftrightarrow C, A \leftrightarrow C$	19956	0.78	2.21	0.0	0.00	0.00	0.84	100.00
$A \leftrightarrow B \leftrightarrow C, A \leftrightarrow C$	19956	0.32	1.56	0.0	0.00	0.00	0.00	100.00

Note: \* - count on threads with at least one  $\Delta$ , 35.50% of the threads.

TABLE 4.1: Descriptive statistics for motif

it is coming from OP, to 2 if it comes from another author, 0 otherwise), and finally the total number of  $\Delta$  given by the author in the thread ( $\Delta$  given).

First of all, neither the level of experience and from whom one received a  $\Delta$  have a significant correlation with the role of the author. However, the status of the author (OP or not) and the number of  $\Delta$  he has given are significantly and positively correlated. We can conclude that, as defined by the rules of the forum, it is mainly OP authors who award authors with a  $\Delta$ .

Moreover, looking at the OP variable, its correlation with the roles confirms previous assumptions. OP is significantly and positively correlated

	#Post	#Author	#Root Answers	#Leaf	Width	Level Width	Height	Modularity	Av. Sackin
#Post	0.94***	0.79***	0.88***	0.97***	0.93***	0.38***	0.41***	-0.54***	0.39***
#Author	0.94***	0.88***	0.97***	0.98***	0.95***	0.26***	0.32***	-0.49***	0.2***
#Root Answers	0.79***	0.88***	0.86***	0.86***	0.91***	0.04***	0.28***	-0.51***	0.03***
#Leaf	0.97***	0.98***	0.86***	0.96***	0.96***	0.3***	0.33***	-0.5***	0.25***
Width	0.93***	0.95***	0.91***	0.96***	0.25***	0.25***	0.33***	-0.54***	0.32***
Level Width	0.38***	0.26***	0.04***	0.3***	0.33***	0.32***	0.32***	-0.74***	-0.74***
Height	0.41***	0.32***	0.28***	0.33***	0.33***	0.32***	0.32***	-0.74***	0.71***
Modularity	-0.54***	-0.49***	-0.51***	-0.5***	-0.54***	-0.33***	-0.74***	-0.51***	-0.51***
Av. Sackin	0.39***	0.2***	0.03***	0.25***	0.17***	0.58***	0.71***	-0.51***	0.12***
Av. Age Author (Month)	0.12***	0.14***	0.09***	0.12***	0.11***	0.09***	0.13***	-0.18***	0.09***
OP Age (Month)	0.03***	0.03***	-0.0	0.02**	0.01	0.05***	0.06***	-0.06***	0.09***
Av. Experience	-0.12***	-0.13***	-0.13***	-0.13***	-0.14***	-0.04***	-0.02***	0.08***	0.01**
OP Experience	-0.0	-0.01	-0.02***	-0.01	-0.01*	0.01	0.03***	-0.01	0.05***
#Delta	0.23***	0.25***	0.22***	0.24***	0.25***	0.07***	0.09***	-0.18***	0.06***
#Delta from OP	0.09***	0.09***	0.09***	0.08***	0.1***	0.03***	0.04***	-0.11***	0.03***
Duration (Hour)	0.36***	0.36***	0.32***	0.37***	0.36***	0.13***	0.15***	-0.21***	0.11***
A ← B → C	0.22***	0.19***	0.09***	0.21***	0.19***	0.18***	0.13***	-0.2***	0.12***
A → B → C	0.05***	0.22***	0.33***	0.17***	0.22***	-0.28***	-0.25***	0.08***	-0.57***
A → B → C	0.09***	0.12***	0.0	0.1***	0.07***	0.1***	-0.02**	0.02**	-0.02**
A ↔ B → C	-0.07***	-0.09***	-0.1***	-0.08***	-0.1***	-0.04***	0.0	-0.01*	-0.03***
A ↔ B → C	0.16***	0.1***	-0.03***	0.11***	0.08***	0.24***	0.17***	-0.11***	0.23***
A → B → C, A → C	-0.04***	-0.02***	-0.03***	-0.02***	-0.03***	-0.02***	-0.07***	0.05***	-0.09***
A ← B → C, A → C	-0.02***	-0.02**	-0.03***	-0.02***	-0.02***	0.0	-0.03***	0.04***	-0.01
A ↔ B ↔ C	-0.06***	-0.21***	-0.25***	-0.17***	-0.19***	0.2***	0.2***	-0.04***	0.52***
A ← B → C, A ↔ C	-0.02**	-0.03***	-0.05***	-0.02**	-0.03***	0.05***	0.02**	-0.02***	0.01**
A → B → C, A ↔ C	-0.04***	-0.04***	-0.05***	-0.03***	-0.04***	-0.01*	-0.02***	0.05***	-0.05***
A → B → C, A ↔ C	-0.02***	-0.03***	-0.05***	-0.03***	-0.03***	0.04***	0.0	0.02**	0.03***
A → B ↔ C, A ↔ C	-0.02***	-0.07***	-0.1***	-0.05***	-0.06***	0.13***	0.07***	-0.02***	0.15***
A ↔ B ↔ C, A ↔ C	-0.01	-0.05***	-0.06***	-0.03***	-0.04***	0.1***	0.06***	-0.03***	0.13***

Note: Significance of correlation test of Pearson: "\*\*\*\*" - p-value < 0.01, "\*\*\*" - p-value < 0.05, "\*\*" - p-value < 0.1, "" - non significant.

TABLE 4.2: Correlation matrix

	Av. Age Author (Month)	OP Age (Month)	Av. Experience	OP Experience	#Delta	#Delta from OP	Duration (Hour)
#Post	0.12***	0.03***	-0.12***	-0.0	0.23***	0.09***	0.36***
#Author	0.14***	0.03***	-0.13***	-0.01	0.25***	0.09***	0.36***
#Root Answers	0.09***	-0.0	-0.13***	-0.02***	0.22***	0.09***	0.32***
#Leaf	0.12***	0.02**	-0.13***	-0.01	0.24***	0.08***	0.37***
Width	0.11***	0.01	-0.14***	-0.01*	0.25***	0.1***	0.36***
Level Width	0.09***	0.05***	-0.04***	0.01	0.07***	0.03***	0.13***
Height	0.13***	0.06***	-0.02***	0.03***	0.09***	0.04***	0.15***
Modularity	-0.18***	-0.06***	0.08***	-0.01	-0.18***	-0.11***	-0.21***
Av. Sackin	0.12***	0.09***	0.01**	0.05***	0.06***	0.03***	0.1***
Av. Age Author (Month)		0.6***	0.31***	0.09***	0.09***	0.11***	-0.03***
OP Age (Month)			0.18***	0.15***	0.01	0.02***	-0.02***
Av. Experience				0.5***	0.01*	0.05***	-0.08***
OP Experience					0.03***	0.04***	-0.0
#Delta						0.91***	0.09***
#Delta from OP							0.01*
Duration (Hour)							
$A \leftarrow B \rightarrow C$	0.01*	-0.02***	-0.07***	-0.0	0.09***	0.01*	0.11***
$A \rightarrow B \leftarrow C$	-0.05***	-0.09***	-0.09***	-0.07***	-0.09***	-0.14***	0.06***
$A \rightarrow B \rightarrow C$	-0.04***	-0.05***	-0.08***	-0.02***	-0.03***	-0.08***	0.08***
$A \leftrightarrow B \leftarrow C$	0.12***	0.07***	0.08***	0.0	0.08***	0.12***	-0.04***
$A \leftrightarrow B \rightarrow C$	-0.07***	-0.05***	-0.08***	-0.02**	-0.03***	-0.08***	0.08***
$A \rightarrow B \leftarrow C, A \rightarrow C$	-0.05***	-0.04***	-0.04***	-0.02***	-0.06***	-0.07***	0.01
$A \leftarrow B \leftarrow C, A \rightarrow C$	-0.01*	-0.02**	0.02***	-0.0	-0.02**	-0.02**	-0.01
$A \leftrightarrow B \leftrightarrow C$	0.0	0.07***	0.08***	0.08***	0.06***	0.11***	-0.08***
$A \leftarrow B \rightarrow C, A \leftrightarrow C$	0.01**	0.01	0.0	0.0	-0.01	-0.01*	0.0
$A \rightarrow B \leftarrow C, A \leftrightarrow C$	-0.06***	-0.04***	-0.03***	-0.02***	-0.06***	-0.07***	-0.0
$A \rightarrow B \rightarrow C, A \leftrightarrow C$	-0.02**	-0.01	-0.0	0.01	0.0	0.0	0.0
$A \rightarrow B \leftrightarrow C, A \leftrightarrow C$	-0.05***	-0.03***	-0.01*	-0.0	-0.02**	-0.01*	-0.01
$A \leftrightarrow B \leftrightarrow C, A \leftrightarrow C$	-0.03***	-0.01	-0.0	0.01**	0.01	0.01	-0.01

Note: Significance of correlation test of Pearson: "\*\*\*" - pvalue < 0.01, "\*\*" - pvalue < 0.05, "\*" - pvalue < 0.1, "" - non significant.

TABLE 4.2: Correlation matrix (con't)

	$A \leftarrow B \rightarrow C$	$A \rightarrow B \leftarrow C$	$A \rightarrow B \rightarrow C$	$A \leftrightarrow B \leftarrow C$	$A \leftrightarrow B \rightarrow C$	$A \rightarrow B \leftarrow C$	$A \leftrightarrow B \rightarrow C$	$A \rightarrow B \leftarrow C$	$A \leftarrow B \leftarrow C$	$A \leftarrow B \rightarrow C$
#Post	0.22***	0.05***	0.09***	-0.07***	0.16***	-0.04***	-0.02***	-0.02***	-0.02***	-0.02***
#Author	0.19***	0.22***	0.12***	-0.09***	0.1***	-0.02***	-0.02***	-0.02***	-0.02***	-0.02***
#Root Answers	0.09***	0.33***	0.0	-0.1***	-0.03***	-0.03***	-0.03***	-0.03***	-0.03***	-0.03***
#Leaf	0.21***	0.17***	0.1***	-0.08***	0.11***	-0.02***	-0.02***	-0.02***	-0.02***	-0.02***
Width	0.19***	0.22***	0.07***	-0.1***	0.08***	-0.03***	-0.03***	-0.03***	-0.03***	-0.03***
Level Width	0.18***	-0.28***	0.1***	-0.04***	0.24***	-0.02***	-0.02***	-0.02***	-0.02***	-0.02***
Height	0.13***	-0.25***	-0.02**	0.0	0.17***	-0.07***	-0.07***	-0.07***	-0.07***	-0.07***
Modularity	-0.2***	0.08***	0.02**	-0.01*	-0.11***	0.05***	0.05***	0.05***	0.05***	0.05***
Av. Sackin	0.12***	-0.57***	-0.02**	-0.03***	0.23***	-0.09***	-0.09***	-0.09***	-0.09***	-0.09***
Av. Age Author (Month)	0.01*	-0.05***	-0.04***	0.12***	-0.07***	-0.05***	-0.05***	-0.05***	-0.05***	-0.05***
OP Age (Month)	-0.0	-0.09***	-0.05***	0.07***	-0.05***	-0.04***	-0.04***	-0.04***	-0.04***	-0.04***
Av. Experience	-0.07***	-0.09***	-0.08***	0.08***	-0.08***	-0.04***	-0.04***	-0.04***	-0.04***	-0.04***
OP Experience	0.0	-0.07***	-0.02***	0.0	-0.02**	-0.02***	-0.02***	-0.02***	-0.02***	-0.02***
#Delta	0.01	-0.09***	-0.03***	0.08***	-0.03***	-0.06***	-0.06***	-0.06***	-0.06***	-0.06***
#Delta from OP	-0.05***	-0.14***	-0.08***	0.12***	-0.08***	-0.07***	-0.07***	-0.07***	-0.07***	-0.07***
Duration (Hour)	0.11***	0.06***	0.08***	-0.04***	0.08***	0.01	-0.01	-0.01	-0.01	-0.01
$A \leftarrow B \rightarrow C$	0.02***	0.12***	0.12***	-0.35***	0.39***	0.2***	0.2***	0.2***	0.2***	0.2***
$A \rightarrow B \leftarrow C$	0.36***	-0.35***	0.36***	-0.14***	0.33***	0.15***	0.15***	0.15***	0.15***	0.15***
$A \rightarrow B \rightarrow C$	0.36***	0.12***	0.12***	-0.23***	0.39***	0.22***	0.22***	0.22***	0.22***	0.22***
$A \leftrightarrow B \leftarrow C$	-0.14***	-0.35***	-0.23***	-0.35***	-0.27***	-0.17***	-0.17***	-0.17***	-0.17***	-0.17***
$A \leftrightarrow B \rightarrow C$	0.33***	-0.04***	0.39***	-0.27***	0.16***	0.16***	0.16***	0.16***	0.16***	0.16***
$A \rightarrow B \leftarrow C, A \rightarrow C$	0.2***	0.15***	0.22***	-0.17***	0.16***	0.02**	0.02**	0.02**	0.02**	0.02**
$A \leftarrow B \leftarrow C, A \rightarrow C$	0.07***	-0.01*	0.16***	-0.03***	0.13***	-0.18***	-0.18***	-0.18***	-0.18***	-0.18***
$A \leftrightarrow B \leftrightarrow C$	-0.16***	-0.77***	-0.31***	-0.17***	-0.14***	-0.05***	-0.05***	-0.05***	-0.05***	-0.05***
$A \leftarrow B \rightarrow C, A \leftrightarrow C$	0.12***	-0.03***	0.08***	0.03***	0.06***	0.07***	0.07***	0.07***	0.07***	0.07***
$A \rightarrow B \leftarrow C, A \leftrightarrow C$	0.06***	0.13***	0.11***	-0.16***	0.13***	0.16***	0.16***	0.16***	0.16***	0.16***
$A \rightarrow B \rightarrow C, A \leftrightarrow C$	0.1***	-0.08***	0.13***	-0.04***	0.18***	0.03***	0.03***	0.03***	0.03***	0.03***
$A \rightarrow B \leftrightarrow C, A \leftrightarrow C$	0.05***	-0.17***	0.06***	-0.03***	0.11***	0.0	0.0	0.0	0.0	0.0
$A \leftrightarrow B \leftrightarrow C, A \leftrightarrow C$	-0.02**	-0.15***	-0.05***	-0.01	0.02***	-0.03***	-0.03***	-0.03***	-0.03***	-0.03***

Note: Significance of correlation test of Pearson: "\*\*\*\*" - p-value < 0.01, "\*\*\*" - p-value < 0.05, "\*\*" - p-value < 0.1, "" - non significant.

TABLE 4.2: Correlation matrix (con't)

	$A \leftrightarrow B \leftrightarrow C$	$A \leftarrow B \rightarrow C$	$A \rightarrow B \rightarrow C$	$A \rightarrow B \leftarrow C$	$A \rightarrow B \leftrightarrow C$	$A \leftrightarrow B \leftrightarrow C$	$A \leftrightarrow B \leftrightarrow C$
#Post	-0.06***	-0.02**	-0.04***	-0.02***	-0.02***	-0.02***	-0.01
#Author	-0.21***	-0.03***	-0.04***	-0.03***	-0.07***	-0.07***	-0.05***
#Root Answers	-0.25***	-0.05***	-0.05***	-0.05***	-0.1***	-0.05***	-0.06***
#Leaf	-0.17***	-0.02**	-0.03***	-0.03***	-0.05***	-0.05***	-0.03***
Width	-0.19***	-0.03***	-0.04***	-0.03***	-0.06***	-0.06***	-0.04***
Level Width	0.2***	0.05***	-0.01*	0.04***	0.13***	0.13***	0.1***
Height	0.2***	0.02**	-0.02***	0.0	0.07***	0.07***	0.06***
Modularity	-0.04***	-0.02***	0.05***	0.02**	-0.02***	-0.02***	-0.03***
Av. Sackin	0.52***	0.01**	-0.05***	0.03***	0.15***	0.15***	0.13***
Av. Age Author (Month)	0.0	0.01**	-0.06***	-0.02**	-0.05***	-0.05***	-0.03***
OP Age (Month)	0.07***	0.01	-0.04***	-0.01	-0.03***	-0.03***	-0.01
Av. Experience	0.08***	0.0	-0.03***	-0.0	-0.01*	-0.01*	-0.0
OP Experience	0.08***	0.0	-0.02***	0.01	-0.0	-0.0	0.01**
#Delta	0.06***	-0.01	-0.06***	0.0	-0.02**	-0.02**	0.01
#Delta from OP	0.11***	-0.01*	-0.07***	0.0	-0.01*	-0.01*	0.01
Duration (Hour)	-0.08***	0.0	-0.0	0.0	-0.01	-0.01	-0.01
$A \leftarrow B \rightarrow C$	-0.16***	0.12***	0.06***	0.1***	0.05***	0.05***	-0.02**
$A \rightarrow B \leftarrow C$	-0.77***	-0.03***	0.13***	-0.08***	-0.17***	-0.17***	-0.15***
$A \rightarrow B \rightarrow C$	-0.31***	0.08***	0.11***	0.13***	0.06***	0.06***	-0.05***
$A \leftrightarrow B \leftarrow C$	-0.17***	0.03***	-0.16***	-0.04***	-0.03***	-0.03***	-0.01
$A \leftrightarrow B \rightarrow C$	-0.14***	0.06***	0.13***	0.18***	0.11***	0.11***	0.02***
$A \rightarrow B \leftarrow C, A \rightarrow C$	-0.18***	0.07***	0.16***	0.03***	0.0	0.0	-0.03***
$A \leftarrow B \leftarrow C, A \rightarrow C$	-0.05***	0.0	0.01	0.04***	0.0	0.0	-0.01
$A \leftrightarrow B \leftrightarrow C$	-0.06***	-0.06***	-0.13***	-0.03***	0.05***	0.05***	0.09***
$A \leftarrow B \rightarrow C, A \leftrightarrow C$	-0.06***	0.04***	0.04***	0.02**	0.04***	0.04***	-0.0
$A \rightarrow B \leftarrow C, A \leftrightarrow C$	-0.13***	0.02**	0.01	0.01	0.01	0.01	-0.02**
$A \rightarrow B \rightarrow C, A \leftrightarrow C$	-0.03***	0.02**	0.01	0.04***	0.04***	0.04***	0.03***
$A \rightarrow B \leftrightarrow C, A \leftrightarrow C$	0.05***	0.04***	0.01	0.04***	0.05***	0.05***	0.05***
$A \leftrightarrow B \leftrightarrow C, A \leftrightarrow C$	0.09***	-0.0	-0.02**	0.03***	0.05***	0.05***	0.05***

Note: Significance of correlation test of Pearson: "\*\*\*\*" - pvalue < 0.01, "\*\*\*" - pvalue < 0.05, "\*\*" - pvalue < 0.1, "\*" - non significant.

TABLE 4.2: Correlation matrix (con't)

with the role  $B$  of the three main motifs:  $a \rightarrow *B* \leftarrow c$ ,  $a \leftrightarrow *B* \leftarrow c$  and  $a \leftrightarrow *B* \leftrightarrow c$ .

Finally, analyzing correlation among role, we observe that some roles are mutually exclusive such as  $*A* \rightarrow b \leftarrow *C*$  with  $*A* \leftrightarrow b \leftarrow c$  or with  $*A* \leftrightarrow b \leftrightarrow *C*$ .

The following and last part presents the results of the clustering algorithms.

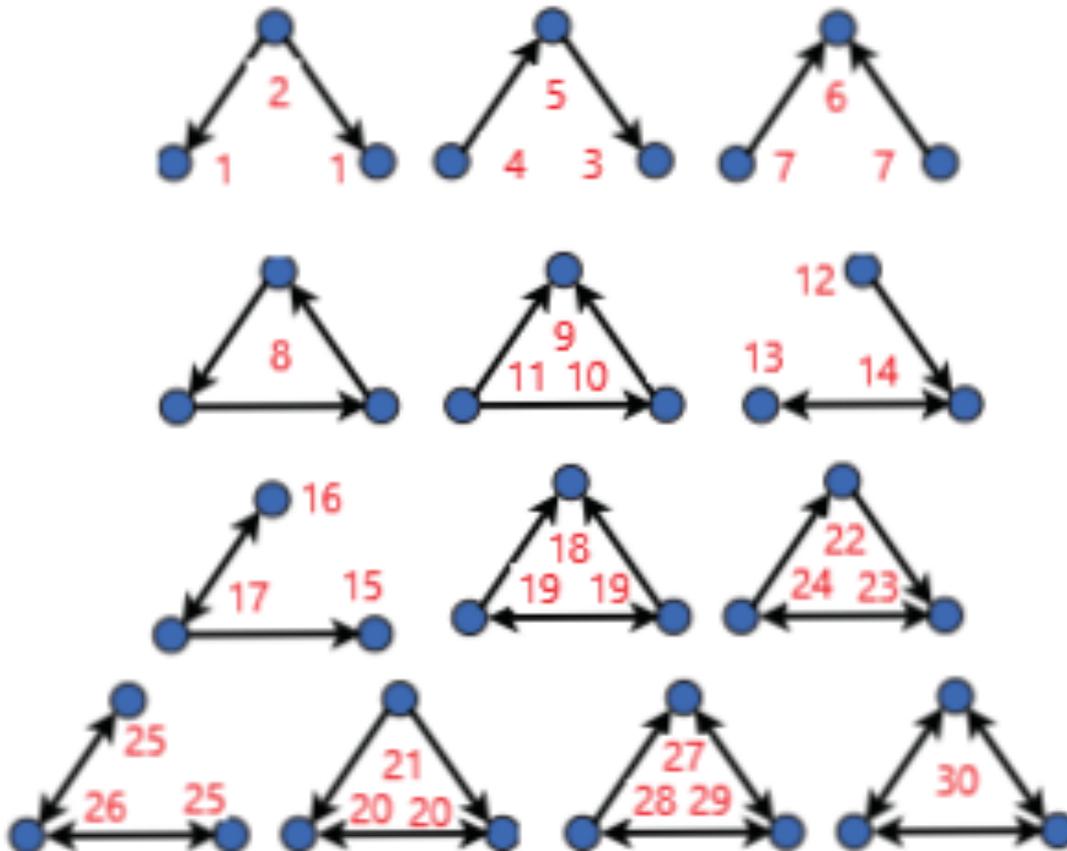


FIGURE 4.5: Role triad dictionary from Cunningham et al. (2013)

## 4.4.2 Clustering analysis

### 4.4.2.1 Optimal clustering

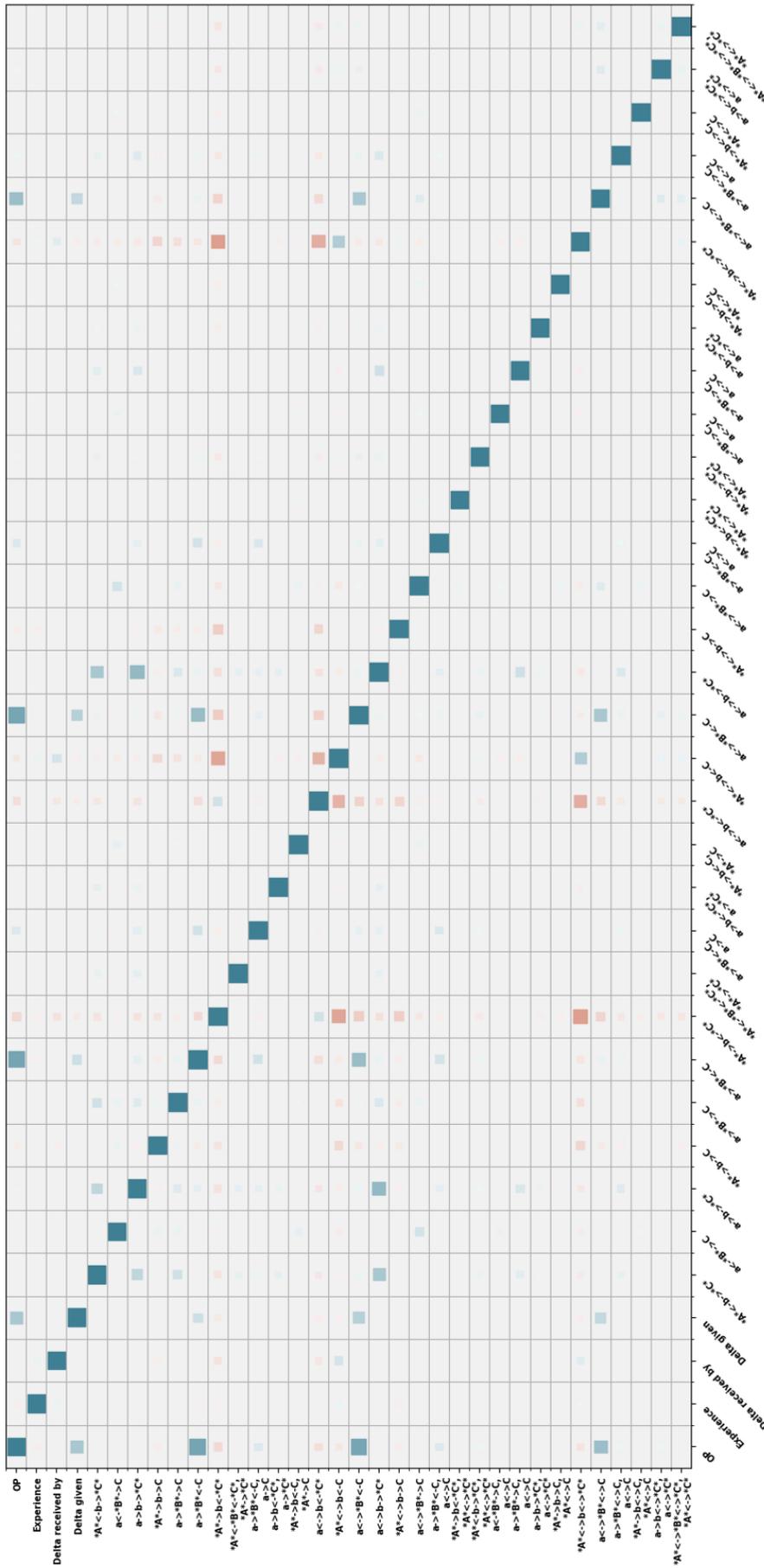
Researchers have to settle the number of clusters they are looking for before running the algorithms presented above. But thanks to the Silhouette criterion described in Section 4.3.2, we are able to select the optimum number of

OP	Experience	$\Delta$ received by	$A \leftarrow B \rightarrow C$	$A \leftarrow B \rightarrow C$	$A \rightarrow B \rightarrow C$										
OP															
Experience	-0.05***	-0.03***	0.43***	-0.03***	-0.06***	-0.02***	0.71***	-0.17***	0.15***	-0.02***	-0.02***	-0.14***	-0.09***		
$\Delta$ received by	0.07***	0.07***	-0.02***	-0.02***	0.01***	-0.04***	-0.04***	-0.04***	-0.01***	-0.03***	-0.03***	0.01***	0.07***		
$\Delta$ given	-0.01***	-0.01***	-0.01***	-0.01***	-0.05***	0.01***	0.01***	-0.12***	0.00**	-0.01***	-0.01***	-0.1***	0.17***		
$A \leftarrow B \rightarrow C$	-0.01***	-0.01***	-0.01***	-0.01***	-0.02***	0.04***	0.23***	-0.04***	0.00**	0.02***	-0.06***	-0.06***	-0.04***		
$A \rightarrow B \rightarrow C$	-0.02***	-0.02***	-0.01***	-0.01***	0.08***	-0.02***	0.04***	-0.04***	0.06***	0.01***	-0.09***	-0.09***	-0.05***		
$A \rightarrow B \rightarrow C$	-0.02***	-0.02***	-0.01***	-0.01***	0.08***	-0.02***	0.11***	-0.12***	0.11***	0.08***	0.02***	-0.11***	-0.06***		
$A \rightarrow B \rightarrow C$	0.01***	0.01***	0.01***	0.01***	-0.04***	0.06***	0.06***	-0.11***	-0.02**	-0.01***	-0.03***	-0.03***	-0.18***		
$A \rightarrow B \rightarrow C$	-0.03***	-0.03***	0.0	0.08***	0.06***	0.13***	0.08***	-0.06***	0.03***	0.03***	0.05***	-0.03***	-0.12***		
$A \rightarrow B \rightarrow C$	-0.04***	-0.04***	0.0	-0.02***	-0.06***	0.08***	0.01***	-0.17***	0.21***	-0.01***	-0.01***	-0.14***	-0.07***		
$A \rightarrow B \rightarrow C$	-0.04***	-0.04***	-0.08***	-0.04***	-0.11***	-0.06***	-0.17***	-0.02***	0.03***	0.02***	0.03***	0.2***	-0.49***		
$A \rightarrow B \rightarrow C$	0.0	0.0	-0.0	0.02***	0.0*	0.02***	0.01***	-0.02***	0.01***	0.0	0.0	-0.01***	-0.02***		
$A \rightarrow B \rightarrow C$	0.15***	0.06**	0.02***	-0.01***	-0.02***	0.03***	0.21***	-0.05***	0.01***	0.03***	0.03***	-0.04***	-0.02***		
$A \rightarrow B \rightarrow C$	-0.02***	-0.01***	-0.01***	0.01***	-0.01***	0.03***	0.01***	0.02***	0.03***	0.03***	0.01***	-0.01***	-0.05***		
$A \rightarrow B \rightarrow C$	-0.02***	-0.01***	-0.01***	0.09***	-0.01***	0.05***	-0.01***	0.03***	-0.0***	-0.0***	0.01***	0.01***	-0.04***		
$A \rightarrow B \rightarrow C$	-0.14***	-0.06**	-0.06**	0.02***	-0.03***	-0.03***	-0.14***	0.2***	-0.04***	-0.04***	-0.01***	0.01***	-0.41***		
$A \rightarrow B \rightarrow C$	0.07***	0.04***	-0.04***	-0.06***	-0.18***	-0.12***	0.07***	-0.49***	-0.02***	-0.02***	-0.04***	-0.41***	-0.03***		
$A \rightarrow B \rightarrow C$	0.7***	0.35***	-0.04***	0.07***	-0.09***	0.06***	0.51***	-0.25***	0.09***	0.09***	-0.01***	-0.21***	-0.06***		
$A \rightarrow B \rightarrow C$	0.02***	-0.02***	-0.01***	-0.01***	0.42***	0.15***	0.08***	-0.13***	0.1***	0.1***	-0.01***	-0.12***	-0.06***		
$A \rightarrow B \rightarrow C$	-0.06***	-0.04***	-0.02***	-0.03***	-0.08***	-0.06***	-0.05***	-0.24***	-0.02***	-0.02***	-0.02***	-0.2***	0.01***		
$A \rightarrow B \rightarrow C$	-0.02***	-0.01***	0.01***	0.19***	0.03***	0.08***	-0.03***	-0.09***	-0.01***	0.01***	0.07***	-0.05***	-0.08***		
$A \rightarrow B \rightarrow C$	0.15***	0.05**	0.02***	-0.01***	-0.02***	0.02***	0.2***	-0.05***	0.01***	0.16***	0.02***	-0.04***	-0.02***		
$A \rightarrow B \rightarrow C$	-0.02***	-0.01***	-0.01***	0.03***	-0.02***	0.02***	-0.01***	0.03***	-0.0***	-0.0***	0.02***	-0.02***	-0.03***		
$A \rightarrow B \rightarrow C$	0.04**	0.01***	0.02***	-0.01***	-0.03***	0.01***	0.06***	-0.08***	0.05***	0.05***	0.0***	-0.07***	0.06***		
$A \rightarrow B \rightarrow C$	-0.02***	-0.01***	-0.01***	0.08***	-0.01***	0.03***	-0.01***	0.02***	-0.0***	-0.0***	0.01***	0.04***	-0.04***		
$A \rightarrow B \rightarrow C$	-0.01***	-0.01***	-0.01***	0.01***	-0.01***	0.02***	0.0***	-0.02***	0.02***	0.02***	0.01***	0.0	-0.05***		
$A \rightarrow B \rightarrow C$	0.04***	0.01***	0.02***	-0.0***	-0.02***	0.01***	0.04***	-0.05***	0.02***	0.02***	0.0***	-0.04***	0.0***		
$A \rightarrow B \rightarrow C$	0.01***	0.02***	0.01***	0.05***	-0.01***	0.03***	-0.01***	-0.05***	-0.0	0.0	0.01***	-0.03***	0.03***		
$A \rightarrow B \rightarrow C$	0.04***	0.11***	-0.05***	-0.07***	-0.14***	-0.14***	-0.1***	-0.54***	-0.03***	-0.05***	-0.05***	-0.44***	0.38***		
$A \rightarrow B \rightarrow C$	0.49***	-0.01***	0.28***	-0.01***	-0.07***	-0.02***	0.08***	-0.2***	0.01***	-0.02***	-0.01***	-0.17***	-0.01***		
$A \rightarrow B \rightarrow C$	0.05***	0.01***	0.01***	-0.01***	-0.04***	0.0	0.06***	-0.1***	0.01***	0.06***	-0.01***	-0.08***	0.03***		
$A \rightarrow B \rightarrow C$	-0.01***	0.0	-0.0	0.05***	-0.02***	0.03***	-0.02***	-0.04***	-0.0	-0.01***	0.01***	-0.02***	0.01***		
$A \rightarrow B \rightarrow C$	0.05***	0.01***	0.03***	-0.01***	-0.03***	-0.01***	0.01***	-0.09***	-0.0	-0.01***	-0.01***	-0.08***	0.07***		
$A \rightarrow B \rightarrow C$	0.03***	0.01***	0.02***	-0.01***	-0.04***	-0.01***	-0.0	-0.1***	-0.0***	-0.0***	-0.01***	-0.09***	0.07***		

Note: Significance of correlation test of Pearson: "\*" - pvalue < 0.05, "\*\*" - pvalue < 0.01, "\*\*\*" - pvalue < 0.001, "ns" - non significant.

TABLE 4.3: Role correlation matrix





Note: The more the coefficient is close to 1, the more the color tends toward a deep blue ; the more the coefficient is close to -1, the more the color tends toward a deep red.

FIGURE 4.6: Role correlation heatmap

clusters per algorithm. Then, we run each algorithm from 2 to 11 clusters, and look at the maximum *Silhouette* coefficients. Solely for the hierarchical clustering algorithm — HCA, we just decide with the corresponding dendrogram. Figure 4.7 presents the *Silhouette* coefficients for the  $k$ -means, Spectral and Ward algorithms, plus the HCA dendrogram.

On the one hand, the  $k$ -means and Spectral algorithms reach the maximum *Silhouette* coefficient with 3 clusters. On the other hand, HCA and Ward (being a specification of one parameter of HCA) reach the optimum for 2 clusters. Nevertheless, as the difference of the coefficient between 2 and 3, for those last two algorithms, is low, we finally select the optimal number of clusters to 3.

Then, we run again the  $k = 3$   $k$ -means, Spectral and Ward algorithms and label each thread with the result. Then, they allocate trees in the database into 3 clusters, as defined in Section 4.3.1. But do they define the same clusters? To answer this question, we calculate the Adjusted Rand Index to compare the results. The Rand Index, developed by Rand (1971), computes a similarity measure between two clustering algorithms by considering all pairs of samples and counting pairs that are assigned in the same or different clusters in the predicted and true clustering. We use the Adjusted Rand Index in order to have a readable indicator. If the coefficient is close to 0, the clustering is made randomly, independently of the number of clusters and samples. Otherwise, if the coefficient is close to 1, the clustering algorithms are identical. Table 4.4 presents the results of the Adjusted Rand Index. The coefficients are really close to 1, indicating a robust clustering among the database. Nevertheless, we look at the cluster distribution, presented in Table 4.5. The results are really close for clusters 2 and 3 and are identical for cluster 1. For the 85 threads which are labeled in different clusters, we follow the result of the majority of algorithms. The final clustering distribution is as follows: cluster 1 groups 3,334 threads, cluster 2 groups 10,293 threads and cluster 3 groups 6,329. The next section analyzes the differences between the clusters.

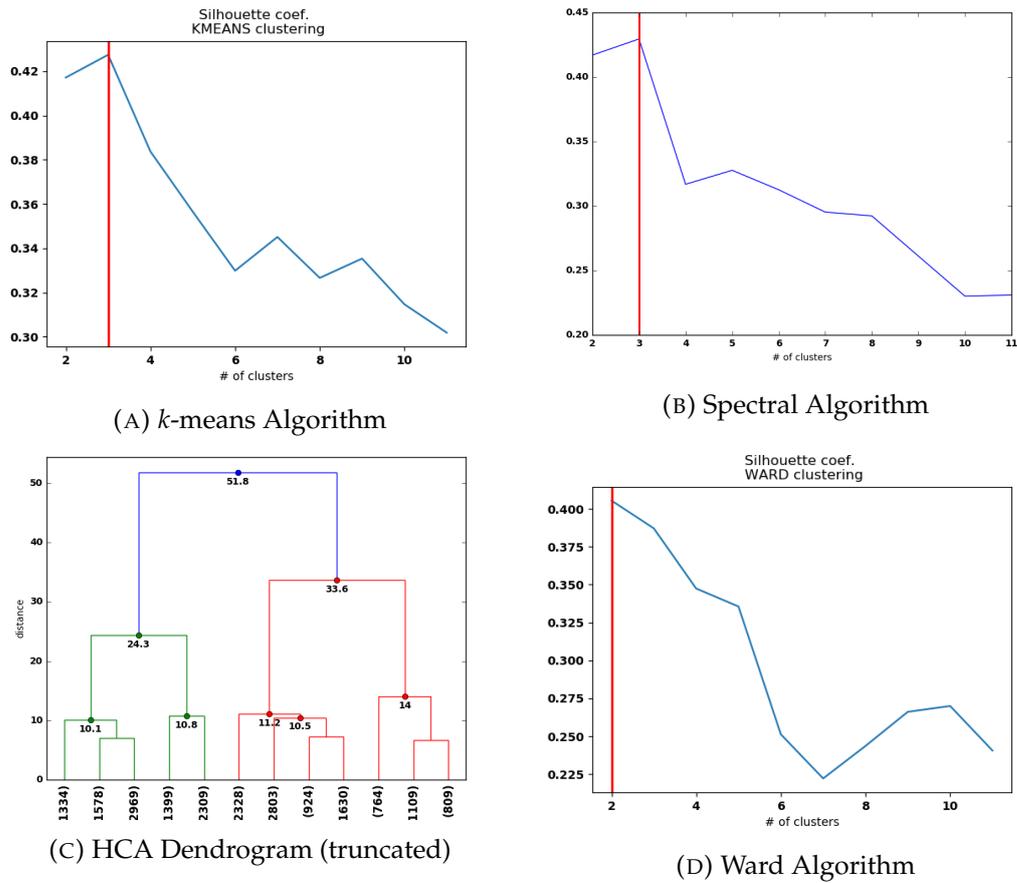


FIGURE 4.7: Optimal number of clusters

	<i>k</i> -means	Spectral	Ward
<i>k</i> -means	N.A.	0.9876	0.9852
Spectral	0.9876	N.A.	0.9976
Ward	0.9852	0.9976	N.A.

TABLE 4.4: Adjust Rand Index

	Cluster 1	Cluster 2	Cluster 3
<i>k</i> -means	3334	10222	6400
Spectral	3334	10293	6329
Ward	3334	10307	6315

P.S.: labels among algorithms have been matched.

TABLE 4.5: Cluster distribution, per algorithm

#### 4.4.2.2 Clustering results

Once the groups have been defined by the algorithm, we want to understand the difference between the clusters. As a reminder, the clustering algorithms solely used as an input the distribution, in percent, of motifs in each thread. Table 4.6 presents the distribution of motifs, per cluster. As one might expect, the motifs  $A \rightarrow B \leftarrow C$ ,  $A \leftrightarrow B \leftarrow C$  and  $A \leftrightarrow B \leftrightarrow C$  count for 90% of the distribution, in each cluster. Similar results have been found in Ashford et al. (2019). Nevertheless, the distribution of these motifs is completely different from one cluster to another.

**Cluster 1** mainly consists of the motif  $A \leftrightarrow B \leftrightarrow C$ , for more than 2/3 of the distribution, and nearly 25% of the distribution consists of the motif  $A \leftrightarrow B \leftarrow C$ . **Cluster 2** motifs' distribution is more balanced. Nearly half of the motifs are  $A \leftrightarrow B \leftarrow C$ , when the rest is equally distributed between  $A \rightarrow B \leftarrow C$  and  $A \leftrightarrow B \leftrightarrow C$ . Finally, **Cluster 3** is mainly characterized by  $A \rightarrow B \leftarrow C$  (59%) and by  $A \leftrightarrow B \leftarrow C$  (24%). We test the difference in distribution thanks to a Student's t-test, presenting the results in Table 4.7.

We interpret the results in assuming that each cluster corresponds with a level of activated *sub-discussions*. We define as a sub-discussion each branch of a tree, corresponding to a new challenging author answering the OP root post. However, analyzing only a snapshot of the database, we must be cautious about the extrapolation of our results.

One of the reason why we observe this would be that Cluster 3 groups threads in which discussions did not take off. On the contrary, Cluster 1 groups discussions where actors are engaged to answer each other. And finally, Cluster 2 lies in between, with threads in which only a part of sub-discussions have been developed. But, why a discussion would not attract enough individuals? We offer several interpretations. Either the discussion is not interesting enough to attract people or the question is *too easy*, and the resolution arises quickly, with few answers or, finally, it might be due to the snapshot of the database itself. Indeed, maybe the snapshot was taken

too early in the *life* of the discussion, before an intense exchange among authors begin. We checked this last hypothesis, analyzing the distribution in timestamp of the *last post* of each thread, per cluster. Indeed, if all threads from Cluster 3 ended (at the time of the snapshot) during the last month of the time frame, we would not be able to affirm that those discussions did not interest enough. Analyzing Figure 4.8, we observe a similar distribution among clusters, matching with the evolution detailed in Section 3.4.2. Then, if we assume, that each cluster corresponds to a different *community*, we can affirm that these communities follow the same evolution step, and that differences between clusters do not come from the fact that the database is a snapshot.

We now wonder if the clustering reveals other differences between threads. In order to do so, we analyze how the tree characteristics differ among clusters. We present the results in Table 4.8.

First of all, except for the variable *Av. Age Author*, Cluster 2 always lies in between the two other clusters. Thus, for the analysis, we focus on Cluster 1 and Cluster 3. Concerning the size of the trees (variables *Post*, *Author*, *Root answers*, *Leaf*) and the time of the discussions (*Duration*), Cluster 1 trees attract fewer people and last less, in time, than Cluster 3 trees. In average, Cluster 3 trees have twice more *sub-discussions*, defined by the variable *Root Answers*, than ones in Cluster 1. Then, we assume that the debate is more intense in Cluster 3, which attracts more authors, who post more.

Second, looking at the shape (variables *Width*, *Level width*, *Height*, *Av. Sackin*), Cluster 3 trees are wider in average (noticed that *Modularity* is equivalent among clusters). But, other characteristics are more important in Cluster 1. Trees from this cluster are higher (which implies a higher level for its width and a higher average Sackin index). We assume that, on the one hand, *Width* is a proxy of the number of sub-discussions which are open, where *Height* is a proxy for the *quality* of the discussions, in the sense that we assume that if a branch of the tree is higher, it implies more exchanges among participants, whereas more branches (*i.e.* a wider tree) highlights the opening of sub-discussions, but not if they have been solved. Thus Cluster 1 attracts

	Cluster 1 $N = 3334$	Cluster 2 $N = 10222$	Cluster 3 $N = 6400$
$A \leftarrow B \rightarrow C$	0.00	0.01	0.01
$A \rightarrow B \leftarrow C$	<b>0.02</b>	<b>0.22</b>	<b>0.59</b>
$A \rightarrow B \rightarrow C$	0.01	0.04	0.05
$A \leftrightarrow B \leftarrow C$	<b>0.23</b>	<b>0.45</b>	<b>0.24</b>
$A \leftrightarrow B \rightarrow C$	0.03	0.04	0.04
$A \rightarrow B \leftarrow C, A \rightarrow C$	0.00	0.00	0.00
$A \leftarrow B \leftarrow C, A \rightarrow C$	0.00	0.00	0.00
$A \leftrightarrow B \leftrightarrow C$	<b>0.68</b>	<b>0.22</b>	<b>0.05</b>
$A \leftarrow B \rightarrow C, A \leftrightarrow C$	0.00	0.00	0.00
$A \rightarrow B \leftarrow C, A \leftrightarrow C$	0.00	0.00	0.00
$A \rightarrow B \rightarrow C, A \leftrightarrow C$	0.00	0.00	0.00
$A \rightarrow B \leftrightarrow C, A \leftrightarrow C$	0.01	0.01	0.00
$A \leftrightarrow B \leftrightarrow C, A \leftrightarrow C$	0.01	0.00	0.00

Note: columns does not add up to 1 due to rounding function. Nevertheless, motif shares below 0.01 are negligible.

TABLE 4.6: Motif distribution, per cluster

more intense and developed discussions than Cluster 3.

Finally, concerning the authors, we observe that there is no difference between the clusters about the age of the authors (*Av. Age Author*). But we observe small differences for the other characteristics (variables *OP age*, *Av. Experience*, *OP Experience*). In Cluster 1, OPs are slightly older and more experienced. The whole population is slightly more expert as well. Moreover, the difference between clusters does not hold concerning the  $\Delta$  system (variables *Delta*, *Delta from OP*).

## 4.5 Conclusion

The current work tackles the issue of coordination in a open Q&A forum: Reddit - Change My View. We analyze the different characteristics of online discussions in order to better understand how online exchanges are conducted. First, we model the discussion through a network framework. We focus on the "author network" where two authors are connected if one answered to the other. From those networks, we extracted sub-networks called motifs. Applying different clustering algorithms on the motifs distribution, two extreme collaborative resolution emerged, a third one appears to be in

	C1 vs C2	C1 vs C3	C2 vs C3
$A \leftarrow B \rightarrow C$	0.0***	0.0***	0.06*
$A \rightarrow B \leftarrow C$	0.0***	0.0***	0.0***
$A \rightarrow B \rightarrow C$	0.0***	0.0***	0.0***
$A \leftrightarrow B \leftarrow C$	0.0***	0.01***	0.0***
$A \leftrightarrow B \rightarrow C$	0.0***	0.0***	0.29
$A \rightarrow B \leftarrow C, A \rightarrow C$	0.0***	0.0***	0.0***
$A \leftarrow B \leftarrow C, A \rightarrow C$	0.0***	0.03**	0.0***
$A \leftrightarrow B \leftrightarrow C$	0.0***	0.0***	0.0***
$A \leftarrow B \rightarrow C, A \leftrightarrow C$	0.0***	0.0***	0.0***
$A \rightarrow B \leftarrow C, A \leftrightarrow C$	0.0***	0.0***	0.0***
$A \rightarrow B \rightarrow C, A \leftrightarrow C$	0.0***	0.04**	0.0***
$A \rightarrow B \leftrightarrow C, A \leftrightarrow C$	0.44	0.0***	0.0***
$A \leftrightarrow B \leftrightarrow C, A \leftrightarrow C$	0.0***	0.0***	0.0***

Note: Significance: "\*\*\*\*" - pvalue < 0.01, "\*\*\*" - pvalue < 0.05, "\*\*" - pvalue < 0.1, "" - non significant.

TABLE 4.7: Student's t-test on percent of motif per cluster

between. Figures 4.10, 4.11 and 4.12 present typical trees and corresponding author networks from each cluster. This clustering, based on motifs solely, meaning on sub-graph shape, highlights differences on all characteristics of a discussion.

The "Chatters" (cluster 1) gather discussions where OP answers to almost all his challenging authors. Indeed those discussions attract few authors and last a short time. Besides, trees (discussion shape) are high but not really wide. And authors are old and experienced in CMV. We conclude that discussions within the "Chatters" are dense, challenging and any new subject is addressed by the OP.

The "Overflowed" (cluster 3) gather discussions with opposed characteristics to those from the "Chatters". Indeed, discussions attract numerous challenging authors, last a long time and OP answers to a little proportion of his challengers. Furthermore, trees are less high but really wide. Finally authors have the same age than those from "Chatters" but less experienced. We conclude that those discussions escape from the lead of the OP, who is overflowed by the number of challengers. Numerous sub-topic are not addressed, global discussion is less dense.

	Cluster	Count	Mean	Std	Min	25%	Median	75%	Max
#Post	1	3334	57.24	52.74	10.00	25.00	44.00	71.00	1043.00
	2	10293	72.13	87.36	10.00	27.00	47.00	80.00	1776.00
	3	6329	78.89	108.24	10.00	23.00	41.00	84.00	1432.00
#Author	1	3334	14.49	11.30	3.00	8.00	12.00	18.00	250.00
	2	10293	25.10	26.65	3.00	12.00	18.00	28.00	397.00
	3	6329	36.82	43.92	4.00	14.00	22.00	39.00	774.00
#Root Answers	1	3334	8.30	5.50	1.00	5.00	7.00	10.00	81.00
	2	10293	11.87	8.68	1.00	6.00	10.00	14.00	124.00
	3	6329	18.40	16.39	2.00	9.00	13.00	21.00	216.00
#Leaf	1	3334	15.88	15.74	1.00	7.00	12.00	20.00	367.00
	2	10293	27.01	33.50	1.00	11.00	17.00	30.00	640.00
	3	6329	37.04	48.97	2.00	12.00	20.00	39.00	874.00
Width	1	3334	10.63	7.97	1.00	6.00	9.00	13.00	140.00
	2	10293	14.87	13.16	1.00	8.00	11.00	17.00	245.00
	3	6329	19.94	18.69	2.00	9.00	14.00	23.00	263.00
Level Width	1	3334	2.09	1.31	0.00	1.00	2.00	3.00	10.00
	2	10293	1.85	1.18	0.00	1.00	1.00	3.00	10.00
	3	6329	1.35	0.85	1.00	1.00	1.00	1.00	9.00
Height	1	3334	8.62	2.01	2.00	7.00	10.00	10.00	10.00
	2	10293	8.22	2.18	2.00	7.00	9.00	10.00	10.00
	3	6329	7.40	2.62	1.00	5.00	8.00	10.00	10.00
Modularity	1	3334	0.06	0.04	0.00	0.03	0.05	0.08	0.20
	2	10293	0.05	0.04	0.00	0.02	0.04	0.07	0.20
	3	6329	0.06	0.05	0.00	0.02	0.05	0.09	0.20
Av. Sackin	1	3334	4.70	1.29	1.80	3.75	4.64	5.57	10.00
	2	10293	3.79	1.08	1.29	3.00	3.71	4.54	10.00
	3	6329	2.87	1.00	1.00	2.11	2.75	3.55	9.00
Av. Age Author (Month)	1	3334	8.44	5.93	0.00	3.91	7.68	11.78	35.92
	2	10293	9.01	5.70	0.00	4.37	8.55	12.81	34.12
	3	6329	8.32	5.58	0.00	3.68	7.33	12.49	30.45
OP Age (Month)	1	3334	5.15	8.08	0.00	0.00	1.00	7.00	43.00
	2	10293	4.84	8.13	0.00	0.00	0.00	6.00	45.00
	3	6329	3.60	7.04	0.00	0.00	0.00	4.00	45.00
Av. Experience	1	3334	12.65	15.78	0.00	4.20	8.08	14.99	175.40
	2	10293	11.64	12.08	0.00	4.31	8.12	14.57	136.50
	3	6329	9.80	8.99	0.00	3.91	7.31	12.59	111.17
OP Experience	1	3334	4.98	25.55	0.00	0.00	0.00	2.00	257.00
	2	10290.0	2.83	17.83	0.00	0.00	0.00	1.00	257.00
	3	6321.0	1.08	9.12	0.00	0.00	0.00	0.00	232.00
#Delta	1	3334	0.68	1.23	0.00	0.00	0.00	1.00	18.00
	2	10293	0.66	1.05	0.00	0.00	0.00	1.00	11.00
	3	6329	0.49	0.92	0.00	0.00	0.00	1.00	17.00
#Delta from OP	1	3334	0.64	1.20	0.00	0.00	0.00	1.00	18.00
	2	10293	0.56	0.97	0.00	0.00	0.00	1.00	11.00
	3	6329	0.31	0.71	0.00	0.00	0.00	0.00	17.00
Duration (Hour)	1	3334	160.54	478.50	0.80	23.39	42.62	84.64	4310.61
	2	10293	266.80	699.63	0.90	28.01	52.45	114.69	9578.17
	3	6329	338.49	812.52	0.95	29.43	53.66	130.29	6888.93

TABLE 4.8: Descriptive statistics, per cluster

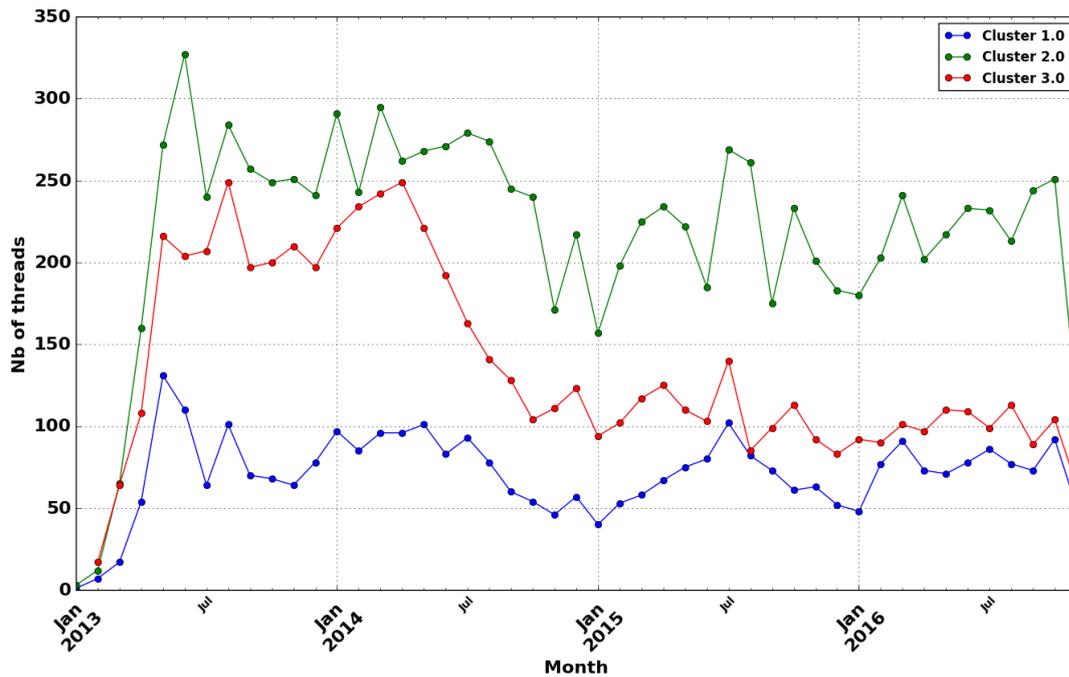


FIGURE 4.8: Last post date frequency, per cluster

The "Orators" (cluster 2) gather discussions being inbetween the previous ones. It gathers more than the half of the discussions in study, and discussions characteristics always fall between discussions characteristics of the "Chatters" and the "Overflowed". Thus we conclude that we lack of informations to properly distinguish behaviors among this cluster, which gathers heterogeneous discussions.

The use of motifs in the author networks of online discussions is an innovative methodology to extract information on the structure of discussions which leads toward a better comprehension of online conversations evolution. This analysis is a first work toward a monitoring tool for CMV moderators. If they aim for dense discussion, where each argument is tackle, then they could settle a maximum number of challenging author, in order to stay close to a "Chatter" discussion. In this case, the OP would be able to adress all the subject. Finally, further work should focus on the power of prediction of motifs distribution in order to anticipate discussion evolution, and here again, help monitoring the forum.

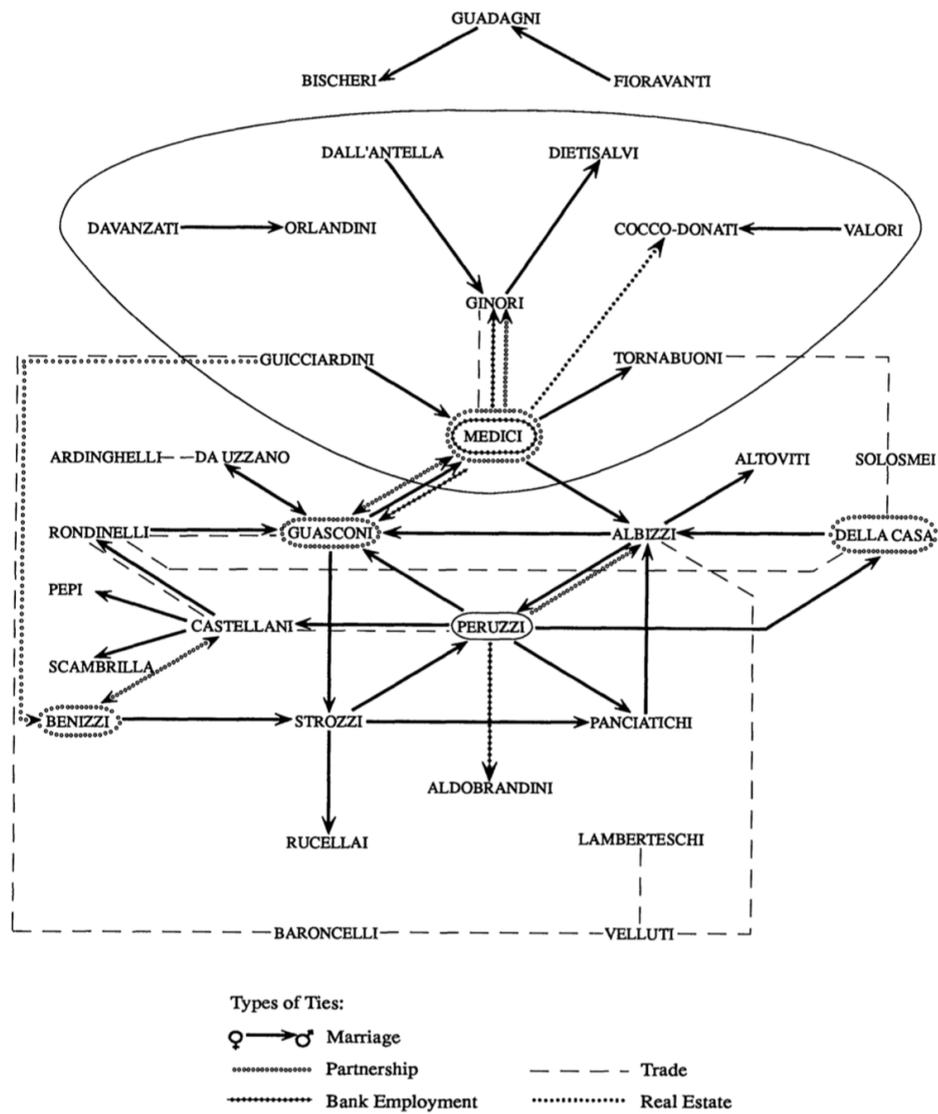
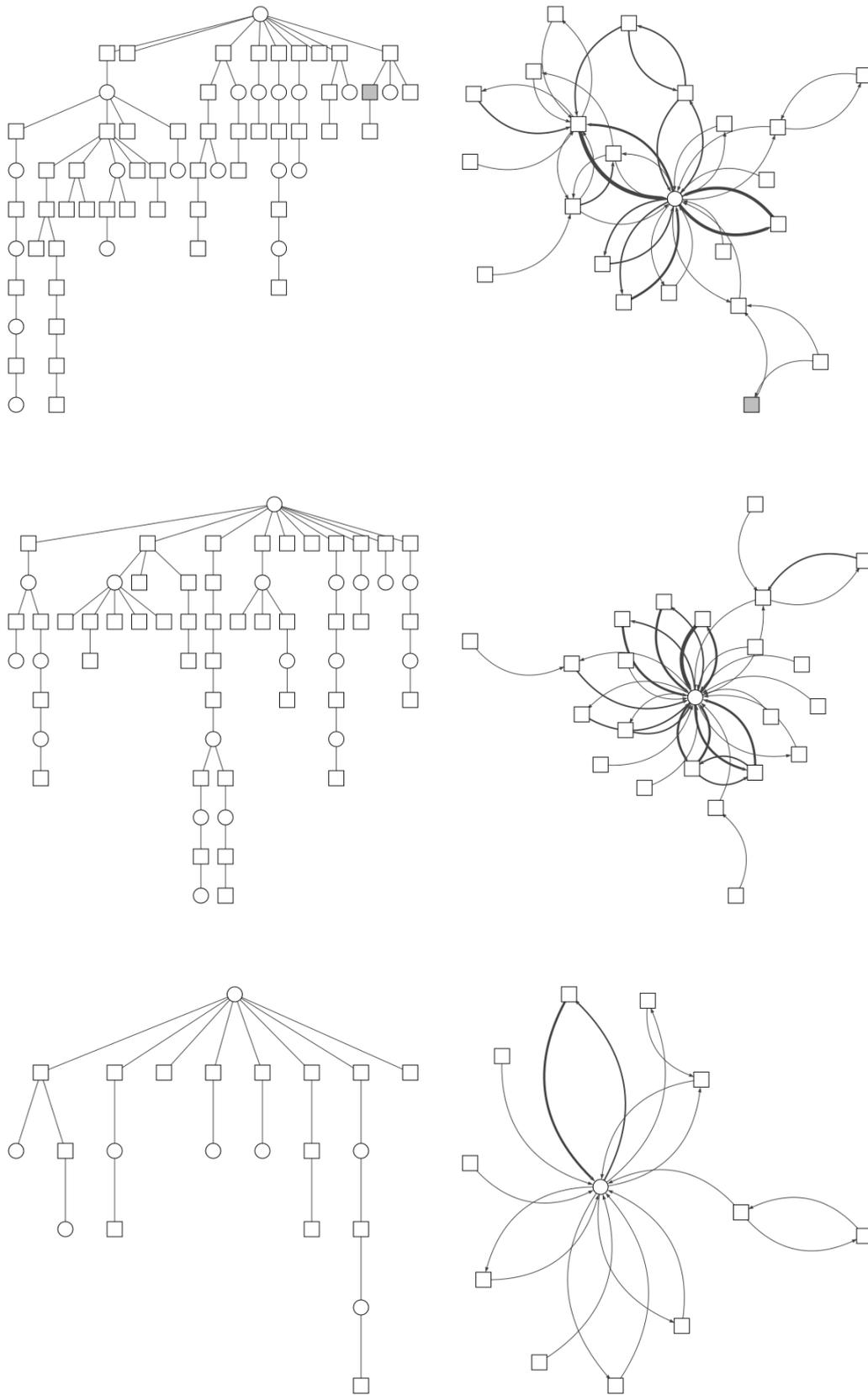


FIGURE 4.9: Marriage and economic network Padgett and Ansell, 1993

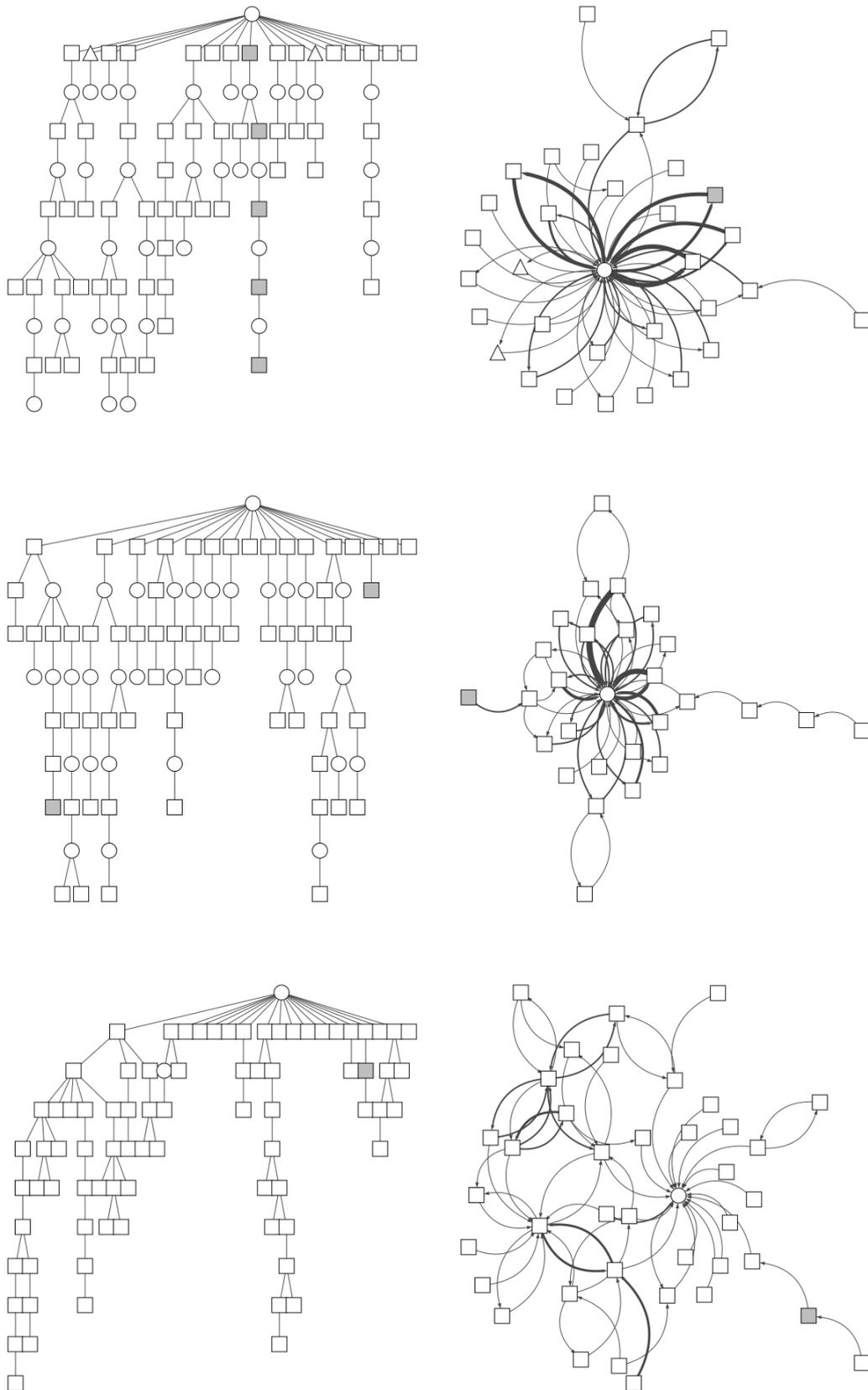


(A) Post network

(B) Author network

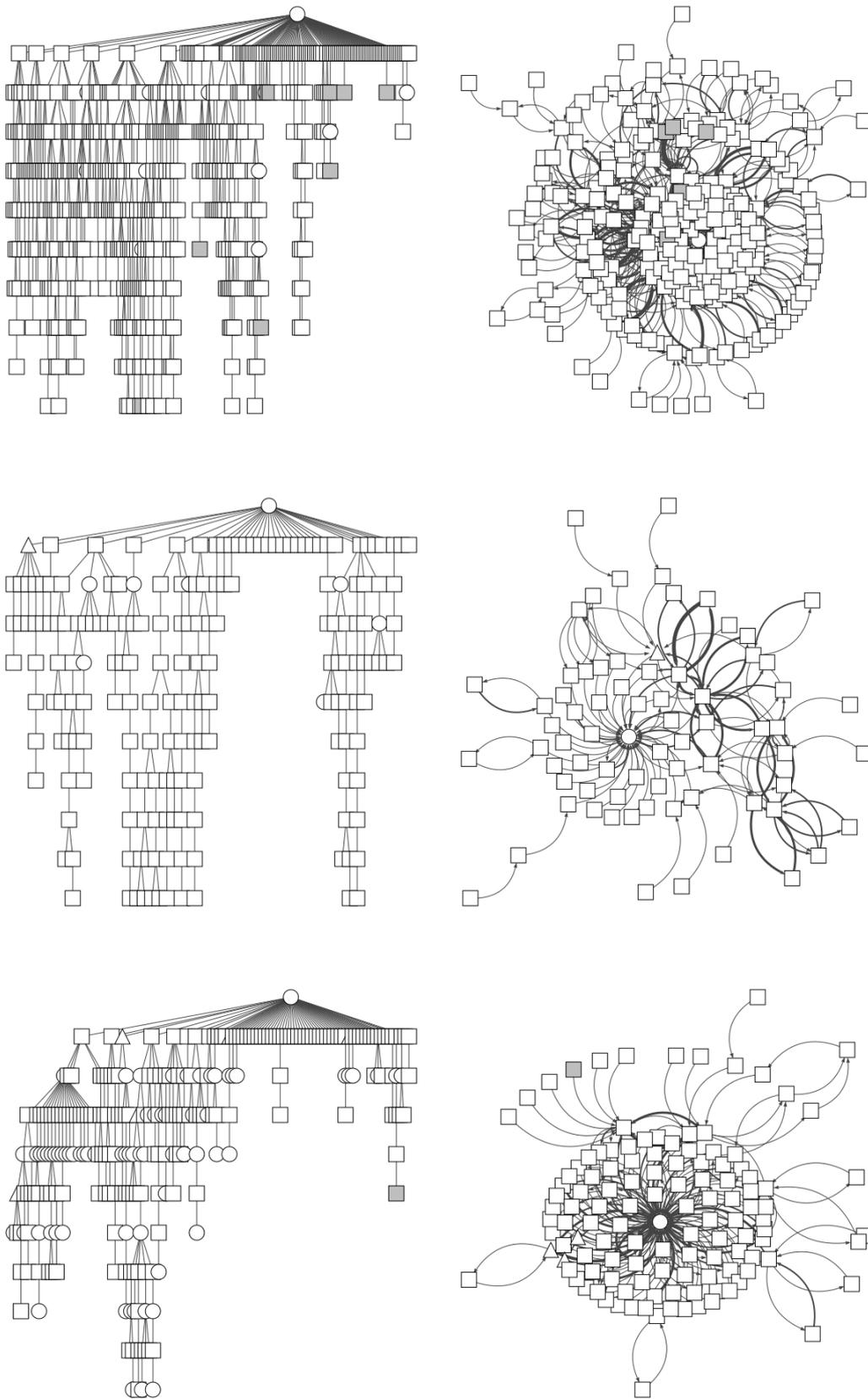
Note:  $\circ = OP$ ,  $\square = non-\Delta$ ,  $\triangle = \Delta$

FIGURE 4.10: Network modeling Cluster 1



(A) Post network (B) Author network  
 Note:  $\circ = OP$ ,  $\square = non-\Delta$ ,  $\triangle = \Delta$

FIGURE 4.11: Network modeling Cluster 2



(A) Post network

(B) Author network

Note:  $\circ = OP$ ,  $\square = non-\Delta$ ,  $\triangle = \Delta$

FIGURE 4.12: Network modeling Cluster 3



# Chapter 5: The Consent of the Crowd

This chapter is an update of Mano, Dalle, and Tomasik (2018).

5.1	Who: Rewarded discussions . . . . .	114
5.2	Why: Individual vs. crowd . . . . .	115
5.3	How: Econometrics . . . . .	117
5.3.1	Reddit – Change My View (CMV) . . . . .	117
5.3.2	First delta context . . . . .	119
5.4	What: The Consent of the Crowd . . . . .	124
5.4.1	Descriptive statistics . . . . .	124
5.4.2	Regression modelling . . . . .	133
5.4.3	Result discussion . . . . .	135
5.4.3.1	Results . . . . .	135
5.4.3.2	Discussion . . . . .	140
5.4.4	Conclusion . . . . .	143

## 5.1 Who: Rewarded discussions

People commonly exchange messages online — from the 1980s and the first *newsgroups* to today's online fora, social networks, chats, etc. The biggest sites where online discussions take place, such as StackOverflow, gather millions of threads and users, calling attention to the importance of such platforms in social life. And, even after the expansion of social networks, still 15% of Internet users in the U.S. exchange actively on fora (Duggan et al., 2015). Basically, online discussions are active on any kind of subject, from politics to video games, about finding a technical answer to a problem through a Q&A site or in order to seek help from fellow students in Massive Open Online Courses (MOOC) fora. Fora are often used to discuss and challenge opinions (Liang, 2014), such as in the sub-Reddit *Change My View* (CMV), which is dedicated to individuals who are looking to challenge and potentially change their view on a given topic.

In parallel, much attention has recently been devoted to what has been called the *Wisdom of the Crowds* (WoC). This concept had first been introduced long ago and way before the Internet, by Galton (1907b), who had provided early evidence of the existence of the WoC, by comparing the median estimation of an ox weight from a crowd with its true weight. The crowd had performed surprisingly well, with an estimation error lower than 1%. Needless to say, the development of information technologies has considerably renewed this interest in the WoC. Many recent articles have investigated this phenomenon in its modern settings highlighting a better performance on a task from a crowd rather than an expert (Surowiecki, 2005; Buecheler et al., 2010; Woolley et al., 2010). Independency is one of the founding characteristics, as defined by Surowiecki (2005) in his seminal book, for the WoC to exist. Indeed, members of the group should not influence each other. Each of them should be able to propose her or his own solution.

What happens then for interacting crowds, such as on online fora? Can they also be wise? Since they rely upon social interactions, fora do not seem to fulfill the WoC condition. All the more so as most of the issues discussed

there are qualitative and not quantitative, far from estimating the weight of an ox. Indeed, what takes place in open fora is closer to another historical analysis of crowds, the one of Tarde (1901) who had developed a theoretical framework about the *opinion of the crowd*, where members of the crowd are linked because they are physically in contact with one another. Although this contact is not anymore physical, Tarde's insights guide us towards a larger understanding of what the crowds can do when they interact and address non-quantitative issues.

What can the crowd do online? In the current chapter, we show that it can notably consent, and sometimes dissent with someone, who often has a prominent role in online fora, such as the moderator, the person who has asked the initial question, or a salient member of the community. We address this issue by taking advantage of CMV's dual reward system: an individual reward system where a reward, *the delta*  $\Delta$ , is given by an individual, and a collective reward system where people vote. By analyzing when discrepancies between both occur, we show that the crowd consents when it has participate to the discussion. If the  $\Delta$  is awarded before the crowd debate, we observe that the post with the  $\Delta$  does not attract the vote reward, the crowd reward. Through those two reward systems we observe the mechanism behind a consensus.

In the next section, we review recent literature that studies dialectic activity within online communities. To progress towards a model adapted to the analysis of online discussions, we then describe CMV and our methodology. Next, we present descriptive statistics and results before discussing them and concluding.

## 5.2 Why: Individual vs. crowd

In his seminal book, Surowiecki (2005) develops a theoretical framework for the WoC. He defines four conditions that characterize wise crowds: diversity, independence, decentralization, and aggregation (cf. Section 2.3.1.1 for more

informations). It is then legitimate to ask the question: "Does forum structure aligns with those characteristics?"

Concerning the diversity, the more diverse the crowd is, the more different are the solutions proposed, which increases the chance for a suitable one to be among the proposals. In their last report, *We Are Social* states that more than four billion of individuals are using the Internet around the world <sup>1</sup>. Of course CMV does not gather half of the human beings, but we assume that with 90,000 participants, diversity is ensured.

In "contest" type of fora (Malone, Laubacher, and Dellarocas, 2009), challengers compete with other participants to win the reward. In this case, they do not have interest to exchange with each other. As CMV is a contest forum, in which each challenger tries to get rewarded with the  $\Delta$  by giving the best argument to make the OP changes is view, we assume the independence condition fulfilled.

The decentralization process is a system where decisions are not fully taken by one agent. Q&A fora do not fulfill this condition. The reward is usually awarded by the problem questioner, who selects which solution he prefers. This is the case of the platform *InnoCentive* where companies ask the community to solve a technical problem and pay a financial recompense to the author of the best solution (Brabham, 2008). On other platforms such as StackOverflow, even after the "best" solution has been selected by the questioner, the crowd can react on it. In Dalle, Devillers, and Besten (2014), authors demonstrate that the number of edits (modifications of a post) is higher for the winning answer than for the second best. In both cases, the crowd improves the selected answers performing a decentralized optimization. Finally, aggregation is ensured through the platform itself: the best answer is usually the first one to be seen when someone visits the question thread or at least, one can sort post according to their score. It is also the case in CMV.

The Wisdom of the Crowd (WoC) is an important area of investigation in Open Innovation (Chesbrough, Vanhaverbeke, and West, 2006) since, under the right circumstances, a company could have an advantage to outsource its

---

<sup>1</sup><https://wearesocial.com/blog/2018/01/global-digital-report-2018>

innovation (Afuah and Tucci, 2012). A main condition for the crowd to perform better than a group of experts is the diversity among its members (Hong and Page, 2004; Page, 2008; Srba and Bielikova, 2015). Thanks to diversity — regarding "intelligence", social background, skills, etc. — the proposed solutions are better. Moreover, it is often the combination of these solutions which performs the best to solve a problem (such as the disappearance of the *USS Scorpion* submarine (Surowiecki, 2005)). Nevertheless, the WoC is not always the optimal mean to get the proper solution. One of the obstacles comes from social interactions (Lorenz et al., 2011). If and when members of the crowd have the possibility to exchange about their proposals, a decrease of the diversity can be observed, due to psychosocial mechanisms such as conformism (Asch, 1951). However, under right circumstances, social influences can have a positive impact on the WoC. In experimental settings, Becker, Brackbill, and Centola (2017) bring to light a positive influence of knowing the group answer — the mean answer of the group, on individual estimates when people could modify their answer. In this case, social influences improved both individual and group estimates.

## 5.3 How: Econometrics

### 5.3.1 Reddit – Change My View (CMV)

Reddit – *Change My view* (CMV)<sup>2</sup> admits discussions and opinions on any subject. Such as *Agora* in Ancient Greece, the initiator — Original Poster (OP) — opens a discussion announcing his idea on a topic. CMV has developed constrained rules to join the debate. To open a topic, one agrees to follow the conversation, at least three hours after its beginning and one *must personally hold the view and be willing to have it changed*<sup>3</sup>. All participants, OP included, can award with a delta  $\Delta$  any argument that allowed him to make a step in the change of his view. The reward attribution should be accompanied by the argumentation why the comment has been convincing. The

<sup>2</sup><https://www.reddit.com/r/changemyview/>

<sup>3</sup><http://www.reddit.com/r/changemyview/wiki/rules>

opinion change may be minor as  $\Delta$  does not have to terminate the thread. We thus observe some threads where several  $\Delta$  were assigned. The DeltaBot, a Reddit bot, confirms the  $\Delta$  attribution through a post.

Figure 3.1 presents an example of a thread in CMV<sup>4</sup>. The boxes indicated with solid blue strip indicate the author who is the OP in this example; This author opens a thread about an investment plan called "tontine". He develops his point of view on the subject. He finishes his post by asking the forum members to change his view (CMV). A first challenger answers (post A.1). Then the OP answers (post A.2), attributing a  $\Delta$  to the post A.1 and providing explanations of his decision. The DeltaBot confirms the delta award in the post A.3. Similarly, another challenger answers the OP (post B.1), which the OP answers (post B.2) before the intervention of another strip with orange horizontal lines challenger (post B.3). This is followed by a discussion between the OP and this strip with orange horizontal lines challenging author (between post B.2 and post B.11). In this example, two other challengers participate (posts A.4 and B.12) with or without an answer from the OP or another author.

In the studied forum, as an author can directly address any post, the "thread" framework preserves the thematic coherence of a discussion. Moreover, we have access to an important quantity of information, such as whether a post was awarded with a  $\Delta$  and by whom and its numbers of up- and down-votes; the number of  $\Delta$ s that a participant has been awarded until now approximates the *experience* of the participant in this forum. The  $\Delta$  system allows us to ascertain on the *quality* of a conversation, despite the following bias. Even if the  $\Delta$  does not ensure a full change of view, it gives some hints on the number of relevant arguments and thus, whether a discussion progresses or stagnates.

Reddit — CMV occurs to be an original study case. On the one hand, it has common features of open online discussions. This forum allows its participants to reply one to another. It provides them with an award system, a vote system on posts, and a user experience system. Furthermore,

<sup>4</sup>[https://www.reddit.com/r/changemyview/comments/3mzc6u/cmv\\_the\\_tontine\\_should\\_be\\_legalized\\_and\\_made\\_a](https://www.reddit.com/r/changemyview/comments/3mzc6u/cmv_the_tontine_should_be_legalized_and_made_a)

the discussion structure ensures that one can answer on any sub-comment level, contrary to a Q&A website such as StackOverflow. On the other hand, CMV requires a unique and strict framework on the discussions, as described above. Finally, while Reddit itself has been the subject of numerous research works, CMV has rarely been chosen as a research object. Altogether, CMV appears to be an appropriate and original candidate to evaluate the consensus process in online discussions.

Moreover, CMV fulfills the four characteristics of a wise crowd in action (Section 5.2): diversity and aggregation are ensured by the forum website type, independence by a contest forum type, and decentralization by a vote system. Still, few research examined the presence of the WoC in open online fora. They have focused on the resolution of a problem, which the solution is an objective outcome (Buecheler et al., 2010; Woolley et al., 2010), in experimental settings. There is no such objective solution to change someone's point of view. Still, the crowd influences the discussion outcomes. Thus, instead to focus on the WoC, we propose to study the power of the crowd on the reward system through its *approval*. We address the following research questions:

- What are the conditions for the crowd to manifest its agreement?
- Does the crowd always consent with the  $\Delta$  reward?

### 5.3.2 First delta context

We focus on the context of a reward in CMV threads and want to know how the crowd may influence the attribution of that reward. Besides, several  $\Delta$ s might occur in a thread. The  $\Delta$  attribution underlines a new step in the discussion. Thus, threads with different numbers of  $\Delta$ s are not at the same evolution level.

To be able to compare the threads awarded with at least one  $\Delta$ , we focus our attention on the discussion leading to the first  $\Delta$  attribution. Firstly, we remove all posts occurring after the post awarding the  $\Delta$ , as they have no impact on the  $\Delta$  attribution. Secondly, we assume that each independent

sub-tree, appearing after the removal of the root of a thread (the opening post), models an independent sub-discussion, where different arguments are developed. Thus we only retain the sub-tree where the  $\Delta$  occurred.

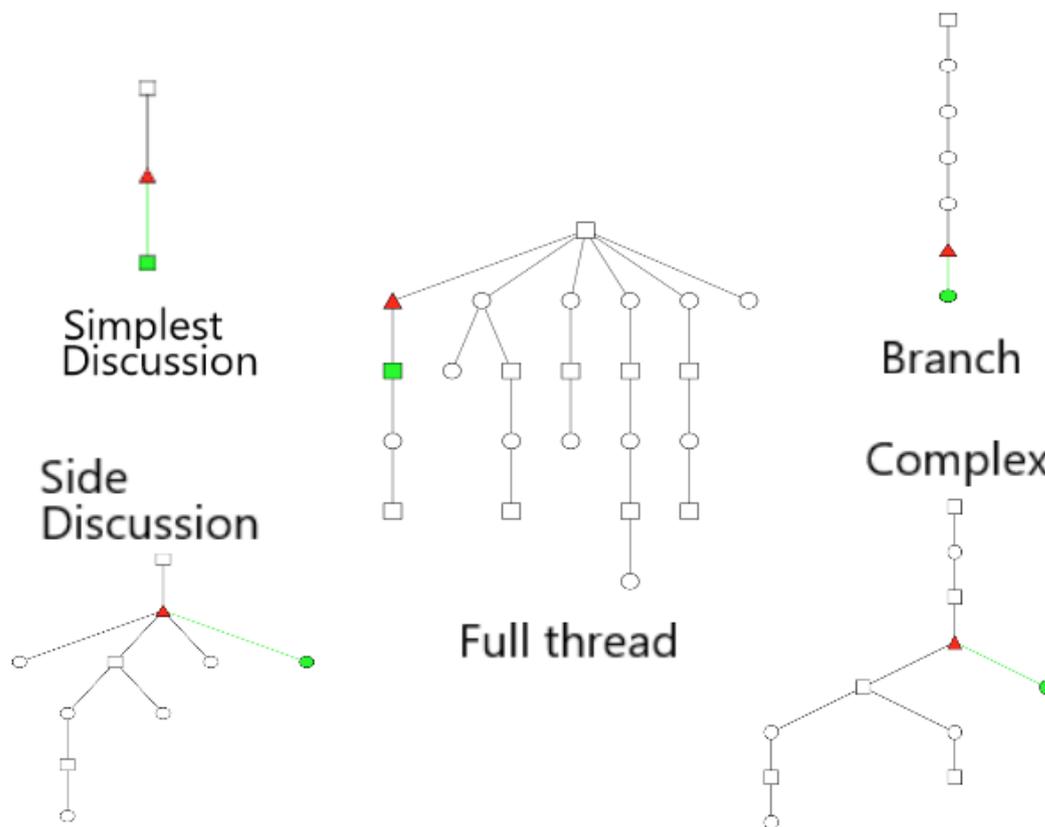
Applying the same pruning process on the 5,773 threads having at least one  $\Delta$ <sup>5</sup>, we observe four types of sub-graphs, presented in Figure 5.2:

1. the *Simple* ones, that are composed solely of the root and two posts, the delta receiving ( $\Delta_r$ ) post and the delta giving ( $\Delta_g$ ) post, 44.10% of the threads — top left corner;
2. the *Branches*, which are composed of several posts between the root and the  $\Delta_r$  post, 37.05% of the threads — top right corner;
3. the *Side discussions*, where a dense sub-discussion occurred between the  $\Delta_r$  and the  $\Delta_g$  posts, 13.81% of the threads — bottom left corner;
4. the *Complex* ones, composed both by a branch and a side discussion, 5.04% of the threads — bottom right corner.

		Simple	Branch	Side	Complex	Row All	% Cluster
Chatters	Freq	405	518	90	63	1076	
	Percent	37.64	48.14	8.36	5.86	100	18.64
Orators	Freq	1451	1254	441	162	3308	
	Percent	43.86	37.91	13.33	4.90	100	57.30
Overflowed	Freq	690	367	266	66	1389	
	Percent	49.68	26.42	19.15	4.75	100	24.06
All	Freq	2546	2139	797	291	5773	
	Percent	44.10	37.05	13.81	5.04	100	100

TABLE 5.1: Discussion Type and Cluster distribution

<sup>5</sup>We also removed one thread where the OP received the  $\Delta$  and threads with "Deleted" account authors, because we do not have enough characteristics on those authors.



Note:  $\square$  = OP posts,  $\circ$  = non-OP posts, red  $\triangle$  =  $\Delta$  receiving post, green =  $\Delta$  giving post.

FIGURE 5.2: Sub-graphs identified after pruning process

For the first type that gather Simple discussion, either the question was easily answered or an expert of the field happened to spot it quickly and gave a strong argument. In both cases, the crowd has no role in the  $\Delta$  attribution process, having no time to do so. Thus, when the sub-tree is a *Simple discussion* type, there are always exactly two authors: the OP and one challenger. We exclude this type of sub-discussions and we only keep those which attracted at least three distinct authors. Indeed, in order to account for crowd characteristics, we keep discussions with a *crowd* composed by at least one participant (other than  $\Delta_r$  and  $\Delta_g$  authors). We also decide to keep crowd composed by only one author, in order to keep all discussions with *external* intervention (in this case, of one author). The new distribution, regrouping 1,590 discussions, is detailed in Table 5.2. We observe mainly *Branch* type (45%), less *Branch* (42%) and few 13% *Complex* discussions.

As complementary information, we look at the distribution of types among

the clusters defined in Chapter 4. As a reminder, cluster *Chatters* covers discussions where the OP answers a large majority of his challengers. *Orators* is similar but OP does not answer all his challengers. And *Overflowed* discussions cover a group where OP could not answer to all requests. In *Chatters*, two third of the discussions are of type *Branch* and one fifth are of the *Side* type. In *Orators* cluster, almost half of discussions are *Branch* against 40% of *Side*. And in *Overflowed*, the majority are *Side* discussion and less than a third are *Branch*. *Complex* discussions are equally represented among the three clusters, grouping more than 13% of discussions.

In a nutshell, the distribution between *Chatters* and *Orators* might highlight a *size* effect. Nevertheless, the stable presence of *Complex* type across clusters emphasizes the fact that when a *crowd* is involved, which it is the case in our subsample, we observe much more *extreme* types (*Branch* and *Side*) rather than mixed one (*Complex*).

		Branch	Side	Complex	Row All	% Cluster
Chatters	Freq	162	54	33	249	
	Percent	65.06	21.69	13.25	100	15.66
Orators	Freq	427	364	123	914	
	Percent	46.72	39.82	13.46	100	57.48
Overflowed	Freq	121	247	59	427	
	Percent	28.34	57.85	13.82	100	26.86
All	Freq	1710	665	215	1590	
	Percent	44.65	41.82	13.52	100	100

TABLE 5.2: Discussion Type and Cluster repartition  $N = 1590$

On the one hand, the *Branch* type attracts significantly fewer authors than the *Side* discussions. In order to detect the influence of the crowd on the rewarding process, we analyze discussion types. The *Branch* type of discussion might reflect an expert intervention, the  $\Delta_g$  author asking for more details before awarding the last post of the expert, the  $\Delta_r$  post. We can therefore hypothesize an absence of crowd's intervention, despite the fact that it has enough time to intervene. On the other hand, the *Side* type attracts significantly more authors. We thus assume that the crowd debates on the content of the  $\Delta_r$  post (before being awarded), which results in the  $\Delta$  occurrence.

Besides the  $\Delta$  mechanism, each participant could up- or down- vote one or several comments (down-votes are allowed but not recommended by the CMV rules<sup>6</sup>). Firstly, a  $\Delta$  reward might be the result of an intense debate between participants, still it reflects the decision of an individual. Secondly, even if the voting action is individual, the visible characteristic of a post is the number of votes it gathered. Thus, the voting system expresses a collective view. We investigate which post received the maximum score in each sub-threads (after the pruning process). This post is assumed as the one which brings along the winning argument for the crowd. The "winner" post may be:

1. the root post, assumed as a popularity choice to point out the interest of the discussion in its entirety;
2. one of the two  $\Delta$  posts, assumed as the consent of the crowd on the reward;
3. any other post, assumed as the disapproval of the crowd. Thus the crowd points out a potential better argument.

If the most popular post is the root, we cannot affirm that the crowd consents with the first  $\Delta$  awarded. Therefore, in this case, we look for the second most popular post occurrence, which is either one of the  $\Delta$  posts or any other post. In Figure 3.1, the  $\Delta_r$  post also received the maximum vote score.

Besides, we must underline that we compare award systems which might occur with a different timing. We know for sure the exact timing of the  $\Delta$  being awarded by a post, for which we get the timestamp. However, it is not the case for the voting system. Thus, the votes might have occur before and/or after the  $\Delta$  attribution. Moreover, the pruning process might cut an important argument occurring after the  $\Delta$  attribution. Still, we observe that the  $\Delta$  posts might also be the posts receiving the most upvotes *in the analyzed sub-discussion*. For this reason, comparing those two award systems allows

---

<sup>6</sup>[https://www.reddit.com/r/changemyview/wiki/guidelineswiki\\_upvoting.2Fdownvoting](https://www.reddit.com/r/changemyview/wiki/guidelineswiki_upvoting.2Fdownvoting)

us to capture a collective decision (vote system) about an individual action ( $\Delta$  reward). From this voting system, we state the following assumption:

**Assumption (A):** If the crowd does not manage to influence the  $\Delta$  rewarding process through posting actions, it will intervene by its vote power.

If the **Assumption** is verified, the *Consent of the Crowd* would be the mechanism through which the crowd will assure the quality of a discussion reward.

## 5.4 What: The Consent of the Crowd

### 5.4.1 Descriptive statistics

#### The variables

We use an open database from Tan et al. (2016)<sup>7</sup>. Authors extracted all discussions from Reddit API from January 2013, the creation of CMV, to November 2016, as widely detailed in Chapter 3. We analyze the following groups of variables in order to understand the dependent variable: the  $\Delta_r$  (or  $\Delta_g$ ) post is or is not the post receiving the largest number of upvotes (labeled afterward *MaxVote*  $\Delta_r$  and *MaxVote*  $\Delta_g$  and equals to 1 if it is the case, 0 otherwise).

Firstly, we capture post activity with a sequence of related variable: a binary variable to detect if the  $\Delta_g$  author is the OP or not (labeled as  $\Delta_g$  OP); the numbers of posts from the OP (*Post OP*) the  $\Delta_g$  author and from the  $\Delta_r$  author (labeled as *Post*  $\Delta_g$  and *Post*  $\Delta_r$ , respectively); the total number of post (*#Post*) and of authors (*#Author*) in the subthread<sup>8</sup>, the number of posts occurring between the root and the  $\Delta_r$  post (labeled as *Post Bfr.*  $\Delta$ ), and the number of posts occurring between the  $\Delta_r$  and the  $\Delta_g$  posts (labeled as *Post Inbtw*  $\Delta$ ); we observe the size of the crowd (how many people participate aside from  $\Delta_r$  and the  $\Delta_g$  authors, then we construct *#Crowd* = *#Author* - 2). Moreover, we observe different temporal variables (in hours): the time

<sup>7</sup><https://chenhaot.com/pages/changemyview.html>

<sup>8</sup>We removed posts from *DeltaBot*, which is moderation posts from a bot developed by creators of CMV.

between  $\Delta_g$  post and the previous one; between the two  $\Delta$  posts ; and how long it takes to award the  $\Delta$  (time between  $\Delta_g$  and the root).

Second, we capture the vote activity: vote received by the root post (first post of the OP),  $\Delta_g$  and  $\Delta_r$  posts (labeled as *Vote Root*, *Vote  $\Delta_g$*  and *Vote  $\Delta_r$* , respectively); we observe the author of the post receiving the maximum number of votes (*MaxVote Author  $\Delta_r$* , *MaxVote Author  $\Delta_g$*  or *MaxVote Author Other*) ; the average number of votes received by a post in the sub-thread (*Mean Vote*) and the high in the tree of the maxvote post (*MaxVote Post High*).

Finally, we measure author characteristics such as experience and age for OP,  $\Delta_r$  and  $\Delta_g$  authors (*OP Exp*,  *$\Delta_r$  Exp* and  *$\Delta_g$  Exp* or *Age*). For the crowd, we compute the average experience (and age) as a proxies for the Wisdom of Crowd (*Crowd Exp/Age*). We also want to control for the diversity of the crowd (Hong and Page, 2004; Page, 2008; Srba and Bielikova, 2015) using the Gini index for this purpose, as it expresses a disparity in the distribution of experience (age), that ranges between 0 and 1. The higher the index is, the more diverse the distribution is (*Crowd Exp/Age Diversity*). The following paragraphs explore Table 5.3 which provides the descriptive statistics of the variables we defined to conduct our analysis.

## Descriptive statistics

### Discussion Types

In 47% of cases, the post which is awarded with the  $\Delta$  is also the one receiving the maximum vote score (plus 1% for  $\Delta_g$  post) and in 71% of the cases, the  $\Delta$  is awarded by the OP. Moreover, even if only 48% of the  $\Delta$  posts received the highest number of votes, 62% of authors who have written the *MaxVote* post are  $\Delta$  authors (58% for  $\Delta_r$  author and 4% for  $\Delta_g$  post). This fact highlights a *second chance* factor, as the crowd may use the vote system as an adjustment for the  $\Delta$  award.

Besides, as seen previously, 45% of the threads are *Branches*, 42% are *Side* and 13% of them are *Complex* types. Moreover, a majority of discussions

belongs to the *Orators* cluster (57%), 27% belong to the *Overflowed* and 16% to *Chatters*.

For the following statistics, we remove from the data set the root post. Indeed, this post brings the same information for all the threads: it is a post written by the OP. It will not help us distinguish the different types of discussions.

### Experience

Concerning the experience, which counts the number of  $\Delta$  since the registration, of main actors, the average experience of the OP is 4, 28 for  $\Delta_r$  authors and 5 for  $\Delta_g$  authors, implying experienced  $\Delta_r$  authors. The average crowd experience level is 17, with the crowd experience diversity index being at 0.58, showing an important heterogeneity.

### Age

Concerning the age, OP are youngest with an average age of 5 months.  $\Delta$  authors are older with an average of 10.6 for the  $\Delta_r$  author and 9.6 for  $\Delta_g$  author, similar to the Crowd age being at 10.4. The diversity in age is lower than in experience, with a Crowd Age Diversity at 0.38, showing a higher homogeneity in Age.

### Vote activity

Besides, we observe vote characteristics. In average, the root post received much more upvotes than any other post (132 against 26 and 5 for  $\Delta_r$  and  $\Delta_g$  posts respectively). The average score is equal to 12. We assume that participants upvote the root as a signal to highlight globally the discussion, in the list of whole threads on Reddit-CMV<sup>9</sup>. Concerning the height of the *MaxVote* post, it is really low (with an average at 1.42 and with a 3<sup>rd</sup> quartile still equals to 1), corresponding to the first level of the answer after the root, highlighting the rapidity of occurring of the  $\Delta$ .

<sup>9</sup><https://www.reddit.com/r/changemyview/>

### Post activity

Finally, concerning *discussion* activities, on average, 4 authors participate in the sub-discussions, writing 6.5 posts in total. 2 participants composed the crowd<sup>10</sup>. The  $\Delta_g$  author participates slightly more than the  $\Delta_r$  author (2.04 posts against 1.80 posts, on average). Moreover, while, on the average, 1.89 posts are written before the  $\Delta_r$  post, there are 2.54 posts in between the  $\Delta$  posts. Besides, times between the important posts are important: in average, 24 hours are necessary to award the  $\Delta$  (*Time btw Root/ $\Delta_g$* ). And the times between  $\Delta_g$ /*previous post* and  $\Delta_g$ / $\Delta_r$  are close to 20 hours.

### Correlation

In this part, we analyze the correlation between variables presented in Table 5.4, summarized in Figure 5.3.

### MaxVote $\Delta$

Our dependent variable *MaxVote  $\Delta$*  is equal to 1 when the post receiving the  $\Delta$  is also the post with the highest number of upvotes, the best score (or the second one when the first one is the root), which is the case in 48% of the discussions. First of all, we observe that such event is negatively correlated with the fact that the  $\Delta$  giver is the OP (-0.09\*\*\*), which is the case in 71%. Besides, its link with the clusters is rather weak. Indeed, it insignificantly correlated with *Orators*, negatively with *Chatters* (-0.17\*\*\*) and positively with *Overflowed* (0.16\*\*\*). On the contrary, we observe a strong correlation with discussion types: -0.59\*\*\* with *Branch*, -0.24\*\*\* with *Complex* and 0.76\*\*\* with *Side*. In both cases, short and few populated discussions seems linked with *MaxVote  $\Delta$*  equal to 0 and, on the contrary, a populated discussion might imply a *convergence* between the  $\Delta$  and vote systems (*MaxVote  $\Delta$*  equals to 1). Besides, experienced OP,  $\Delta_g$  author and Crowd are negatively correlated with this *convergence*, whereas a diversity in crowd experience is positively

<sup>10</sup>This value does not reflect the heterogeneity of the size of the discussions. The biggest one gathered 49 authors, who have written 124 posts, and the crowd size is up to 47.

correlated with it (0.11\*\*\*). We observe similar but weaker trends with age variables. Furthermore, an important vote activity is positively correlated with the *convergence* (up to 0.26\*\*\* for the  $\Delta_r$  vote score). Concerning posts activity, individual (OP is a  $\Delta$  author) activity is negatively correlated with the *convergence*. Conversely, global activity (number of posts, authors - in total and in the crowd) are positively correlated with it. Finally, we observe a really strong correlation with the number of posts *Bfr*  $\Delta_r$  (-0.68\*\*\*) and *Inbtw*  $\Delta$  (0.28\*\*\*) highlighting again the importance of the shape of the discussion on the dependent variable.

### $\Delta_g$ OP

The variable  $\Delta_g$  OP informs us of the fact that the OP awarded the  $\Delta$ . It is positively correlated with experienced OP,  $\Delta_r$  author and Crowd (around 0.1\*\*\*) but negatively with Crowd Experience Diversity (-0.12\*\*\*), and almost insignificantly correlated with age variables. On the contrary, it is negatively correlated with vote variables. It is logically positively correlated with OP,  $\Delta_g$  author and *Bfr*  $\Delta_r$  number of posts and negatively correlated with all other post and time variables.

### Cluster

Concerning cluster variables, we observe an insignificant correlation with *Complex* type. Similarly, *Orators* is weakly correlated with other variables. But we observe an exact opposite correlation between *Branch* and *Side*. When *Chatters* and *Branch* are positively correlated (0.18\*\*\*), *Overflowed* and *Side* are as well positively correlated (0.2\*\*\*). Here again, those correlations *group* discussions with similar shape. Clusters are insignificantly or weakly correlated with experience and age variables. On the contrary, we observe a significant correlation between cluster types and vote activity. *Chatters* and *Orators* follow the same trends: negative correlation (around -0.1\*\*\*) with root and  $\Delta$

posts vote score. *Overflowed* followed opposite trends. Concerning post activity, except a positive correlation with *#Post OP* (0.08\*\*\*), *Orators* is insignificantly (or really weakly) correlated with posts activities. This could be due to the important heterogeneity within this cluster. *Chatters* is positively correlated with *OP*,  $\Delta_g$  and *Bfr*  $\Delta_r$  number of posts and negatively correlated with a high number of authors (in total either as in crowd) and with *#Post Inbtw*  $\Delta$ . Furthermore, *Chatters* is not correlated with time variables. Finally, *Overflowed* followed the exact opposite trends and is positively correlated with time variables. As a confirmation of Chapter 4, what we have uncovered only with motifs characteristic highlight *shape* characteristics of the trees.

### *Discussion types*

Concerning *Discussion type* variables (presented in Figure 5.2), *Complex* is insignificantly correlated with experience and age variables except with *Crowd Exp Diversity* (0.11\*\*\*) and *Crowd Age* (-0.06\*\*). *Branch* is positively correlated with *OP*,  $\Delta_g$  and *Crowd* experiences, but negatively correlated with *Crowd Experience Diversity*. Beside *Branch* is positively correlated with *OP* and *Crowd Age* (0.07\*\*\* and 0.14\*\*\* respectively) and negatively correlated with *Crowd Age Diversity* (-0.08\*\*\*). *Side* follows opposite trends than *Branch* for both experience and age variables. In a nutshell, experienced and older participants seem to be more present in *Branch* than in *Side*. Furthermore, *Complex* is insignificantly or weakly and lowly correlated with vote activities. With those variables, we observe again opposite trends between *Branch* and *Side*. *Branch* is negatively correlated with *OP*, both *Delta* authors and mean vote posts, and when the *MaxVote* author is  $\Delta_r$  author (-0.57\*\*\*). It is positively correlated with *MaxVote Post High* (0.21\*\*\*), and when the *MaxVote* post author is either a  $\Delta_g$  or a non- $\Delta$  author (0.12\*\*\* and 0.53\*\*\* respectively). Finally, we observe that *Branch* and *Side* follow opposite trends concerning post activities as well. On one hand, *OP*,  $\Delta_g$  and *Bfr*  $\Delta_r$  number of posts are strongly positively correlated with *Branch* (0.16\*\*\*, 0.33\*\*\* and 0.70\*\*\* respectively). On the other,  $\Delta_r$ , *Inbtw*  $\Delta$  and total number of posts, of authors

(in total and in crowd) and time variables are strongly negatively correlated with *Branch* (up to  $-0.41^{***}$  with *Inbtw*  $\Delta$  variable). Concerning *Complex*, it is strongly positively correlated with those variables, except for the number of authors (total and crowd) and with time variables. These observations confirm the *physical* characteristics of discussion types, at least for the *extreme* shapes *Branch* and *Side*.

### *Experience*

Concerning *Experience* variables, we observe that more experienced OP attract more experienced challenger. OP and  $\Delta_g$  author and Crowd experiences (average experience of individuals composing the crowd), which are positively correlated between each other (up to  $0.85^{***}$  between OP and  $\Delta_g$  author),  $\Delta_r$  author experience being insignificantly correlated with the others. Furthermore, Crowd Experience Diversity is negatively correlated with OP Experience ( $-0.06^{***}$ ) and positively with  $\Delta_r$  Experience ( $0.13^{***}$ ). Besides, experience comes with age. Crowd Experience Diversity is insignificantly correlated with age variables, except a positive correlation with Crowd Age Diversity ( $0.09^{***}$ ).

OP, and  $\Delta$  authors' experiences are not significantly correlated with vote variables. But it is with the *MaxVote Post Author*  $\Delta_r$ . Nevertheless, Crowd Experience and Diversity are correlated with vote variables (except for *Vote*  $\Delta_g$ ), in opposite trends. Crowd Experience is positively correlated with *MaxVote Post High* and *MaxVote Author Other*, and negatively with other variables. An experienced crowd will less upvote first posts (OP and  $\Delta$  posts) and maybe dig deeper in the discussion.

Finally, what is the link between experience and post activity? OP,  $\Delta$  authors and Crowd experiences are significantly positively correlated with OP and  $\Delta_g$  and *Bfr*  $\Delta_r$  number of posts, but insignificantly correlated with global activity and time variables. Besides, *Crowd Experience* is negatively correlated with the number of authors (total and in the crowd), confirming the trend with *Crowd Experience Diversity*, which is positively and strongly correlated

with the number of authors. The bigger the crowd is, the less experienced it is and the more we observe diversity in experience among its participants.

### *Age*

Same as experience variables, old attract old. Age variables are strongly and highly positively correlated with each other, except with *Crowd Age Diversity* which is negatively correlated to other variables. Age is not strongly correlated with Vote variables, except for  $\Delta_g$  Age being correlated with vote activities (equal or close to 0.09\*\*\*). *MaxVote* authors variables are insignificantly correlated with age variables except for Crowd Age (average of the age of individuals of the crowd) which is negatively correlated with  $\Delta_r$  author (0.08\*\*\*) and positively with non- $\Delta$  authors (0.08\*\*\*).

Finally, OP,  $\Delta_g$  ages and *Crowd Age Diversity* are positively correlated with all post activities. The elders post more and participate to bigger discussions. Furthermore, they take less time to post (even if the correlation is insignificant with OP age and weakly significant with  $\Delta_g$  age, it is strongly - yet lowly - with  $\Delta_r$  Age). *Crowd Age* is only significant with OP,  $\Delta_g$  and *Bfr*  $\Delta_r$  number of posts.

### *Vote activity*

Votes attract votes: strong highly positive correlation between Root,  $\Delta_s$  and Mean vote, up to 0.6 \* \*\* between  $\Delta_r$  and  $\Delta_g$ . But it is negatively with *MaxVote Post High* highlighting the fact that the higher the MaxVote post is, the less it might be the Root (which is level 0 exclusively) or  $\Delta$  posts. Besides, Root and  $\Delta_r$  posts vote follow the same trends concerning post activities: negatively correlated with OP,  $\Delta_g$  and *Bfr*  $\Delta_r$  number of posts, and positively with other post variables.  $\Delta_g$  and Mean Vote follow the same trends as the first two variables but with weaker and lower correlation, except for the negative one between Mean Vote and Post  $\Delta_r$ , which is equal to  $-0.07 * **$ . Vote variables are not significantly correlated with Time variables.

### *Post activity*

Finally, post activity calls for post activity. Almost all variables are positively correlated to each other. Still, different observations must be highlighted. OP and  $\Delta_g$  number of posts are highly correlated ( $0.74 * **$ ), such as any variables between those two roles. As a reminder,  $\Delta_g$  author is the OP in 71% of the discussions. A particular vigilance must be set when modeling in the regression those two roles. Furthermore, correlation between a number of posts and authors (in total and in the crowd) variables are as well very high (up to 1 between authors and crowd). Knowing that a regression modeling is sensible to size effect, we will normalize proper variables with those. Concerning time variables, when significantly correlated with variables, they are weak and low. We remove those in the regression model.

### **Descriptive statistics conclusion**

In order to conclude this first step, we would like to summarize what we have learned from the descriptive statistics. First, all these *mean* statistics hide an important heterogeneity in the variables, presented through the standard deviation and quartile statistics. The database is mainly composed of small discussions (7 posts produced by 4 authors up to the 3<sup>rd</sup> quartile). But as it is common in web characteristics (and in numerous other fields such as physics, biology, economics, ...), the distribution of those are close to a power-law (Barabási and Albert, 1999; Cooper and Frieze, 2003; Newman, 2005; Ben-Naim and Krapivsky, 2009; Aliakbary, Habibi, and Movaghar, 2014) even if some voices raise against this assertion (Bu and Towsley, 2002; Sowe, Stamelos, and Angelis, 2008; Broido and Clauset, 2018; Klarreich, 2018). In order to properly capture effects of independent variables over the dependent one, we must take into account the *size* effects. Especially as we study characteristics which are strongly dependent on a shape of the discussion, as suggested by correlations with *cluster* and *type* variables. Secondly, as OP is the  $\Delta_g$  author in 71% of discussions, we have a particular challenge if we

want still to distinguish those two roles. The following section will present our modeling approach and how to deal with those issues.

### 5.4.2 Regression modelling

To test assumption **A1**, we apply a logistic regression as defined in the following model:

$$P(\text{MaxVote } \Delta = 1) = \frac{e^{\alpha + \beta X}}{1 + e^{\alpha + \beta X}}$$

with  $X$  the covariates matrix including the variables described above with some modifications in order to take into account issues raised in the previous section.

We did not keep the *type discussions* characteristics, being strongly correlated with *Post Bfr*  $\Delta$  and *Post Inbtw*  $\Delta$ . But we keep cluster variables, setting *Orators* as the reference. We keep all experience and age variables. We only keep the average vote score posts and the height of the post with maximum vote score.

In order to control for size effect, we divided *OP*, *Delta<sub>r</sub>* and *Delta<sub>g</sub>* posts variables by the total number of posts in the sub-threads. Due to the closeness of *OP* and  $\Delta_g$  author, we choose to keep only  $\Delta_g$  author characteristics and add the cross-variable the  $\Delta_g$  *OP*  $\times$   $\Delta_g$  author variables.

For *Post Bfr*  $\Delta_r$  and *Post Inbtw*  $\Delta$ , we divided by the number of non- $\Delta$  posts which is exactly equal to the sum of those two variables, making them collinear. We then only keep *Post Inbtw*  $\Delta$ /*Crowd Posts* variable, defining the percent of posts occurring between the two  $\Delta$  posts. We keep *Crowd Size* variable, dividing it with *Authors*, number of participants in the subthread. Furthermore, we study the different levels of *Crowd Size*, from 1 to 5 or more, setting 2 as the reference. The distribution of this variable is presented in Table 5.5. By construction,  $\text{CrowdSize} = \text{Authors} - 2$ , then,  $\text{CrowdSize} = 1$  always matches with  $\text{CrowdSize} / \text{Authors} = 33\%$ .

Finally, we remove *Authors* and *Posts* variables.



Crowd Size	1	2	3	4	$\geq 5$
Crowd Size/#Authors (%)	33	50	60	66	$\geq 70$
Frequency	1020	293	124	57	96

TABLE 5.5: Crowd Size distribution

Our final selection is confirmed by correlation in Table 5.6<sup>11</sup>, summarized in Figure 5.4.

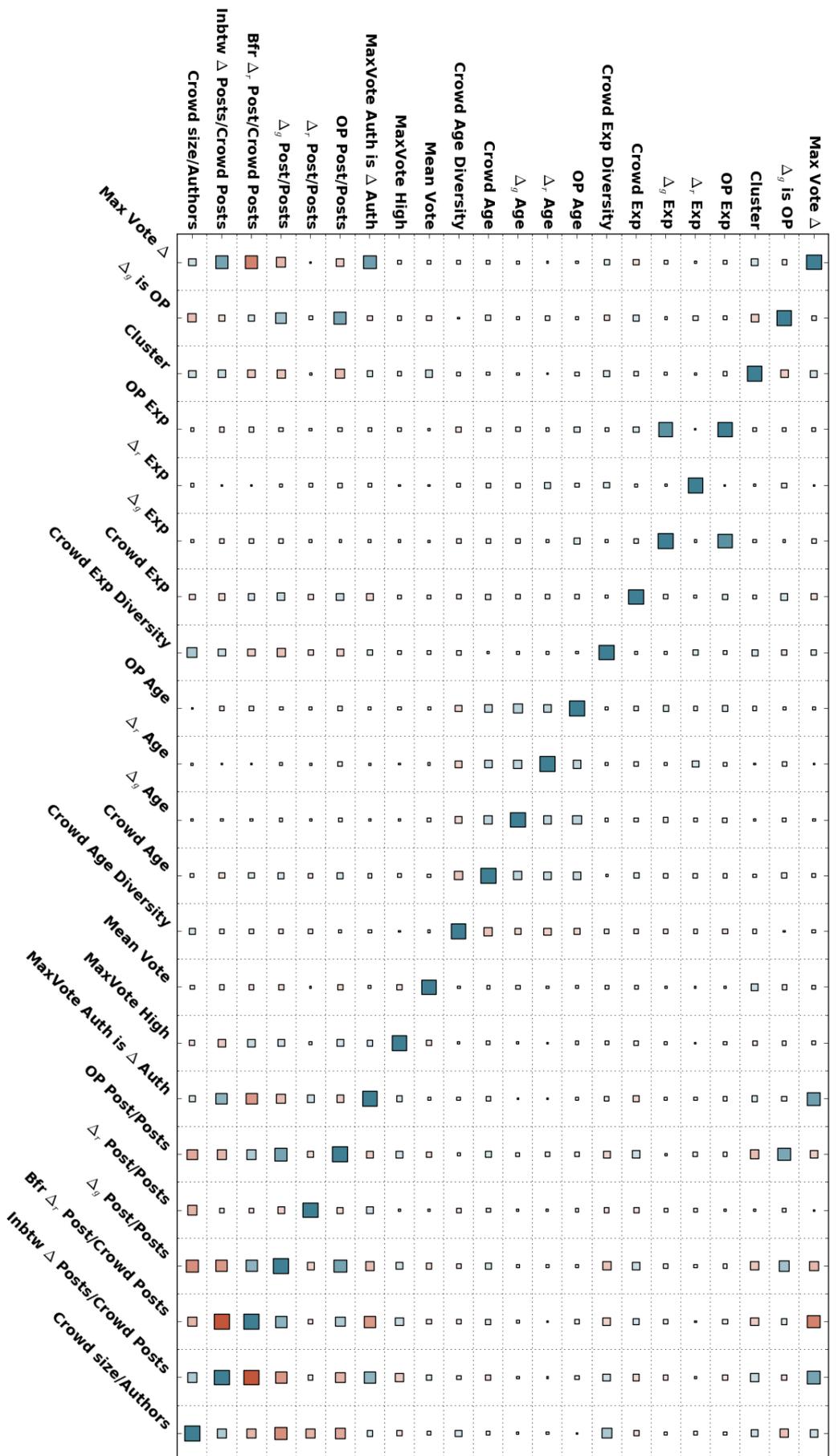
### 5.4.3 Result discussion

#### 5.4.3.1 Results

Table 5.7 presents the results of the regression modeling. A positive coefficient underline an increases in the probability for the post which received the maximum vote score to be a  $\Delta$  post. To ease the reading, we use the odds ratio. To get the odds ratio, one converts a variable coefficient  $b$  by the following formula:  $O.R._b = exp(b)$ , we analyze the inverse  $1/exp(b)$  if  $b$  is negative. As a robustness check, we provide the results from different covariates combinations.

Models from 1 to 6 present the regression by group of variables: 1 — cluster ; 2 — experience ; 3 — age ; 4 to 6 — post activity. The post activity variables play an important role. In modeling 4, we do not include *Crowd Size/# Authors*, added in modeling 5. In modeling 6, we use the *Crowd Size* level variable, with 2 as the reference. The last 3 modelings (columns 7 to 9) add all groups of variables, with *Crowd Size* different variables (7 — none, 8 — percentage , 9 — leveling). In order to select the best modeling, we look at the Akaike Information Criterion (Akaike, 1974) (which we want to minimize) and the McFadden  $R^2$  (McFadden, 1973) (which we want to maximize). Following those robustness indicators, we analyze the last three models. If a variable is significant in a model, but not in all the others, we might face hidden specification. For instance, cluster variables are significant only in

<sup>11</sup>The differences solely concern the post activity variables.



Note: The more the coefficient is close to one, the more the color tends toward a deep blue ; the more the coefficient is close to -1, the more the color tends toward a deep red.

FIGURE 5.4: Correlation heatmap for regression  $N = 1590$

the first model, where we used only those in the regression. But this modeling performs far less well than modeling 9. Thus, we do not keep the first modeling.

How does each variable affect our dependent variable? Neither cluster, nor individual experience, nor age, nor  $\Delta_g$  OP variables has a significant impact.

The experience of the crowd is weakly significant ( $-0.01 * *$ ). If *Crowd Experience* increases by one, the odds of *Convergence* - of the two rewarding systems - decrease by a factor of 1.01 ( $O.R._{Crowd Exp} = \frac{1}{\exp(Coeff_{Crowd Exp})}$ ). But, *Exp Crowd Diversity* is significantly negatively impacting the *Convergence* ( $-1.48 * *$ ): it decreases by a factor of 4.4 the odds of *Convergence*.

Furthermore, *Crowd Size* variables added important informations. Models including this characteristic perform better than without. But it might change significantly specific variables. Nevertheless, when we look at *Crowd Size/#Authors*, it has a strong negative effect on the *Convergence* ( $-6.25 * **$ ): it decreases by a factor of 518 the odds of *Convergence*. But when we take a closer look to it, binarizing it as explained in Section 5.4.2, we observe the exact size of the crowd to be really important. Setting the reference *Crowd Size* = 2 the following analysis compares the impact of the *Crowd Size* = X on the *Convergence*, with respect to what happens when *Crowd Size* = 2. Then, when the Crowd is composed by one individual, it has a high strong effect on the *Convergence* ( $1.61 * **$ ,  $O.R. = 5$ ), in respect with the case where the Crowd is composed by two persons. On the contrary, when *Crowd Size* = 3, it decreases by a factor of 2.3 the odds of the *Convergence*. Higher sizes present negative impact as well, but coefficient are not significant.

Besides, variables having the strongest and highest impact on the *Convergence*<sup>12</sup> are those characterizing the shape of the tree: *posts activity* and *MaxVote Post High*. On the one hand, both counts of posts from  $\Delta_r$  and  $\Delta_g$  authors have a strongly negative impact on the odds of *Convergence* (O.R. are equal to 2440 and 28 respectively), even if  $\Delta_g$  Posts coefficients are less stable and with a much lower impact than  $\Delta_r$  variable. On the other, both *MaxVote*

<sup>12</sup>Labelling *MaxVote*  $\Delta_g = 1$ .

*Post High* and *Inbtw  $\Delta$  Posts/#Crowd Posts* have a strong positive impact on the odds of *Convergence* (O.R. are equal to 3 and 292 respectively).

We learn several lessons from these results. First, the role of the  $\Delta_g$  author (being the OP or not) has no impact on the *Convergence*. As seen in the Chapter 4, OP attracts more attention and interacts a lot with challengers. But once the  $\Delta$  is awarded, challengers do not look attentively who attributed it.

Furthermore, we observe the same absence of impact from experience and age. While, within a discussion<sup>13</sup>, experience is present next to the pseudonym of the author, age is knowable only on the author page<sup>14</sup>. Then, while arguing, only the experience might work as an *appeal to authority*, as a *digital signal* of popularity (Connelly et al., 2010).

Numerous research study the role of popularity and virtual rewards on fora, increasing contributions (Deterding et al., 2011; Trockman et al., 2018; Merchant et al., 2019). On a problem solving forum, such as StackOverflow, research highlights the importance for participants to acquire an important reputation (matching with our *experience* variable), attracting even recruiters (Dabbish et al., 2012; Dabbish et al., 2013) from outside the forum. In *CMV*, the stacks are different, and the goal is to convince someone, at least to change his view at some point. But it is interesting to observe that an experienced author influence a discussion, still through his argument, and not by its *experience signal*.

Nevertheless, experience does have an impact as a crowd characteristic. And not only the average experience of the crowd - its coefficient being significant with  $p < 0.05$  and really low  $-0.01$  - but also the distribution of experience among individuals. Indeed, the results highlight the importance of *homogeneity* in experience to increase the odds of the *Convergence*. Put another way, the more the crowd is composed by people with different levels of experience, the more the crowd diverges from the  $\Delta_g$  author, through the vote system. Heterogeneity or diversity is a well-known characteristic in Wisdom

<sup>13</sup>e.g. [https://www.reddit.com/r/changemyview/comments/fq02gk/meta\\_temporary\\_suspension\\_of\\_covid19coronavirus/](https://www.reddit.com/r/changemyview/comments/fq02gk/meta_temporary_suspension_of_covid19coronavirus/)

<sup>14</sup>textite.g. <https://www.reddit.com/user/dublea/>

of Crowd. Surowiecki (2005) explicitly highlights diversity as a proper characteristic of wise crowds. Page (2008) develops a whole thesis in favor of *difference* in problem solving teams. In the current research, we do not rule in favor nor in disfavor of the *Convergence*. But we must conclude through this result that the more diverse a crowd is, the less it will confirm  $\Delta$  choice, and will highlight another post as Crowd's *best answer*.

Moreover, *Crowd Size* variables add valuable information. The rough variable *Crowd Size/#Authors* indicates a massive negative impact on the *Convergence*. The bigger the crowd is, the less it agrees with  $\Delta_g$  author. The continuous crowd variable *Crowd Size/#Authors* signals a linear effect, where each new member to the crowd highly decreases the odds of *Convergence* by a factor of (518). The level variables *Crowd Size = X* add a new information. Actually, when the crowd is in the minority in the discussion (size = 1), the odds of *Convergence* increases by a factor of 5, compared to a situation where the crowd is as numerous as the  $\Delta$  authors (size = 2). When the crowd includes a majority of participants (size = 3, in respect when it is as numerous as the  $\Delta$  authors), the odds decreases by a factor of 2. But an additional member of the crowd (size > 3) does not impact the odds. We conclude that a bigger crowd has a tendency to less agree with the  $\Delta_g$  author choice and highlights, through the vote system, another answer. The straightforward explanation is the following: the bigger the crowd is, the more difficult it is to it to agree on the same post, furthermore, on the  $\Delta_g$  author choice.

Finally, post activity variables impact the *Convergence*. First, the more  $\Delta$  authors post, the more the odds of *Convergence* decreases (up to a factor of 2440 for  $\Delta_r$  author and to 28 for  $\Delta_g$  author). As a reminder, those variables are percent of  $\Delta$  authors posts over the sub-thread posts. Then, the more those two particular actors take space in the discussion, the less we observe a *Convergence*. On the other hand, we observe an increase of the odds when the max vote score post is *highest* in the discussion (*MaxVote Post High*) AND when the crowd is more active in the discussion (*Inbtw  $\Delta$  Posts/#Crowd Posts* before the  $\Delta$  attribution. Besides, those two variables are negatively correlated ( $-0.26 * **$ ). But, in both cases, more time is spent in the discussion,

because more posts are created. Then, more discussion increases the odds of *Convergence*. Furthermore, the *Inbtw* variable is opposed to *Bfr*  $\Delta_r$  Posts count. The increase of the odds of *Convergence* by a factor of 260 of the *Inbtw* variable is in respect with a low number of posts before the  $\Delta_r$  one. This variable highlight the shape of the discussion between *Side Discussion* type (if *Inbtw* variable is high) and *Branch* type (otherwise) — cf. Figure 5.2. This result allows us to conclude on the fact that: the more  $\Delta_r$  is discussed before receiving the  $\Delta$ , the more the Crowd **consents**, through the vote system, to the  $\Delta$  reward.

This result validates our initial **Assumption**: when the crowd participates to the debate before  $\Delta$  attribution, it validates  $\Delta_r$  post through the vote system.

#### 5.4.3.2 Discussion

The obtained results indicate the power of votes in the open forum CMV. Comparing a discussion to a one-against-all game, if one challenger plays alone and wins a reward before enough other challengers have the time to argue, the crowd will punish her by voting for another argument. Moreover, the crowd acts like the blind justice: we find that it does not take into account who are the  $\Delta$  authors or the experience signal, but rather if the author awarded with  $\Delta$  leaves free room to other challengers and whether the author who attributed a  $\Delta$  gives enough time for others to intervene. Additionally, we find that the wiser and the more diverse the crowd is, the more it is prone to disagree. The OP would really benefit from a discussion with different challengers about the same argument. The more challengers have the possibility to express themselves, deliberating among them and with the OP, the more they will consent with the ultimate reward: the  $\Delta$ . **The Consent of the Crowd (CoC) reflects a quality measure on the reward  $\Delta$ .**

### Limits

Even if important mechanisms emerged from that analysis, we need to investigate the limitations. First, in this work we estimate the vote mechanism as a response to the  $\Delta$  reward. Data at our disposal is, however, a snapshot. We do not know the timestamp nor the author of a vote. Therefore, another possibility is that the crowd highlights a pertinent argument with votes in order to help the OP to find it. The question is whether the vote of the  $\Delta_r$  post occurred before the  $\Delta$  attribution.

Moreover, a complementary limitation concerns the meaning of upvoting. We assume that numerous upvotes on a post reflects the quality of a post (Dalle, Devillers, and Besten, 2014). However, as mentioned in the CMV rules<sup>15</sup>, moderators encourage to upvote an argument which one finds not good enough, in order to attract other challengers to refute it, as mentioned in the precedent argument. We are in front of research framework limits. Access to vote timestamps would solve it.

Besides, some econometrics results seem counter intuitive. First, the more the discussion shape is close to a *Branch* type (low *Inbtw  $\Delta$  Posts/Crowd Posts*, equivalent to a high *Bfr  $\Delta_r$  Posts/Crowd Posts*), the more vote system diverges from  $\Delta$  system. The higher max vote score post is, however, in the discussion (a high *MaxVote High* variable), the more reward systems converge. And, *MaxVote High* is positively correlated with *Bfr  $\Delta_r$  Posts/Crowd Posts* ( $0.26^{***}$ ). Then *Branch* type leads to both increase and decrease of the convergence. Actually, in the first situation, the result of *Inbtw  $\Delta$  Posts* variable is in respect with *Bfr  $\Delta$  Posts*. Thus, both results highlight the same fact: more exchanges among participants, both before  $\Delta_r$  post and in between  $\Delta$  posts, increase the odds of *Convergence*. But discussion in between  $\Delta$  posts is much more effective than before  $\Delta_r$  post. Besides, we observed that crowd size has a negative impact on the odds of *Convergence*. Thus, on the one hand, the more active the crowd is the more reward systems converge toward the same answer. But on the other, the biggest the crowd is, the more the reward systems diverge. And

<sup>15</sup>[http://www.reddit.com/r/changemyview/wiki/guidelineswiki\\_upvoting.2Fdownvoting](http://www.reddit.com/r/changemyview/wiki/guidelineswiki_upvoting.2Fdownvoting)

we observe a positive correlation between those two variables ( $0.39^{***}$ ). We conclude in the sense that the bigger the crowd is, the more difficult to reach a convergence between reward systems, *i.e.*, a *consensus*. Indeed, the bigger it is, the more views on a subject might be different. Especially on *Change My View*. If someone posts, it is either to offer a new view, or to bring precision in favor or in disfavor of another one. Nevertheless, each participant expresses his view, and might divide the vote pool. But, the expression of each one view, goes along with discussion and confrontation of those views. And finally, participants converge toward a same conclusion of the discussion: the  $\Delta_r$  post. This process is close to another present in the *Deliberative Democracy* theory. Jürgen Habermas, major contributor to the theory (Habermas, 1996; Habermas and Burger, 1998), highlights the importance, for a democracy, to let each member participate in political decisions. Cohen (1989) offers his own *deliberative process*, composed by specific characteristics. We observe a deliberative process when:

1. we observe argumentation, critical exchanges among participants
2. it is open to everybody, no one may be excluded in principle
3. it is free of external coercion and only the participants can make the final decision
4. it is free of internal coercion: each has an equal opportunity to be heard

Cohen pursues with characteristics specific to the political sphere. But we can already find those characteristics on *Reddit - CMV*. Thanks to the strict rules, administrators achieved to settle a framework which goes along a deliberative democracy.

### **Impact for moderators**

Nevertheless, our results highlight behaviors that constraint this framework. Those results can be translated into moderation enhancements to improve

and smooth the deliberative process. *Reddit - CMV's* framework defines specific rules to conduct a discussion. When an OP begins a discussion, he commits himself to *stay*, to follow the discussion at least during its first three hours. As a reminder, in this forum, the goal is to *change the view* of the OP, and exclusively of the OP. Then, ensuring his presence *in* the discussion seems a pertinent constraint. Nevertheless, our results highlight somehow *haste* in OP's behaviors, at least from the Crowd point of view. As developed by Habermas and Burger (1998) and by Cohen (1989), a proper deliberative process let everybody *take his, her chance* to express his/her idea about the topic discussed.

Then, we offer to complete the *Convince the OP* rule by the following: *Convince the OP, with your co-challenger*. In order to do so, *Reddit - CMV* moderators could add a minimum amount of time before the OP is able to *award the  $\Delta$* .

Furthermore, a duality emerges. On the one hand, we need an incentive to stimulate challengers. This is done through the  $\Delta$  system. On the other hand, to reach the meta-goal to convince the OP, we need him not to be drowned by the number of challengers. To work out this dual issue, we suggest limiting the creation of sub-branches of the tree, which means, the creation of sub-discussions as we have defined it. To do so, each argument could be presented to new challengers to see if their argument has not already been discussed.

#### 5.4.4 Conclusion

We defined as the *Consent of the Crowd* the power of a crowd in a open online forum to shine a light on a collective result, rather than an individual one. Focusing on a particular period of the discussion, from the beginning up to the first reward, our results bring in evidence the use of the upvote reward system as a punishment against the challenger rewarded with the  $\Delta$ . This happened when crowd members do not get the chance to join the discussion before the  $\Delta$  attribution. From the research point of view, we explain how a

crowd takes actions if it perceived an injustice. Moreover, this result should urge CMV moderators to foster multi-challengers sub-discussions.

Our future work will focus on the extension of the analysis on different kinds of fora, such as StackOverflow. This Q&A forum is not concerned by the first limitation mentioned in Section 5.4.3. It does not suffer from the second limitation as the down-votes are fully accepted. Moreover, we will confront the texts of  $\Delta$  post and *MaxVote* post when they differ (Adamic et al., 2008) in order to compare arguments.

In particular for Stackoverflow, we have access to the timestamp of the vote, and thus might be able to investigate the first limitation we mentioned in. Moreover, we will investigate the second limitation by confronting the texts of  $\Delta_r$  post and *MaxVote* post, when they differ.

As a general conclusion, our results highlight a form of collective intelligence in open fora, increasing the comprehension of participants interactions in online fora.

	Freq	Percent						
MaxVote $\Delta_r$ post	740	47						
MaxVote $\Delta_g$ post	18	01						
MaxVote Other	832	52						
MaxVote Author $\Delta_r$	917	58						
MaxVote Author $\Delta_g$	70	04						
MaxVote Author Other	603	38						
$\Delta_g$ OP	1132	71						
Chatters	249	16						
Orators	914	57						
Overflowed	427	27						
Branch	710	45						
Side	665	42						
Complex	215	13						
	count	mean	std	min	25%	50%	75%	max
OP Exp	1590.0	3.94	24.33	0.00	0.00	0.00	1.00	257.00
$\Delta_r$ Exp	1590	28.22	53.61	0.00	1.00	6.00	21.00	257.00
$\Delta_g$ Exp	1590	5.17	28.12	0.00	0.00	0.00	1.00	257.00
Crowd Exp	1590	17.17	36.67	0.00	0.00	3.00	15.00	257.00
Crowd Exp Diversity	1590	0.58	0.15	0.00	0.50	0.61	0.67	0.95
OP Age	1590	5.03	8.46	0.00	0.00	0.00	7.00	44.00
$\Delta_r$ Age	1590	10.56	11.29	0.00	1.00	7.00	17.00	46.00
$\Delta_g$ Age	1590	9.62	10.83	0.00	1.00	5.00	16.00	45.00
Crowd Age	1590	10.35	10.41	0.00	2.00	7.00	16.00	45.00
Crowd Age Diversity	1590	0.38	0.18	0.00	0.26	0.37	0.51	0.86
Vote root	1590	132.04	270.30	0.00	7.00	23.00	94.00	2276.00
Vote $\Delta_r$	1590	26.46	70.89	-31.00	3.00	7.00	21.75	1201.00
Vote $\Delta_g$	1590	5.44	18.27	-17.00	1.00	2.00	4.00	483.00
Mean vote	1590	11.92	25.38	-1.00	2.38	4.75	10.25	438.67
MaxVote Post High	1590	1.42	0.97	1.00	1.00	1.00	1.00	8.00
#Post OP	1590	1.90	1.41	0.00	1.00	2.00	3.00	17.00
#Post $\Delta_r$	1590	1.80	1.15	1.00	1.00	1.00	2.00	14.00
#Post $\Delta_g$	1590	2.04	1.20	1.00	1.00	2.00	3.00	13.00
#Post	1590	6.42	5.26	3.00	4.00	6.00	7.00	124.00
#Author	1590	3.93	2.49	3.00	3.00	3.00	4.00	49.00
#Crowd	1590	1.93	2.49	1.00	1.00	1.00	2.00	47.00
#Post Bfr $\Delta_r$	1590	1.89	2.01	0.00	0.00	2.00	4.00	7.00
#Post Inbtw $\Delta$	1590	2.54	5.58	0.00	0.00	1.00	3.00	122.00
Time btw $\Delta_g$ /previous post	1590	18.57	130.96	0.01	0.61	1.74	7.32	2631.61
Time btw $\Delta_r/\Delta_g$	1590	19.78	132.39	0.01	0.47	2.21	8.97	2651.26
Time btw Root/ $\Delta_g$	1590	24.45	133.33	0.24	2.22	5.93	15.23	2655.16

Note: we provide the percent of threads where *MaxVote* and  $\Delta_g$  OP variables are equal to one

TABLE 5.3: Descriptive Statistics  $N = 1590$

	MaxVote $\Delta$	$\Delta_s$ OP	Chatters	Orators	Overflowed	Branch	Side	Complex	OP Exp	$\Delta_r$ Exp
MaxVote $\Delta$	-0.09***	-0.17***	-0.02	0.16***	-0.23***	-0.59***	0.76***	-0.24***	-0.07***	0.0
$\Delta_s$ OP	-0.17***	0.16***	0.16***	0.09***	-0.23***	0.14***	-0.11***	-0.04*	0.06**	0.1***
Chatters	-0.17***	0.09***	-0.5***	-0.5***	-0.26***	0.18***	-0.18***	-0.0	0.04	0.0
Orators	-0.02	-0.5***	-0.7***	-0.7***	-0.7***	0.05*	-0.05*	-0.0	0.04	0.01
Overflowed	0.16***	-0.23***	-0.26***	-0.2***	-0.2***	0.2***	0.01	-0.07***	-0.07***	-0.01
Branch	-0.59***	0.14***	0.18***	0.05*	-0.2***	-0.76***	-0.76***	-0.36***	0.09***	-0.02
Side	0.76***	-0.11***	-0.18***	-0.05*	0.2***	-0.36***	-0.34***	-0.34***	-0.07***	-0.0
Complex	-0.24***	-0.04*	-0.0	-0.0	0.01	0.09***	-0.07***	-0.03	-0.03	0.03
OP Exp	-0.07***	0.06**	0.04	0.04	-0.07***	-0.02	-0.03	0.03	0.0	0.0
$\Delta_r$ Exp	0.0	0.1***	0.0	0.01	-0.01	-0.02	-0.0	0.03	0.0	0.0
$\Delta_s$ Exp	-0.07***	-0.02	0.02	0.02	-0.04	0.1***	-0.07***	-0.05*	0.85***	0.02
Crowd Exp	-0.15***	0.17***	0.02	0.06**	-0.09***	0.16***	-0.17***	0.02	0.14***	0.03
Crowd Exp Diversity	0.11***	-0.12***	-0.11***	-0.04	0.14***	-0.23***	0.16***	0.11***	-0.06**	0.13***
OP Age	-0.02	0.03	0.04*	0.01	-0.05**	0.07***	-0.06**	-0.02	0.14***	0.07***
$\Delta_r$ Age	0.01	0.08***	0.0	0.01	-0.01	0.01	-0.0	-0.01	0.05**	0.17***
$\Delta_s$ Age	-0.03	0.04*	-0.03	0.01	0.01	0.02	-0.03	0.02	0.08***	0.07***
Crowd Age	-0.06**	0.11***	0.03	0.01	-0.04	0.14***	-0.1***	-0.06**	0.08***	0.08***
Crowd Age Diversity	0.05*	0.0	-0.05**	-0.01	0.05**	-0.08***	0.07***	0.01	-0.11***	-0.06**
Vote root	0.07***	-0.18***	-0.17***	-0.14***	0.29***	-0.14***	0.11***	0.04	-0.0	-0.01
Vote $\Delta_r$	0.26***	-0.14***	-0.12***	-0.1***	0.2***	-0.23***	0.28***	-0.06**	-0.02	-0.0
Vote $\Delta_s$	0.1***	0.02	-0.08***	-0.05**	0.13***	-0.1***	0.1***	-0.01	0.0	0.01
Mean vote	0.07***	-0.09***	-0.14***	-0.1***	0.23***	-0.09***	0.09***	0.01	0.01	0.0
MaxVote Post High	0.06**	0.07***	0.03	0.06**	-0.09***	0.21***	-0.27***	0.08***	0.04	-0.0
MaxVote Author $\Delta_r$	-0.09***	-0.1***	-0.13***	-0.01	0.12***	-0.57***	-0.11***	-0.02	0.02	0.05**
MaxVote Author $\Delta_s$	-0.24***	0.02	0.03	0.0	-0.03	0.12***	-0.11***	-0.03	0.02	-0.0
MaxVote Author Other	-0.75***	0.1***	0.12***	0.01	-0.11***	0.53***	-0.56***	0.04	0.06**	-0.05**
#Post OP	-0.24***	0.43***	0.22***	0.08***	-0.26***	0.16***	-0.26***	0.15***	0.07***	0.07***
#Post $\Delta_r$	-0.05*	-0.14***	0.0	-0.0	0.0	-0.19***	0.04	0.22***	-0.0	0.05*
#Post $\Delta_s$	-0.39***	0.32***	0.21***	0.04*	-0.22***	0.33***	-0.42***	0.13***	0.07***	0.05**
#Post	0.04*	-0.14***	-0.02	-0.01	0.02	-0.16***	0.1***	0.1***	-0.01	-0.01
#Author	0.2***	-0.24***	-0.11***	-0.04	0.13***	-0.26***	0.25***	0.02	-0.03	-0.04
#Crowd	0.2***	-0.24***	-0.11***	-0.04	0.13***	-0.26***	0.25***	0.02	-0.03	-0.04
#Post Btr $\Delta_r$	-0.68***	0.08***	0.17***	0.04	-0.18***	0.7***	-0.8***	0.13***	0.1***	0.02
#Post Btr $\Delta_s$	0.28***	-0.17***	-0.08***	-0.02	0.09***	-0.41***	0.38***	0.05*	-0.04	-0.02
Time btr $\Delta_s$ /previous post	0.0	-0.07***	-0.03	-0.04*	0.07***	-0.05*	0.03	0.03	-0.01	-0.03
Time btr $\Delta_r$ / $\Delta_s$	0.02	-0.07***	-0.03	-0.05*	0.07***	-0.07***	0.05*	0.03	-0.01	-0.02
Time btr Root/ $\Delta_s$	0.0	-0.08***	-0.02	-0.05*	0.07***	-0.05**	0.03	0.03	-0.01	-0.03

Note: Significance of correlation test of Pearson: "\*\*\*\*" - pvalue < 0.01, "\*\*\*" - pvalue < 0.05, "\*\*" - pvalue < 0.1, "" - non significant.

TABLE 5.4: Correlation matrix  $N = 1590$

	$\Delta_g$ Exp	Crowd Exp	Crowd Exp Diversity	OP Age	$\Delta_r$ Age	$\Delta_g$ Age	Crowd Age	Crowd Age Diversity	Vote root
MaxVote $\Delta$	-0.07***	-0.15***	0.11***	-0.02	0.01	-0.03	-0.06**	0.05*	0.07***
$\Delta_g$ OP	-0.02	0.17***	-0.12***	0.03	0.08***	0.04*	0.11***	0.0	-0.18***
Chatters	0.02	0.02	-0.11***	0.04*	0.0	-0.03	0.03	-0.05**	-0.17***
Orators	0.02	0.06**	-0.04	0.01	0.01	0.01	0.01	-0.01	-0.14***
Overflowed	-0.04	-0.09***	0.14***	-0.05**	-0.01	0.01	-0.04	0.05**	0.29***
Branch	0.1***	0.16***	-0.23***	0.07***	0.01	0.02	0.14***	-0.08***	-0.14***
Side	-0.07***	-0.17***	0.16***	-0.06**	-0.0	-0.03	-0.1***	0.07***	0.11***
Complex	-0.05*	0.02	0.11***	-0.02	-0.01	0.02	-0.06**	0.01	0.04
OP Exp	0.85***	0.14***	-0.06**	0.14***	0.05**	0.08***	0.08***	-0.11***	-0.0
$\Delta_r$ Exp	0.02	0.03	0.13***	0.07***	0.17***	0.07***	0.08***	-0.06**	-0.01
$\Delta_g$ Exp	0.09***	0.09***	-0.04	0.13***	0.04	0.09***	0.07***	-0.09***	0.02
Crowd Exp	0.09***	-0.04	-0.04	0.09***	0.08***	0.08***	0.12	-0.09***	-0.09***
Crowd Exp Diversity	-0.04	-0.04	-0.02	-0.02	0.03	0.03	-0.01	0.03	0.14***
OP Age	0.13***	0.09***	-0.02	0.3***	0.3***	0.33***	0.25***	-0.18***	0.06**
$\Delta_r$ Age	0.04	0.08***	0.03	0.33***	0.3***	0.3***	0.28***	-0.21***	0.09***
$\Delta_g$ Age	0.09***	0.08***	0.03	0.33***	0.3***	0.3***	0.3***	-0.18***	0.09***
Crowd Age	0.07***	0.12***	-0.01	0.25***	0.28***	0.3***	0.3***	-0.18***	0.09***
Crowd Age Diversity	-0.09***	-0.09***	0.09***	-0.18***	-0.21***	-0.18***	-0.3***	-0.3***	0.05**
Vote root	0.02	-0.09***	0.14***	0.06**	0.09***	0.09***	0.05**	0.03	0.03
Vote $\Delta_r$	-0.0	-0.08***	0.15***	0.03	0.04*	0.09***	-0.02	0.05*	0.47***
Vote $\Delta_g$	0.01	-0.04	0.01	0.03	-0.0	0.08***	-0.05*	0.05**	0.36***
Mean vote	0.02	-0.06**	0.05**	0.03	0.03	0.08***	-0.05*	0.03	0.54***
MaxVote Post High	0.02	0.07***	-0.05**	0.04*	-0.0	0.02	0.05**	-0.01	-0.08***
MaxVote Author $\Delta_r$	-0.08***	-0.17***	0.12***	-0.04	-0.01	-0.01	-0.08***	0.06**	0.05**
MaxVote Author $\Delta_g$	0.1***	-0.01	-0.06**	0.02	0.05*	0.0	0.01	-0.03	-0.01
MaxVote Author Other	0.04	0.17***	-0.1***	0.03	-0.01	0.01	0.08***	-0.05*	-0.05**
#Post OP	0.03	0.17***	-0.0	0.11***	0.07***	0.09***	0.14***	0.05*	-0.09***
#Post $\Delta_r$	-0.01	-0.06**	0.17***	0.03	0.01	0.04*	-0.03	0.06**	0.08***
#Post $\Delta_g$	0.08***	0.19***	-0.06**	0.06**	0.04	0.07***	0.14***	0.01	-0.12***
#Post	0.01	-0.03	0.28***	0.08***	0.01	0.09***	0.02	0.12***	0.14***
#Author	-0.01	-0.1***	0.35***	0.04*	-0.01	0.06**	-0.01	0.14***	0.22***
#Crowd	-0.01	-0.1***	0.35***	0.04*	-0.01	0.06**	-0.01	0.14***	0.22***
#Post Bfr $\Delta_r$	0.1***	0.15***	-0.1***	0.09***	-0.01	0.06**	0.13***	-0.05*	-0.07***
#Post Inbtw $\Delta$	-0.03	-0.09***	0.3***	0.04	0.01	0.07***	-0.02	0.13***	0.16***
Time btw $\Delta_g$ /previous post	-0.01	-0.02	-0.0	-0.02	-0.07***	-0.04*	-0.02	0.01	0.03
Time btw $\Delta_r$ / $\Delta_g$	-0.02	-0.02	0.0	-0.02	-0.07***	-0.04*	-0.02	0.02	0.04
Time btw Root/ $\Delta_g$	-0.02	-0.02	0.0	-0.02	-0.08***	-0.05*	-0.02	0.01	0.03

Note: Significance of correlation test of Pearson: "\*\*\*\*" - pvalue < 0.01, "\*\*\*" - pvalue < 0.05, "\*\*" - pvalue < 0.1, "\*" - non significant.

TABLE 5.4: Correlation matrix  $N = 1590$

	Vote $\Delta_r$	Vote $\Delta_g$	Mean vote	MaxVote Post High	MaxVote Author $\Delta_r$	MaxVote Author $\Delta_g$	MaxVote Author Other	#Post OP	#Post $\Delta_r$	#Post $\Delta_g$
MaxVote $\Delta$	0.26***	0.1***	0.07***	0.06**	0.77***	-0.09***	-0.75***	-0.24***	-0.05*	-0.39***
$\Delta_g$ OP	-0.14***	0.02	-0.09***	0.07***	-0.1***	0.02	0.1***	0.43***	-0.14***	0.32***
Chatters	-0.12***	-0.08***	-0.14***	0.03	-0.13***	0.03	0.12***	0.22***	0.0	0.21***
Orators	-0.1***	-0.05**	-0.1***	0.06**	-0.01	0.0	0.01	0.08***	-0.0	0.04*
Overflowed	0.2***	0.13***	0.23***	-0.09***	0.12***	-0.03	-0.11***	-0.226***	0.0	-0.22***
Branch	-0.23***	-0.1***	-0.09***	0.21***	-0.57***	0.12***	0.53***	0.16***	-0.19***	0.33***
Side	0.28***	0.1***	0.09***	-0.27***	0.59***	-0.11***	-0.56***	-0.26***	0.04	-0.42***
Complex	-0.06**	-0.01	0.01	0.08***	0.03	-0.02	0.15***	0.15***	0.22***	0.13***
OP Exp	-0.02	0.0	0.01	0.04	-0.07***	0.02	0.06**	0.07***	-0.0	0.07***
$\Delta_r$ Exp	-0.0	0.01	0.0	-0.0	0.05**	-0.0	-0.05**	0.07***	0.05*	0.05**
$\Delta_g$ Exp	-0.0	0.01	0.02	0.02	-0.08***	0.1***	0.04	0.03	-0.01	0.08***
Crowd Exp	-0.08***	-0.04	-0.06**	0.07***	-0.17***	-0.01	0.17***	0.17***	-0.06**	0.19***
Crowd Exp Diversity	0.15***	0.01	0.05**	-0.05**	0.12***	-0.06**	-0.1***	-0.0	0.17***	-0.06**
OP Age	0.03	0.03	0.03	0.04*	-0.04	0.02	0.03	0.11***	0.03	0.06**
$\Delta_r$ Age	0.04*	-0.0	0.03	-0.0	-0.01	0.05*	-0.01	0.07***	0.01	0.04
$\Delta_g$ Age	0.09***	0.08***	0.08***	0.02	-0.01	0.0	0.08***	0.09***	0.04*	0.07***
Crowd Age	-0.02	-0.05*	-0.04*	0.05**	0.06**	-0.03	-0.05*	0.05*	-0.03	0.14***
Crowd Age Diversity	0.05*	0.05**	0.03	-0.01	0.05**	-0.01	-0.05*	0.05*	0.06**	0.01
Vote root	0.47***	0.36***	0.74***	-0.08***	0.22***	-0.07***	-0.19***	-0.08***	0.07***	-0.12***
Vote $\Delta_r$	0.6***	0.6***	0.77***	-0.06**	0.1***	-0.02	-0.09***	-0.06**	-0.03	-0.09***
Vote $\Delta_g$	0.74***	0.77***	0.77***	-0.12***	0.06**	-0.05**	-0.03	-0.14***	-0.07***	-0.16***
Mean vote	-0.11***	-0.06**	-0.12***	0.05*	0.05*	0.22***	0.14***	0.17***	0.02	0.22***
MaxVote Post High	0.22***	0.1***	0.06**	0.22***	-0.25***	-0.25***	-0.14***	-0.16***	0.14***	-0.3***
MaxVote Author $\Delta_r$	-0.07***	-0.02	-0.05**	-0.14***	-0.91***	0.05**	0.13***	0.05**	0.0	0.13***
MaxVote Author $\Delta_g$	-0.19***	-0.09***	-0.03	-0.14***	-0.16***	-0.17***	-0.17***	0.05**	-0.14***	0.25***
MaxVote Author Other	-0.08***	-0.06**	-0.14***	0.17***	0.14***	0.0	0.14***	0.24***	-0.14***	0.74***
#Post OP	0.07***	-0.03	-0.07***	0.02	0.14***	0.0	-0.07***	0.24***	0.15***	0.15***
#Post $\Delta_r$	-0.19***	-0.09***	-0.16***	0.22***	-0.3***	0.13***	0.25***	0.74***	0.59***	0.19***
#Post $\Delta_g$	0.2***	-0.03	-0.04*	-0.01	0.08***	-0.01	-0.07***	0.4***	0.32***	0.19***
#Author	0.33***	0.01	0.04*	-0.1***	0.16***	-0.06**	-0.13***	0.12***	0.32***	-0.1***
#Crowd	0.33***	0.01	0.04*	-0.1***	0.16***	-0.06**	-0.13***	0.12***	0.32***	-0.1***
#Post Btr $\Delta_r$	-0.24***	-0.1***	-0.11***	0.27***	-0.53***	0.12***	0.49***	0.38***	0.1***	0.6***
#Post Btrw $\Delta$	0.27***	0.01	-0.0	-0.11***	0.26***	-0.06**	-0.24***	0.24***	0.52***	-0.04
Time btrw $\Delta_g$ /previous post	-0.02	-0.03	-0.02	-0.01	-0.0	-0.01	-0.01	-0.08***	-0.01	-0.06**
Time btrw $\Delta_r$ / $\Delta_g$	-0.01	-0.03	-0.02	-0.02	0.02	-0.01	-0.01	-0.07***	0.01	-0.06**
Time btrw Root/ $\Delta_g$	-0.02	-0.03	-0.03	-0.01	-0.0	-0.0	0.0	-0.07***	0.01	-0.05*

Note: Significance of correlation test of Pearson: "\*\*\*\*" - pvalue < 0.01, "\*\*\*" - pvalue < 0.05, "\*\*" - pvalue < 0.1, "" - non significant.

TABLE 5.4: Correlation matrix  $N = 1590$

	#Post	#Author	#Crowd	#Post Bfr $\Delta_r$	#Post Inbtw $\Delta$	Time btw $\Delta_g$ / previous post	Time btw $\Delta_r$ / $\Delta_g$	Time btw Root / $\Delta_g$
MaxVote $\Delta$	0.04*	0.2***	0.2***	-0.68***	0.28***	0.0	0.02	0.0
$\Delta_g$ OP	-0.14***	-0.24***	-0.24***	0.08***	-0.17***	-0.07***	-0.07***	-0.08***
Chatters	-0.02	-0.11***	-0.11***	0.17***	-0.08***	-0.03	-0.03	-0.02
Orators	-0.01	-0.04	-0.04	0.04	-0.02	0.04*	-0.05*	-0.05*
Overflown	0.02	0.13***	0.13***	-0.18***	0.09***	0.07***	0.07***	0.07***
Branch	-0.16***	-0.26***	-0.26***	0.7***	-0.41***	-0.05*	-0.07***	-0.05**
Side	0.1***	0.25***	0.25***	-0.8***	0.38***	0.03	0.05*	0.03
Complex	0.1***	0.02	0.02	0.13***	0.05*	0.03	0.03	0.03
OP Exp	-0.01	-0.03	-0.03	0.1***	-0.04	-0.01	-0.01	-0.01
$\Delta_r$ Exp	-0.01	-0.04	-0.04	0.02	-0.02	-0.03	-0.02	-0.03
$\Delta_g$ Exp	0.01	-0.01	-0.01	0.1***	-0.03	-0.01	-0.02	-0.02
Crowd Exp	-0.03	-0.1***	-0.1***	0.15***	-0.09***	-0.02	-0.02	-0.02
Crowd Exp Diversity	0.28***	0.35***	0.35***	-0.1***	0.3***	-0.0	0.0	0.0
OP Age	0.08***	0.04*	0.04*	0.09***	0.04	-0.02	-0.02	-0.02
$\Delta_r$ Age	0.01	-0.01	-0.01	-0.01	0.01	-0.07***	-0.07***	-0.08***
$\Delta_g$ Age	0.09***	0.06**	0.06**	0.06**	0.07***	-0.04*	-0.04*	-0.05*
Crowd Age	0.02	-0.01	-0.01	0.13***	-0.02	-0.02	-0.02	-0.02
Crowd Age Diversity	0.12***	0.14***	0.14***	-0.05*	0.13***	0.01	0.02	0.01
Vote root	0.14***	0.22***	0.22***	-0.07***	0.16***	0.03	0.04	0.03
Vote $\Delta_r$	0.2***	0.33***	0.33***	-0.24***	0.27***	-0.02	-0.01	-0.02
Vote $\Delta_g$	-0.03	0.01	0.01	-0.1***	0.01	-0.03	-0.03	-0.03
Mean vote	-0.04*	0.04*	0.04*	-0.11***	-0.0	-0.02	-0.02	-0.03
MaxVote Post High	-0.01	-0.1***	-0.1***	0.27***	-0.11***	-0.01	-0.02	-0.01
MaxVote Author $\Delta_r$	0.08***	0.16***	0.16***	-0.53***	0.26***	-0.0	0.02	-0.0
MaxVote Author $\Delta_g$	-0.01	-0.06**	-0.06**	0.12***	-0.06**	-0.01	-0.01	-0.0
MaxVote Author Other	-0.07***	-0.13***	-0.13***	0.49***	-0.24***	0.01	-0.01	0.0
#Post OP	0.4***	0.12***	0.12***	0.38***	0.24***	-0.08***	-0.07***	-0.07***
#Post $\Delta_r$	0.59***	0.32***	0.32***	0.1***	0.52***	-0.01	0.01	0.01
#Post $\Delta_g$	0.15***	-0.1***	-0.1***	0.6***	-0.04	-0.06**	-0.06**	-0.05*
#Post		0.86***	0.86***	0.03	0.93***	-0.02	-0.0	-0.0
#Author	0.86***	1.0***	1.0***	-0.19***	0.88***	0.02	0.02	0.01
#Crowd	0.86***	1.0***	1.0***	-0.19***	0.88***	0.02	0.02	0.01
#Post Bfr $\Delta_r$	0.03	-0.19***	-0.19***	-0.33***	-0.33***	-0.05**	-0.07***	-0.05*
#Post Inbtw $\Delta$	0.93***	0.88***	0.88***	0.0	0.0	1.0***	1.0***	0.99***
Time btw $\Delta_g$ / previous post	-0.02	-0.0	-0.0	-0.05**	0.0	0.03	0.03	0.02
Time btw $\Delta_r$ / $\Delta_g$	-0.0	0.02	0.02	-0.07***	0.03	1.0***	1.0***	0.99***
Time btw Root / $\Delta_g$	-0.0	0.01	0.01	-0.05*	0.02	0.99***	0.99***	0.99***

Note: Significance of correlation test of Pearson: "\*\*\*\*" - pvalue < 0.01, "\*\*\*" - pvalue < 0.05, "\*\*" - pvalue < 0.1, "\*" - non significant.

TABLE 5.4: Correlation matrix  $N = 1590$

	Max Vote $\Delta$	$\Delta_{\delta}$ OP	Cluster	OP Exp	$\Delta_r$ Exp	$\Delta_{\delta}$ Exp	Crowd Exp	Crowd Exp Diversity	OP Age	$\Delta_r$ Age	$\Delta_{\delta}$ Age
Max Vote $\Delta$	-0.09***	-0.09***	0.21***	-0.07***	0.0	-0.07***	-0.15***	0.11***	-0.02	0.01	-0.03
$\Delta_{\delta}$ OP	0.21***	-0.25***	-0.25***	0.06**	0.1***	-0.02	0.17***	-0.12***	0.03	0.08***	0.08***
Cluster	-0.07***	0.06**	-0.07***	-0.07***	-0.01	-0.04	-0.07***	0.16***	-0.06**	-0.01	0.08***
OP Exp	0.0	-0.07***	-0.01	0.0	0.0	0.02	0.03	-0.01	0.05**	0.08***	0.08***
$\Delta_r$ Exp	0.0	0.1***	-0.01	0.0	0.0	0.02	0.03	0.09***	0.07***	0.09***	0.09***
$\Delta_{\delta}$ Exp	-0.07***	-0.02	-0.04	0.85***	0.02	0.09***	0.09***	-0.04	-0.04	-0.04	-0.02
Crowd Exp	-0.15***	0.17***	-0.07***	0.14***	0.03	0.09***	0.09***	-0.04	0.09***	-0.02	-0.02
Crowd Exp Diversity	0.11***	-0.12***	0.16***	-0.06**	0.13***	-0.04	-0.04	0.09***	-0.02	0.03	0.33***
OP Age	-0.02	0.03	-0.06**	0.14***	0.07***	0.13***	0.08***	-0.02	0.03	0.33***	0.33***
$\Delta_r$ Age	0.01	0.08***	-0.01	0.05**	0.17***	0.04	0.08***	0.03	0.03	0.33***	0.33***
$\Delta_{\delta}$ Age	-0.03	0.04*	0.02	0.08***	0.07***	0.09***	0.08***	0.03	0.33***	0.33***	0.33***
Crowd Age	-0.06**	0.11***	-0.05*	0.08***	0.08***	0.07***	0.12***	-0.01	0.25***	0.28***	0.28***
Crowd Age Diversity	0.05*	0.0	0.06**	-0.11***	-0.06**	-0.09***	-0.09***	0.09***	-0.18***	-0.21***	-0.18***
Mean Vote	0.07***	0.0	0.23***	0.01	0.0	0.02	-0.06**	0.05**	0.03	0.03	0.08***
MaxVote High	0.06**	0.07***	-0.08***	0.07***	-0.0	0.02	0.07***	-0.05**	0.04*	0.02	0.02
MaxVote Auth is $\Delta$ Auth	0.75***	-0.1***	0.14***	-0.06**	0.05**	-0.04	-0.17***	0.1***	-0.03	0.01	-0.01
OP Post/Posts	-0.25***	0.69***	-0.36***	0.07***	0.08***	0.02	0.24***	-0.2***	0.08***	0.09***	0.06**
$\Delta_r$ Post/Posts	-0.0	-0.06**	0.01	-0.02	0.07***	-0.04*	-0.11***	-0.11***	-0.04*	-0.01	-0.01
$\Delta_{\delta}$ Post/Posts	-0.39***	0.47***	-0.3***	0.06***	0.04	0.07***	0.24***	-0.31***	0.05**	0.04	0.04
Bfr $\Delta_r$ Post/Crowd Posts	-0.71***	0.13***	-0.26***	0.09***	-0.0	0.09***	0.18***	-0.22***	0.07***	0.0	0.03
Inbtw $\Delta$ Posts/Crowd Posts	0.71***	-0.13***	0.26***	-0.09***	0.0	-0.09***	-0.18***	0.22***	-0.07***	-0.0	-0.03
Crowd size/Authors	0.23***	-0.32***	0.22***	-0.04	-0.05*	-0.03	-0.13***	0.42***	-0.01	-0.02	0.01

Note: Significance of correlation test of Pearson: "\*\*\*" - pvalue < 0.01, "\*\*" - pvalue < 0.05, "\*" - pvalue < 0.1, "-" non significant.

TABLE 5.6: Correlation matrix for regression  $N = 1590$

	Crowd Age	Crowd Age Diversity	Mean Vote	MaxVote High	MaxVote Auth is $\Delta$ Auth	OP Post Posts	$\Delta_r$ Post/ Posts	$\Delta_g$ Post/ Posts	Bfr $\Delta_r$ Post/ Crowd Posts	Inbtw $\Delta$ Posts/ Crowd Posts	Crowd size/ Authors
Max Vote $\Delta$	-0.06**	0.05*	0.07***	0.06**	0.75***	-0.25***	-0.0	-0.39***	-0.71***	0.71***	0.23***
$\Delta_g$ OP	0.11***	0.0	-0.09***	0.07***	-0.1***	0.69***	-0.06**	0.47***	0.13***	-0.13***	-0.32***
Cluster	-0.05*	0.06**	0.23***	-0.08***	0.14***	-0.36***	0.01	-0.3***	-0.26***	0.26***	0.22***
OP Exp	0.08***	-0.11***	0.01	0.04	-0.06**	0.07***	-0.02	0.06***	0.09***	-0.09***	-0.04
$\Delta_r$ Exp	0.08***	-0.06**	0.0	-0.0	0.05**	0.08***	0.07***	0.04	-0.0	0.0	-0.05*
$\Delta_g$ Exp	0.07***	-0.09***	0.02	0.02	-0.04	0.02	-0.04*	0.07***	0.09***	-0.09***	-0.03
Crowd Exp	0.12***	-0.09***	-0.06**	0.07***	-0.17***	0.24***	-0.11***	0.24***	0.18***	-0.18***	-0.13***
Crowd Exp Diversity	-0.01	0.09***	0.05**	-0.05**	0.1***	-0.2***	-0.11***	-0.31***	-0.22***	0.22***	0.42***
OP Age	0.25***	-0.18***	0.03	0.04*	-0.03	0.08***	-0.04*	0.05**	0.07***	-0.07***	-0.01
$\Delta_r$ Age	0.28***	-0.21***	0.03	-0.0	0.01	0.09***	-0.01	0.04	0.0	-0.0	-0.02
$\Delta_g$ Age	0.3***	-0.18***	0.08***	0.02	-0.01	0.06**	-0.01	0.04	0.03	-0.03	0.01
Crowd Age	-0.3***	-0.3***	-0.04*	0.05**	-0.08***	0.16***	-0.08***	0.16***	0.13***	-0.13***	-0.05**
Crowd Age Diversity	-0.04*	0.03	0.03	-0.12***	0.05*	-0.03	-0.08***	-0.09***	-0.08***	0.08***	0.18***
Mean Vote	0.05**	0.03	-0.12***	0.14***	0.03	-0.13***	0.0	-0.13***	-0.1***	0.1***	0.07***
MaxVote High	-0.08***	0.05*	0.03	0.14***	0.03	0.18***	-0.02	0.2***	0.26***	-0.26***	-0.12***
MaxVote Auth is $\Delta$ Auth	0.16***	-0.03	-0.13***	0.18***	-0.22***	-0.22***	0.22***	-0.34***	-0.58***	0.58***	0.14***
$\Delta_r$ Post/Posts	-0.08***	-0.08***	0.0	-0.02	0.22***	-0.16***	-0.16***	0.71***	0.39***	-0.39***	-0.47***
$\Delta_g$ Post/Posts	0.16***	-0.09***	-0.13***	0.2***	0.22***	-0.16***	-0.21***	-0.21***	-0.09***	0.09***	-0.4***
Bfr $\Delta_r$ Post/Crowd Posts	0.13***	-0.08***	-0.1***	0.26***	-0.58***	0.39***	-0.09***	0.58***	-0.58***	-0.58***	-0.63***
Inbtw $\Delta$ Posts/Crowd Posts	-0.13***	0.08***	0.1***	-0.26***	0.58***	-0.39***	0.09***	-0.58***	-1.0***	-1.0***	-0.39***
Crowd size/ Authors	-0.05**	0.18***	0.07***	-0.12***	0.14***	-0.47***	-0.4***	-0.63***	-0.39***	0.39***	0.39***

Note: Significance of correlation test of Pearson: "\*\*\*\*" - pvalue < 0.01, "\*\*\*" - pvalue < 0.05, "\*\*" - pvalue < 0.1, "\*" - non significant.

TABLE 5.6: Correlation matrix for regression N = 1590

	Dependent variable:								
	MaxVote $\Delta = 1$								
	Convergence								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Chatters vs. Orators	-0.78*** (0.16)						-0.29 (0.25)	-0.32 (0.25)	-0.35 (0.25)
Overflowed vs. Orators	0.59*** (0.12)						0.08 (0.19)	0.07 (0.19)	0.04 (0.20)
$\Delta_g$ OP		-0.28** (0.12)	-0.64*** (0.15)	-0.24 (0.39)	-0.32 (0.38)	-0.21 (0.39)	-0.28 (0.45)	-0.38 (0.45)	-0.22 (0.46)
Exp $\Delta_r$		0.0001 (0.001)					0.001 (0.001)	0.001 (0.002)	0.001 (0.002)
Exp $\Delta_g$		-0.01* (0.004)					-0.001 (0.01)	-0.0001 (0.01)	-0.002 (0.01)
$\Delta_g$ OP x Exp $\Delta_g$		0.003 (0.005)					0.0000 (0.01)	-0.001 (0.01)	0.001 (0.01)
Exp Crowd		-0.01*** (0.002)					-0.01** (0.002)	-0.01** (0.003)	-0.01** (0.003)
Exp Crowd Diversity		1.44*** (0.36)					-2.22*** (0.60)	-1.32** (0.63)	-1.48** (0.63)
Age $\Delta_r$			0.01 (0.005)				0.003 (0.01)	0.002 (0.01)	0.002 (0.01)
Age $\Delta_g$			-0.02** (0.01)				-0.01 (0.01)	-0.01 (0.01)	-0.02 (0.02)
$\Delta_g$ OP x Age $\Delta_g$			0.03** (0.01)				0.01 (0.02)	0.01 (0.02)	0.01 (0.02)
Age Crowd			-0.01 (0.01)				0.01 (0.01)	0.01 (0.01)	0.01 (0.01)
Age Crowd Diversity			0.46 (0.30)				-0.39 (0.47)	0.01 (0.48)	-0.11 (0.49)
Vote Mean							0.01 (0.004)	0.005 (0.004)	0.01 (0.004)
MaxVote Post High							1.08*** (0.09)	1.08*** (0.09)	1.10*** (0.09)
Post $\Delta_r$ / #Posts				-2.18*** (0.72)	-6.49*** (1.01)	-6.46*** (1.01)	-3.25*** (0.82)	-7.77*** (1.13)	-7.80*** (1.15)
Post $\Delta_g$ / #Posts				0.85 (1.02)	-3.26*** (1.24)	-2.79** (1.23)	0.10 (1.23)	-4.22*** (1.47)	-3.33** (1.47)
$\Delta_g$ OP x Post $\Delta_g$ / #Posts				0.31 (1.17)	0.70 (1.19)	0.16 (1.20)	0.35 (1.38)	0.81 (1.40)	-0.23 (1.43)
Inbtw $\Delta$ Posts / #Crowd Posts				4.19*** (0.21)	4.39*** (0.22)	4.47*** (0.23)	5.42*** (0.27)	5.56*** (0.28)	5.68*** (0.29)
Crowd Size / #Authors					-5.76*** (0.92)			-6.25*** (1.06)	
Crowd Size = 1 (Ref = 2)						1.38*** (0.24)			1.61*** (0.26)
Crowd Size = 3 (Ref = 2)						-0.80*** (0.29)			-0.86*** (0.31)
Crowd Size = 4 (Ref = 2)						-0.45 (0.41)			-0.24 (0.45)
Crowd Size $\geq 5$ (Ref = 2)						-0.60 (0.42)			-0.52 (0.46)
Constant	-0.14** (0.07)	-0.56** (0.24)	0.23 (0.18)	-1.76*** (0.41)	3.32*** (0.90)	-0.02 (0.56)	-2.10*** (0.72)	2.64** (1.07)	-0.96 (0.81)
Observations	1,590	1,590	1,590	1,590	1,590	1,590	1,590	1,590	1,590
Log Likelihood	-1,065.59	-1,065.72	-1,086.29	-641.94	-622.02	-613.68	-535.44	-517.28	-506.01
McFadden's Pseudo- $R^2$	0.0316	0.0315	0.0128	0.4166	0.4347	0.4423	0.5134	0.5299	0.5402
Akaike Inf. Crit.	2,137.17	2,145.44	2,186.58	1,295.87	1,258.05	1,247.36	1,110.87	1,076.56	1,060.02

Note:

\* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$ TABLE 5.7: Regression results  $N = 1590$

## Chapter 6: Conclusion

Chapter 2 presents fundamental research about concepts at stake. First, we presented the evolution what embody a **crowd**, before Internet, and since its introduction in our daily life. Second, we present the motivation for an individual to participate on such online exchange. Third, we extensively present core concepts of this PhD work: the Wisdom of the Crowd, and closed concept of Swarm Intelligence and Collective Intelligence. The main questions of the current research is how the whole is more than the sum of its parts? How human behave together toward the same goal? Those questions already arised in Antiquity, and are a part of: how human society hold together through the political regime of democracy? We then present the major schools of thought in this area. Finally, we present the adding-values to society for a crowd to work together.

Chapter 3 presents the case study: Reddit - Change My View (CMV). CMV is a subreddit, constraint by the same rule than Reddit. Contrary to social networks, in Reddit, users follow **contents**, and not accounts. They can collectively put forward contents by up-voting it, which make it appear in the home page. Or they can bury a content by down-voting it. Furthermore, discussions arise about those contents. When Kal Turnbull created CMV in 2013, he observed:

*Conversations on the internet are broken.* — Kal Turnbull, founder of *Change My View*, 2013.

His idea then was to develop a space where friendly debates occure, where

people coming from all over the world would fill up his mind with perspectives he couldn't have in his Scottish Highlands. Then he dedicated an important part of his time to open this space on Reddit. In order to do so, he settled ground rules to debate. One individual, the Original Poster - OP, presents his idea to the community, looking for arguments which make him *change his view*. The first rule then is the will to change its own view. Secondly, the OP agrees to engage in a discussion for at least three hours. This ensure a time of lively exchange among participants, and diminish the risk for one to give up his discussion. The third rule concerns argumentation itself: in order to participate, you have to argument in detail and precisely what is your point, both for change a view or acknowledge a change in view. Besides the vote system presents in Reddit, Change My View added an individual reward system: the delta  $\Delta$ . One acknowledges, with arguments, that someone post has change his view, at least a few. On the one hand, this allows to highlight relevant post/argument for the discussion. On the other, by not putting an end to the discussion, it allows it to pursue and new argument to appear.

Through this constraint rules, Kal Turnbull defines a proper application to the *third space* framed by Wright (2012). By construction, CMV is a proper case study to analyze how members of a same society could exchange, freely, and convince each other for the common good.

Our first inquiry is to understand at which step of its development the CMV community is. The entirety of the empirical study of this PhD thesis is based on a database coming from this subreddit, extracted by Tan et al. (2016)<sup>1</sup>. The database covers threads from January 2013, date of the subreddit creation, to November 2016. It then covers different steps of online community evolution, as defined by Iriberry and Leroy (2009). The first months covers both the *creation* and *growth* phases. The next years are cycle of *maturity* phase. This phase matches with a need for a more formal organization. This is when new features are developed such as a reward system, subgroups. Furthermore, new members arrive, old left: the community is

---

<sup>1</sup>An update version of the database, used in the current PhD thesis, can be found here: <https://chenhaot.com/papers/changemyview.html>

constantly moving and managers of the forum have to integrate those newcomers or the community does not survive and *die* due to a lack of participation Jarvenpaa, Knoll, and Leidner (1998).

Chapter 4 challenges Kal Turnbull's dream when he developed CMV. Has he succeed to create a third space opens to discussion, allowing anyone to exchange with any fellow members of the community, attracting interest from the community about a subject and not about himself, contrary to social networks where members usually follow someone. We study the heterogeneity in coordination modes for discussions inside the forum, where rules are defined to frame one unique way to coordinate. In order to do so, we apply different clustering algorithms to distinguish different coordination modes. Thus, we model a discussion as a rooted tree (type of network), extracting the corresponding graph whose vertices model authors. Then, we apply the clustering algorithms on particular sub-network called motifs Milo et al. (2002). The different clustering methods deliver a robust result: we observe , two extreme collaborative resolution emerged, a third one appears to be in between. Furthermore, this clustering, based only on network characteristics, highlights authors characteristics.

The "Chatters" gather discussions where OP answers to almost all his challenging authors, where authors are old and experienced in CMV. We conclude that discussions within the "Chatters" are dense, challenging and any new subject is addressed by the OP. The "Overflowed" gather discussions with opposed characteristics to those from the "Chatters". Indeed, discussions attract numerous challenging authors, last a long time and OP answers to a little proportion of his challengers. Finally authors are less experienced than those from "Chatters". We conclude that those discussions escape from the lead of the OP, who is overflowed by the number of challengers. Numerous sub-topic are not addressed, global discussion is less dense. Inbetween, we observe the "Orators". It gathers more than the half of the discussions in study, and discussions characteristics always fall between discussions characteristics of the "Chatters" and the "Overflowed". Thus we conclude that we lack of informations to properly distinguish behaviors among this cluster, which

gathers heterogeneous discussions.

The use of motifs in the author networks of online discussions is an innovative methodology to extract information on the structure of discussions which leads toward a better comprehension of online conversations evolution. This analysis is a first work toward a monitoring tool for CMV moderators. If they aim for dense discussion, where each argument is tackle, then they could settle a number maximum of challenging author, in order to stay close to a "Chatter" discussions characteristics. In this case, the OP would be able to adress all the subject.

Chapter 5 focuses on discussion which seem to work, in the meaning that  $\Delta$  have been rewarded, relevant arguments have been made and thus a level of consensus has been reached, at least between the one receiving the  $\Delta$  and the one awarding it. Reddit - Change My View offers a dual reward system. On the one hand, the  $\Delta$  which an participant award another one. On the other, the vote system, allowing anyone to up-vote a post. Even if it is still an individual action, such as the  $\Delta$ , others members observe the total of vote, and so a collective action, contrary to the  $\Delta$  which is rewarded by a unique author, labelled as such to the community. In the current chapter, we show that the crowd can notably consent, and sometimes dissent with someone, who often has a prominent role in online fora, such as the moderator, the person who has asked the initial question, or a salient member of the community.

We defined as the *Consent of the Crowd* the power of a crowd in a open online forum to shine a light on a collective result, rather than an individual one. Focusing on a particular period of the discussion, from the beginning up to the first reward, our results bring in evidence the use of the upvote reward system as a punishment against the challenger rewarded with the  $\Delta$ . This happened when crowd members do not get the chance to join the discussion before the  $\Delta$  attribution. From the research point of view, we explain how a crowd takes actions if it perceived an injustice. Moreover, this result should urge CMV moderators to foster multi-challengers sub-discussions.

Our main results from Chapters 4 and 5 provide the managing team with

enhancements in order to improve the discussion framework defined by Kal Turnbull. First, in order to help the original poster to follow the discussion he opened, they should put a maximum number of challengers to join in. The OP is, by construction, the only one to manage the discussion. As long as the forum need someone to change his view, it should allow him to do so. Second, we defend the strength of the many over the one. Even if the goal of a CMV thread is to change the OP view, he might select *bad* argument. The crowd has a role of gatekeeper on the quality of the discussion. If the managing team aim for a collective confirmation of a *good* argument, it must let time to the crowd to participate to the discussion. Theory would suggest that the biggest the crowd is, the better the discussion would be. Nevertheless, those two results seem contradictory. Actually, it depends which goal we pursue. If the main objective is to offer a proper space to the maximum of individuals, in order to let them express their idea, and have a proper discussion about it, we do not need the OP to follow all his challengers. New opposition could arise among challengers themselves. But we still need the crowd to vote in order to signal proper argument, unbiased one. If the goal is to offer to one individual a space where he could challenge himself to a crowd, then this crowd must be not too big. Furthermore, if the goal is to achieve to a consensus, where all members agreed on the problem and its resolution, then we need the Consent of the Crowd to operate. And in order to do so, settle a minimum amount of time, or similar rule, in order to let the crowd express its idea on the problem, before letting anyone award an individual reward, here the  $\Delta$ .

Nevertheless, this research has its own limits. Chapter 3 provides a temporal analysis of CMV evolution. It allows us to understand at which step of community evolution CMV is. But through the network analysis, framework chosen for this research, we have not taken into account this temporal evolution. Research in the field still lack of results to properly apply a temporal network analysis, even if first attempts have been made (Tang et al., 2009; Blonder et al., 2012; Holme, 2015).

Especially within Chapter 5 we define CMV as a democratic space where

consensus is a must have. This assumption is strong and must be balanced. As depicted previously, must the consensus be the goal of every discussion? Kal Turnbull wanted, above all, a space where he can have argument to change his view. But actually, what is the most important? The possibility to have a strong argument discussion, about any subject, without being attacked by malicious users (Binns, 2012). Or to reach a resolution to the discussion, at all cost. Even if we labelled the action of the crowd to support the  $\Delta$  reward "the consent of the crowd", it is based upon the reward system of Reddit. It has been pointed out with a lack of transparency (Pardes, 2019). Indeed, someone could upvote a post he agrees with, and vice versa. This system, by construction, does not take into account the nuance of the debate. Nor the  $\Delta$  reward does. As detailed in its presentation, a  $\Delta$  reflects a change of view from a change of an inch to a complete reverse point of view. No system could highlight that.

This reflects the lack of semantic analysis by the forum itself. And by the current work. A whole part of the research focuses not only on how something is said, such as the current work, but what it is said. In Musi (2018) and Musi, Ghosh, and Muresan (2018), Elena Musi offers in depth semantic analysis, on CMV discussions, to better understand the function of argumentation strategies in order to convince the other participant, which is visible through the award of a  $\Delta$ .

Future works should be interested to go further on both approaches. On the one hand, introduce temporal network analysis to improve the results observed in this work. Furthermore, research would benefit from a comparative analysis of winning and losing argument. Chapter 5 focuses on the question of consensus, but what characteristics distinguish a post which receives a  $\Delta$  with one which does not. On the other hand, semantic approach is promising and needs to be developed in order to help the comprehension of the mechanism behind the discussion process.

# Bibliography

- Adamic, Lada A. et al. (2008). "Knowledge sharing and Yahoo Answers: Everyone knows something". In: *Proceedings of WWW'08*. ACM, pp. 665–674.
- Adamic, Lada A. et al. (Feb. 2010). "Individual focus and knowledge contribution". In: *arXiv:1002.0561 [cs]*. arXiv: 1002.0561. URL: <http://arxiv.org/abs/1002.0561> (visited on 04/05/2018).
- Afuah, Allan and Christopher L. Tucci (July 2012). "Crowdsourcing As a Solution to Distant Search". en. In: *Academy of Management Review* 37.3, pp. 355–375. ISSN: 0363-7425, 1930-3807. DOI: [10.5465/amr.2010.0146](https://doi.org/10.5465/amr.2010.0146). URL: <http://amr.aom.org/cgi/doi/10.5465/amr.2010.0146> (visited on 02/14/2018).
- Aggarwal, Ishani and Anita Williams Woolley (Sept. 2013). "Do you see what I see? The effect of members' cognitive styles on team processes and errors in task execution". In: *Organizational Behavior and Human Decision Processes* 122.1, pp. 92–99. ISSN: 0749-5978. DOI: [10.1016/j.obhdp.2013.04.003](https://doi.org/10.1016/j.obhdp.2013.04.003). URL: <http://www.sciencedirect.com/science/article/pii/S074959781300054X> (visited on 04/19/2018).
- Aggarwal, Ishani et al. (2015). "Cognitive Diversity, Collective Intelligence, and Learning in Teams". en. In: p. 4.
- Ahituv, Niv and Seev Neumann (1986). *Principles of Information Systems for Management (2Nd Ed.)* Dubuque, IA, USA: William C. Brown Publishers. ISBN: 978-0-697-08267-1.

- Akaike, H. (Dec. 1974). "A new look at the statistical model identification". en. In: *IEEE Transactions on Automatic Control* 19.6, pp. 716–723. ISSN: 0018-9286. DOI: [10.1109/TAC.1974.1100705](https://doi.org/10.1109/TAC.1974.1100705). URL: <http://ieeexplore.ieee.org/document/1100705/> (visited on 04/15/2020).
- Akerlof, George A. (Aug. 1970). "The Market for "Lemons": Quality Uncertainty and the Market Mechanism". en. In: *The Quarterly Journal of Economics* 84.3, pp. 488–500. ISSN: 0033-5533. DOI: [10.2307/1879431](https://doi.org/10.2307/1879431). URL: <https://academic.oup.com/qje/article/84/3/488/1896241> (visited on 08/17/2018).
- Aliakbary, Sadegh, Jafar Habibi, and Ali Movaghar (July 2014). "Feature Extraction from Degree Distribution for Comparison and Analysis of Complex Networks". In: *arXiv:1407.3386 [cs]*. arXiv: 1407.3386. URL: <http://arxiv.org/abs/1407.3386> (visited on 09/21/2015).
- Allcott, Hunt and Matthew Gentzkow (May 2017). "Social Media and Fake News in the 2016 Election". en. In: *Journal of Economic Perspectives* 31.2, pp. 211–236. ISSN: 0895-3309. DOI: [10.1257/jep.31.2.211](https://doi.org/10.1257/jep.31.2.211). URL: <https://www.aeaweb.org/articles?id=10.1257/jep.31.2.211> (visited on 03/01/2019).
- Aman, Mohammed M. and Tina J. Jayroe (2013). "ICT, Social Media, and the Arab Transition to Democracy: From Venting to Acting". en. In: *Digest of Middle East Studies* 22.2, pp. 317–347. ISSN: 1949-3606. DOI: [10.1111/dome.12024](https://doi.org/10.1111/dome.12024). URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/dome.12024> (visited on 03/01/2019).
- Anderson, Ashton et al. (2012). "Discovering value from community activity on focused question answering sites: a case study of stack overflow". In: *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, pp. 850–858. URL: <http://dl.acm.org/citation.cfm?id=2339665> (visited on 09/10/2015).
- Anderson, Monica et al. (July 2018). *Activism in the social media age*. en-US. URL: <http://www.pewinternet.org/2018/07/11/public-attitudes-toward-political-engagement-on-social-media/> (visited on 08/16/2018).

- Andrews, D., J. Preece, and M. Turoff (Jan. 2001). "A conceptual framework for demographic groups resistant to online community interaction". In: *Proceedings of the 34th Annual Hawaii International Conference on System Sciences*, 10 pp.–. DOI: [10.1109/HICSS.2001.927044](https://doi.org/10.1109/HICSS.2001.927044).
- Andrews, Dorine (2001). "Considerations in the Development of Commercially Based Online Communities". en. In: p. 7.
- Aral, Sinan, Erik Brynjolfsson, and Erik Brynjolfsson (2006). "Network Structure & Information Advantage: Structural Determinants of Access to Novel Information". In:
- Aristote (2007). *Éthique à Eudème*. Translated from the Greek. Trans. by Vianney Décarie and Renée Houde-Sauvé. OCLC: 654775447. Paris: J. Vrin. ISBN: 978-2-7116-0025-0.
- Aristotle (1944). *Politics*. Trans. by H. Rackham. Vol. 21. Loeb classical library 264. Cambridge, Mass: Harvard University Press. ISBN: 978-0-674-99291-7.
- Arrow, Kenneth J. (June 2012). *Social Choice and Individual Values: Third Edition*. en. Google-Books-ID: y\_rkX6QWOYMC. Yale University Press. ISBN: 978-0-300-18698-7.
- Arthur, David and Sergei Vassilvitskii (2007). "k-means++: The Advantages of Careful Seeding". en. In: p. 9.
- Asch, Solomon E. (1951). "Effects of group pressure upon the modification and distortion of judgments". In: *Groups, Leadership, and Men*. Ed. by H Guetzkow. Carnegie Press, pp. 177–190.
- Ashford, James R. et al. (Apr. 2019). "Understanding the Signature of Controversial Wikipedia Articles through Motifs in Editor Revision Networks". en. In: *arXiv:1904.08139 [cs]*. arXiv: 1904.08139. URL: <http://arxiv.org/abs/1904.08139> (visited on 02/26/2020).
- Awazu, Yukika and Kevin C. Desouza (Sept. 2004). "Open knowledge management: Lessons from the open source revolution". en. In: *Journal of the American Society for Information Science and Technology* 55.11, pp. 1016–1019. ISSN: 1532-2882, 1532-2890. DOI: [10.1002/asi.20050](https://doi.org/10.1002/asi.20050). URL: <http://doi.wiley.com/10.1002/asi.20050> (visited on 06/21/2017).

- Bail, Christopher Andrew (Sept. 2016). "Combing natural language processing and network analysis to examine how advocacy organizations stimulate conversation on social media". en. In: *Proceedings of the National Academy of Sciences*, p. 201607151. ISSN: 0027-8424, 1091-6490. DOI: [10.1073/pnas.1607151113](https://doi.org/10.1073/pnas.1607151113). URL: <http://www.pnas.org/lookup/doi/10.1073/pnas.1607151113> (visited on 10/05/2016).
- Barabási, Albert-László and Réka Albert (Oct. 1999). "Emergence of Scaling in Random Networks". en. In: *Science* 286.5439, pp. 509–512. ISSN: 0036-8075, 1095-9203. DOI: [10.1126/science.286.5439.509](https://doi.org/10.1126/science.286.5439.509). URL: <http://www.sciencemag.org/content/286/5439/509> (visited on 09/10/2015).
- Bartlett, Jamie and Heather Grabbe (2015). "e-democracy in the EU: the opportunities for digital politics to re-engage voters and the risks of disappointment". en. In: p. 47.
- Baumol, William J. (2004). "Welfare Economics and the Theory of the State". en. In: *The Encyclopedia of Public Choice*. Ed. by Charles K. Rowley and Friedrich Schneider. Boston, MA: Springer US, pp. 937–940. ISBN: 978-0-306-47828-4. DOI: [10.1007/978-0-306-47828-4\\_214](https://doi.org/10.1007/978-0-306-47828-4_214). URL: [https://doi.org/10.1007/978-0-306-47828-4\\_214](https://doi.org/10.1007/978-0-306-47828-4_214) (visited on 06/19/2020).
- Becker, Gary S. (Oct. 1962). "Investment in Human Capital: A Theoretical Analysis". In: *Journal of Political Economy* 70.5, Part 2, pp. 9–49. ISSN: 0022-3808. DOI: [10.1086/258724](https://doi.org/10.1086/258724). URL: <https://www.journals.uchicago.edu/doi/abs/10.1086/258724> (visited on 02/15/2019).
- Becker, Joshua, Devon Brackbill, and Damon Centola (June 2017). "Network dynamics of social influence in the wisdom of crowds". en. In: *Proceedings of the National Academy of Sciences*, p. 201615978. ISSN: 0027-8424, 1091-6490. DOI: [10.1073/pnas.1615978114](https://doi.org/10.1073/pnas.1615978114). URL: <http://www.pnas.org/lookup/doi/10.1073/pnas.1615978114> (visited on 01/03/2018).
- Beenen, Gerard et al. (Nov. 2004). "Using social psychology to motivate contributions to online communities". In: *Proceedings of the 2004 ACM conference on Computer supported cooperative work*. CSCW '04. Chicago, Illinois,

- USA: Association for Computing Machinery, pp. 212–221. ISBN: 978-1-58113-810-8. DOI: [10.1145/1031607.1031642](https://doi.org/10.1145/1031607.1031642). URL: <https://doi.org/10.1145/1031607.1031642> (visited on 04/27/2020).
- Ben-Naim, E. and P. L. Krapivsky (May 2009). “Stratification in the Preferential Attachment Network”. In:
- Benghozi, Pierre-Jean (2006). “Les communautés virtuelles: structuration sociale ou outil de gestion?” In: *Entreprises et histoire* 2, pp. 67–81. URL: <http://www.cairn.info/revue-entreprises-et-histoire-2006-2-page-67.htm> (visited on 06/21/2017).
- Benghozi, Pierre-Jean et al. (2001). “Le mode communautaire : vers une nouvelle forme d’organisation”. In: *3ème Colloque ICUST e-usages*. Paris, France. URL: <https://hal.archives-ouvertes.fr/hal-00262785>.
- Benkler, Yochai (2016). “Peer production and cooperation”. en. In: *Handbook on the Economics of the Internet*. Ed. by Johannes Bauer and Michael Latzer. Edward Elgar Publishing. ISBN: 978-0-85793-985-2. DOI: [10.4337/9780857939852](https://doi.org/10.4337/9780857939852). URL: <http://www.elgaronline.com/view/9780857939845.xml> (visited on 04/24/2018).
- Beshers, Samuel N. and Jennifer H. Fewell (2001). “Models of Division of Labor in Social Insects”. In: *Annual Review of Entomology* 46.1. \_eprint: <https://doi.org/10.1146/annurev.ento.46.1.413>, pp. 413–440. DOI: [10.1146/annurev.ento.46.1.413](https://doi.org/10.1146/annurev.ento.46.1.413). URL: <https://doi.org/10.1146/annurev.ento.46.1.413> (visited on 04/29/2020).
- Besten, Matthijs den, Jean-Michel Dalle, and Fabrice Galia (2008). “The Allocation of Collaborative Efforts in Open-Source Software”. In: *Information Economics and Policy* 20, pp. 316–322. URL: <http://works.bepress.com/mldb/2> (visited on 04/09/2015).
- Besten, Matthijs den et al. (Sept. 2010). “Using Metadata Signals to Support Stigmergy”. In: *2010 Fourth IEEE International Conference on Self-Adaptive and Self-Organizing Systems Workshop*, pp. 131–135. DOI: [10.1109/SASOW.2010.28](https://doi.org/10.1109/SASOW.2010.28).
- Bialik, Kristen (Aug. 2018). *14% of Americans have changed their mind about an issue because of something they saw on social media*. en-US. URL: <http://www.pewresearch.org/fact-tank/2018/08/01/14-percent-of-americans-have-changed-their-mind-about-an-issue-because-of-something-they-saw-on-social-media/>

- [//www.pewresearch.org/fact-tank/2018/08/15/14-of-americans-have-changed-their-mind-about-an-issue-because-of-something-they-saw-on-social-media/](http://www.pewresearch.org/fact-tank/2018/08/15/14-of-americans-have-changed-their-mind-about-an-issue-because-of-something-they-saw-on-social-media/) (visited on 08/16/2018).
- Bielaczyc, Katerine and Allan Collins (1999). "Learning Communities in Classrooms: A Reconceptualization of Educational Practice". en. In: p. 16.
- Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch (1992). "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades". In: *Journal of Political Economy* 100.5, pp. 992–1026. ISSN: 0022-3808. URL: <http://www.jstor.org/stable/2138632> (visited on 04/24/2018).
- Binns, Amy (Aug. 2012). "Don't Feed the Trolls!" In: *Journalism Practice* 6.4. Publisher: Routledge \_eprint: <https://doi.org/10.1080/17512786.2011.648988>, pp. 547–562. ISSN: 1751-2786. DOI: [10.1080/17512786.2011.648988](https://doi.org/10.1080/17512786.2011.648988). URL: <https://doi.org/10.1080/17512786.2011.648988> (visited on 10/14/2020).
- Blonder, Benjamin et al. (2012). "Temporal dynamics and network analysis". en. In: *Methods in Ecology and Evolution* 3.6. \_eprint: <https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/j.2041-210X.2012.00236.x>, pp. 958–972. ISSN: 2041-210X. DOI: [10.1111/j.2041-210X.2012.00236.x](https://doi.org/10.1111/j.2041-210X.2012.00236.x). URL: <https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/j.2041-210X.2012.00236.x> (visited on 10/14/2020).
- Bobbio, Norberto (2005). *Liberalism and Democracy*. en. Google-Books-ID: koyUXKVzte0C. Verso. ISBN: 978-1-84467-062-8.
- Bogers, Marcel and Joel West (2012). "Managing Distributed Innovation: Strategic Utilization of Open and User Innovation". en. In: *Creativity and Innovation Management* 21.1, pp. 61–75. ISSN: 1467-8691. DOI: [10.1111/j.1467-8691.2011.00622.x](https://doi.org/10.1111/j.1467-8691.2011.00622.x). URL: <http://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-8691.2011.00622.x> (visited on 06/04/2018).
- Bolici, Francesco, James Howison, and Kevin Crowston (2009). "Coordination without discussion? Socio-technical congruence and Stigmergy in Free and Open Source Software projects". In: *Socio-Technical Congruence Workshop in conj Intl Conf on Software Engineering, Vancouver, Canada*. (Visited on 01/16/2017).

- Borgatti, Stephen P and Martin G Everett (Oct. 2000). "Models of core/periphery structures". In: *Social Networks* 21.4, pp. 375–395. ISSN: 0378-8733. DOI: [10.1016/S0378-8733\(99\)00019-2](https://doi.org/10.1016/S0378-8733(99)00019-2). URL: <http://www.sciencedirect.com/science/article/pii/S0378873399000192> (visited on 02/19/2019).
- Bossetta, Michael (June 2018). "The Digital Architectures of Social Media: Comparing Political Campaigning on Facebook, Twitter, Instagram, and Snapchat in the 2016 U.S. Election". en. In: *Journalism & Mass Communication Quarterly* 95.2, pp. 471–496. ISSN: 1077-6990, 2161-430X. DOI: [10.1177/1077699018763307](https://doi.org/10.1177/1077699018763307). URL: <http://journals.sagepub.com/doi/10.1177/1077699018763307> (visited on 03/01/2019).
- Brabham, Daren C. (Feb. 2008). "Crowdsourcing as a Model for Problem Solving: An Introduction and Cases". en. In: *Convergence* 14.1, pp. 75–90. ISSN: 1354-8565. DOI: [10.1177/1354856507084420](https://doi.org/10.1177/1354856507084420). URL: <https://doi.org/10.1177/1354856507084420> (visited on 01/25/2018).
- Brabham, Daren C, Thomas W Sanchez, and Keith Bartholomew (2009). "Crowdsourcing Public Participation in Transit Planning: Preliminary Results from the Next Stop Design Case". en. In: p. 20.
- Broido, Anna D. and Aaron Clauset (Jan. 2018). "Scale-free networks are rare". In: *arXiv:1801.03400 [physics, q-bio, stat]*. arXiv: 1801.03400. URL: <http://arxiv.org/abs/1801.03400> (visited on 02/26/2018).
- Brown, John Seely and Paul Duguid (Feb. 1991). "Organizational Learning and Communities-of-Practice: Toward a Unified View of Working, Learning, and Innovation". In: *Organization Science* 2.1, pp. 40–57. ISSN: 1047-7039. DOI: [10.1287/orsc.2.1.40](https://doi.org/10.1287/orsc.2.1.40). URL: <https://pubsonline.informs.org/doi/abs/10.1287/orsc.2.1.40> (visited on 03/08/2019).
- Bu, Tian and Don Towsley (2002). *On distinguishing between Internet power law topology generators*. URL: [http://www.academia.edu/1452980/On\\_distinguishing\\_between\\_Internet\\_power\\_law\\_topology\\_generators](http://www.academia.edu/1452980/On_distinguishing_between_Internet_power_law_topology_generators) (visited on 09/17/2015).
- Buecheler, Thierry et al. (2010). "Crowdsourcing, Open Innovation and Collective Intelligence in the Scientific Method-A Research Agenda and Operational Framework." In: *ALIFE*, pp. 679–686.

- Butler, Brian et al. (Jan. 2007). "Community Effort in Online Groups: Who Does the Work and Why?" en. In: Publisher: Carnegie Mellon University. DOI: [10.1184/R1/6469925.v1](https://doi.org/10.1184/R1/6469925.v1). URL: [https://kilthub.cmu.edu/articles/Community\\_Effort\\_in\\_Online\\_Groups\\_Who\\_Does\\_the\\_Work\\_and\\_Why\\_/6469925](https://kilthub.cmu.edu/articles/Community_Effort_in_Online_Groups_Who_Does_the_Work_and_Why_/6469925) (visited on 04/27/2020).
- Chadwick, Andrew (Nov. 2003). "Bringing E-Democracy Back In: Why it Matters for Future Research on E-Governance". en. In: *Social Science Computer Review* 21.4, pp. 443–455. ISSN: 0894-4393. DOI: [10.1177/0894439303256372](https://doi.org/10.1177/0894439303256372). URL: <https://doi.org/10.1177/0894439303256372> (visited on 03/01/2019).
- Chambers, Simone (June 2009). "Rhetoric and the Public Sphere , Rhetoric and the Public Sphere: Has Deliberative Democracy Abandoned Mass Democracy? , Has Deliberative Democracy Abandoned Mass Democracy?" en. In: *Political Theory* 37.3, pp. 323–350. ISSN: 0090-5917. DOI: [10.1177/0090591709332336](https://doi.org/10.1177/0090591709332336). URL: <https://doi.org/10.1177/0090591709332336> (visited on 07/31/2018).
- Charband, Yeganeh and Nima Jafari Navimipour (Dec. 2016). "Online knowledge sharing mechanisms: a systematic review of the state of the art literature and recommendations for future research". en. In: *Information Systems Frontiers* 18.6, pp. 1131–1151. ISSN: 1387-3326, 1572-9419. DOI: [10.1007/s10796-016-9628-z](https://doi.org/10.1007/s10796-016-9628-z). URL: <http://link.springer.com/10.1007/s10796-016-9628-z> (visited on 07/04/2017).
- Chesbrough, Henry, Wim Vanhaverbeke, and Joel West (Aug. 2006). *Open Innovation: Researching a New Paradigm*. en. Google-Books-ID: RdcSDAAAQBAJ. OUP Oxford. ISBN: 978-0-19-929072-7.
- Chiu, Chao-Min et al. (Feb. 2007). "Understanding knowledge sharing in virtual communities: An integration of expectancy disconfirmation and justice theories". In: *Online Information Review* 35.1, pp. 134–153. ISSN: 1468-4527. DOI: [10.1108/14684521111113623](https://doi.org/10.1108/14684521111113623). URL: <http://www.emeraldinsight.com/doi/abs/10.1108/14684521111113623>.
- Christiano, Thomas (July 1996). *The Rule of the Many: Fundamental Issues in Democratic Theory*. English. 1 edition. Boulder, Colo: Westview Press. ISBN: 978-0-8133-1455-6.

- (May 2008). *The Constitution of Equality: Democratic Authority and Its Limits*. Anglais. Oxford ; New York: OUP Oxford. ISBN: 978-0-19-829747-5.
- Cohen, Joshua (1986). "An Epistemic Conception of Democracy". In: *Ethics* 97.1. Publisher: University of Chicago Press, pp. 26–38. ISSN: 0014-1704. URL: <https://www.jstor.org/stable/2381404> (visited on 05/27/2020).
- (1989). "Deliberation and Democratic Legitimacy". In: *The Good Polity: Normative Analysis of the State*. New York: Blackwell.
- (Oct. 2009). *Philosophy, Politics, Democracy – Selected Essays*. Anglais. 1st ed. Cambridge, Mass: Harvard University Press. ISBN: 978-0-674-03448-8.
- Coleman, Jules and John Ferejohn (Oct. 1986). "Democracy and Social Choice". In: *Ethics* 97.1. Publisher: The University of Chicago Press, pp. 6–25. ISSN: 0014-1704. DOI: [10.1086/292814](https://doi.org/10.1086/292814). URL: <https://www.journals.uchicago.edu/doi/abs/10.1086/292814> (visited on 05/27/2020).
- Conaldi, Guido and Alessandro Lomi (May 2013). "The dual network structure of organizational problem solving: A case study on open source software development". en. In: *Social Networks* 35.2, pp. 237–250. ISSN: 03788733. DOI: [10.1016/j.socnet.2012.12.003](https://doi.org/10.1016/j.socnet.2012.12.003). URL: <http://linkinghub.elsevier.com/retrieve/pii/S037887331200069X> (visited on 01/16/2017).
- Condorcet, Jean-Antoine-Nicolas de Caritat (1743-1794 ; marquis de) Auteur du texte (1785). *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix ([Reprod.]) / par M. le marquis de Condorcet,...* FR. URL: <https://gallica.bnf.fr/ark:/12148/bpt6k417181> (visited on 05/03/2020).
- Connelly, Brian L. et al. (Dec. 2010). "Signaling Theory: A Review and Assessment." en. In: *Journal of Management*. Publisher: SAGE PublicationsSage CA: Los Angeles, CA. DOI: [10.1177/0149206310388419](https://doi.org/10.1177/0149206310388419). URL: <https://journals.sagepub.com/doi/10.1177/0149206310388419> (visited on 04/18/2020).
- Constant, David, Sara Kiesler, and Lee Sproull (1994). "What's Mine Is Ours, or Is It? A Study of Attitudes about Information Sharing". In: *Information Systems Research* 5.4, pp. 400–421. ISSN: 1047-7047. URL: <http://www.jstor.org/stable/23010604> (visited on 04/03/2018).

- Cooper, Colin and Alan Frieze (May 2003). "A General Model of Web Graphs". In: *Random Struct. Algorithms* 22.3, pp. 311–335. ISSN: 1042-9832. DOI: [10.1002/rsa.10084](https://doi.org/10.1002/rsa.10084). URL: <http://dx.doi.org/10.1002/rsa.10084> (visited on 06/23/2016).
- Corbet, Jonathan and Greg Kroah-Hartman (2017). "Linux Kernel Development Report". en. In: p. 31.
- Cox, Alan (Oct. 1998). *Cathedrals, Bazaars and the Town Council*. URL: <https://news.slashdot.org/story/98/10/13/1423253/featurecathedrals-bazaars-and-the-town-council>.
- Cunningham, PáDraig et al. (Aug. 2013). "Characterizing ego-networks using motifs". en. In: *Network Science* 1.02, pp. 170–190. ISSN: 2050-1242, 2050-1250. DOI: [10.1017/nws.2013.12](https://doi.org/10.1017/nws.2013.12). URL: [http://www.journals.cambridge.org/abstract\\_S205012421300012X](http://www.journals.cambridge.org/abstract_S205012421300012X) (visited on 06/09/2016).
- Dabbish, Laura et al. (2012). "Social coding in GitHub: transparency and collaboration in an open software repository". In: *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*. ACM, pp. 1277–1286. URL: <http://dl.acm.org/citation.cfm?id=2145396> (visited on 01/16/2017).
- Dabbish, Laura et al. (Jan. 2013). "Leveraging Transparency". In: *IEEE Software* 30.1. Conference Name: IEEE Software, pp. 37–43. ISSN: 1937-4194. DOI: [10.1109/MS.2012.172](https://doi.org/10.1109/MS.2012.172).
- Dahl, Robert A. (Nov. 1989). *Democracy and Its Critics*. Anglais. New Haven: Yale University Press. ISBN: 978-0-300-04409-6.
- Dahl, Robert A. and Ian Shapiro (Oct. 2008). *On Democracy*. Anglais. Second Edition, 2. Yale University Press.
- Dahlberg, Lincoln (Mar. 2001). "Extending the public sphere through cyberspace: The case of Minnesota E-Democracy". en-US. In: *First Monday* 6.3. ISSN: 13960466. DOI: [10.5210/fm.v6i3.838](https://doi.org/10.5210/fm.v6i3.838). URL: <https://firstmonday.org/ojs/index.php/fm/article/view/838> (visited on 03/01/2019).

- Dai, Peng, Mausam, and Daniel Sabby Weld (Aug. 2011). "Artificial Intelligence for Artificial Intelligence". en. In: *Twenty-Fifth AAAI Conference on Artificial Intelligence*. URL: <https://www.aaai.org/ocs/index.php/AAAI/AAAI11/paper/view/3775> (visited on 04/24/2018).
- Dalle, Jean-Michel and Paul A. David (2003). "The allocation of software development resources in 'open source' production mode". In: *SIEPR-Project NOSTRA Working Paper, (15th February)[Accepted for publication in Joe Feller, Brian Fitzgerald, Scott Hissam, Karim Lakhani, eds., Making Sense of the Bazaar, forthcoming from MIT Press in 2004]*. URL: <http://econwpa.repec.org/eps/io/papers/0502/0502011.pdf> (visited on 10/26/2015).
- Dalle, Jean-Michel, Martin Devillers, and Matthijs den Besten (2014). "Answer Editing in StackOverflow". In: *Proceedings of the 2nd International Conference on Knowledge Commons. NYU Engelberg*. URL: <http://iasckc.nyuengelberg.org/s/Jean-Michel-Dalle-Martin-Devillers-Matthijs-den-Besten.pdf> (visited on 09/10/2015).
- De Borda, Jean-Charles (1781). "Mémoire sur les élections au scrutin". In: *Histoire de l'Académie royale des sciences*. URL: <http://asklepios.chez.com/XIX/borda.htm> (visited on 05/27/2020).
- Del Vicario, Michela et al. (July 2017). "Mapping social dynamics on Facebook: The Brexit debate". In: *Social Networks* 50, pp. 6–16. ISSN: 0378-8733. DOI: 10.1016/j.socnet.2017.02.002. URL: <http://www.sciencedirect.com/science/article/pii/S0378873316304166> (visited on 03/01/2019).
- Deterding, Sebastian et al. (May 2011). "Gamification. using game-design elements in non-gaming contexts". In: *CHI '11 Extended Abstracts on Human Factors in Computing Systems. CHI EA '11. Vancouver, BC, Canada: Association for Computing Machinery*, pp. 2425–2428. ISBN: 978-1-4503-0268-5. DOI: 10.1145/1979742.1979575. URL: <https://doi.org/10.1145/1979742.1979575> (visited on 04/17/2020).
- Dewey, John (1927). "The public and its problems An Essay in Political Inquiry". en. In: p. 241.
- Dorat, Rémi et al. (2007). "Multi-level analysis of an interaction network between individuals in a mailing-list". In: *Annales des télécommunications*.

- Vol. 62. Springer, pp. 325–349. URL: <http://link.springer.com/article/10.1007/BF03253264> (visited on 06/24/2016).
- Dryzek, John S. and Christian List (Jan. 2003). "Social Choice Theory and Deliberative Democracy: A Reconciliation". en. In: *British Journal of Political Science* 33.1. Publisher: Cambridge University Press, pp. 1–28. ISSN: 1469-2112, 0007-1234. DOI: [10.1017/S0007123403000012](https://doi.org/10.1017/S0007123403000012). URL: <https://www.cambridge.org/core/journals/british-journal-of-political-science/article/social-choice-theory-and-deliberative-democracy-a-reconciliation/B3AC55710D81F60C8F3817A3BD8B3099> (visited on 05/03/2020).
- Duggan, Maeve et al. (2015). *Mobile Messagin and social 2015*. en.
- Erdős, Paul and Alfréd Rényi (1960). "On the Evolution of Random Graphs". In: *PUBLICATION OF THE MATHEMATICAL INSTITUTE OF THE HUNGARIAN ACADEMY OF SCIENCES*, pp. 17–61.
- Estlund, David (1998). "The Insularity of the Reasonable: Why Political Liberalism Must Admit the Truth". In: *Ethics* 108.2. Publisher: The University of Chicago Press, pp. 252–275. ISSN: 0014-1704. DOI: [10.1086/233804](https://doi.org/10.1086/233804). URL: <https://www.jstor.org/stable/10.1086/233804> (visited on 05/27/2020).
- (Aug. 2009). *Democratic Authority: A Philosophical Framework*. en. Google-Books-ID: YIHGF166Kd4C. Princeton University Press. ISBN: 978-1-4008-3154-8.
- Estlund, David M. (2003). "Why Not Epistocracy?" In: *Desire, Identity and Existence: Essays in honnor of T.M. Penner*. Ed. by Naomi Reshotko. Kelowna, British Columbia: Academic Printing and Publishin. URL: <https://www.brown.edu/academics/philosophy/sites/brown.edu/academics/philosophy/files/uploads/WhyNotEpistocracy.pdf> (visited on 05/03/2020).
- Faraj, Samer, Sirkka L. Jarvenpaa, and Ann Majchrzak (Oct. 2011). "Knowledge Collaboration in Online Communities". en. In: *Organization Science* 22.5, pp. 1224–1239. ISSN: 1047-7039, 1526-5455. DOI: [10.1287/orsc.1100.0614](https://doi.org/10.1287/orsc.1100.0614). URL: <http://pubsonline.informs.org/doi/abs/10.1287/orsc.1100.0614> (visited on 06/20/2017).

- Farrell, S. (Sept. 2011). "Social influence benefits the wisdom of individuals in the crowd". en. In: *Proceedings of the National Academy of Sciences* 108.36, E625–E625. ISSN: 0027-8424, 1091-6490. DOI: [10.1073/pnas.1109947108](https://doi.org/10.1073/pnas.1109947108). URL: <http://www.pnas.org/cgi/doi/10.1073/pnas.1109947108> (visited on 06/25/2018).
- Fishkin, James S. (1995). *The Voice of the People: Public Opinion and Democracy*. en. Yale University Press. ISBN: 978-0-300-06556-5.
- Forlizzi, Jodi and Katja Battarbee (2004). "Understanding Experience in Interactive Systems". In: *Proceedings of the 5th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques*. DIS '04. event-place: Cambridge, MA, USA. New York, NY, USA: ACM, pp. 261–268. ISBN: 978-1-58113-787-3. DOI: [10.1145/1013115.1013152](https://doi.org/10.1145/1013115.1013152). URL: <http://doi.acm.org/10.1145/1013115.1013152> (visited on 03/08/2019).
- Galbraith, Jay R. (Mar. 2014). *Designing Organizations: Strategy, Structure, and Process at the Business Unit and Enterprise Levels*. Anglais. 3rd Edition. San Francisco: John Wiley & Sons. ISBN: 978-1-118-40995-4.
- Galton, Francis (Mar. 1907a). "The Ballot-Box". en. In: *Nature* 75.1952, pp. 509–510. ISSN: 1476-4687. DOI: [10.1038/075509f0](https://doi.org/10.1038/075509f0). URL: <https://www-nature-com.inshs.bib.cnrs.fr/articles/075509f0> (visited on 02/21/2018).
- (Mar. 1907b). "Vox Populi". en. In: *Nature* 75.1949, pp. 450–451. ISSN: 0028-0836, 1476-4687. DOI: [10.1038/075450a0](https://doi.org/10.1038/075450a0). URL: <http://www.nature.com/articles/075450a0> (visited on 04/24/2018).
- Garnier, Simon, Jacques Gautrais, and Guy Theraulaz (Oct. 2007). "The biological principles of swarm intelligence". en. In: *Swarm Intelligence* 1.1, pp. 3–31. ISSN: 1935-3812, 1935-3820. DOI: [10.1007/s11721-007-0004-y](https://doi.org/10.1007/s11721-007-0004-y). URL: <http://link.springer.com/10.1007/s11721-007-0004-y> (visited on 04/23/2018).
- Gherardi, Silvia (Jan. 1999). "Learning as Problem-Driven or Learning in the Face of Mystery?" en. In: *Organization Studies* 20.1, pp. 101–123. ISSN: 0170-8406, 1741-3044. DOI: [10.1177/0170840699201005](https://doi.org/10.1177/0170840699201005). URL: <http://journals.sagepub.com/doi/10.1177/0170840699201005> (visited on 06/19/2020).

- Gherardi, Silvia and Davide Nicolini (Mar. 2000). "The Organizational Learning of Safety in Communities of Practice". en. In: *Journal of Management Inquiry* 9.1, pp. 7–18. ISSN: 1056-4926, 1552-6542. DOI: [10.1177/105649260091002](https://doi.org/10.1177/105649260091002). URL: <http://journals.sagepub.com/doi/10.1177/105649260091002> (visited on 06/19/2020).
- Ginsburg, M. and S. Weisband (Jan. 2004). "A framework for virtual community business success: the case of the Internet Chess Club". In: *37th Annual Hawaii International Conference on System Sciences, 2004. Proceedings of the*, 10 pp.–. DOI: [10.1109/HICSS.2004.1265466](https://doi.org/10.1109/HICSS.2004.1265466).
- Gordon, Deborah M. (Mar. 2010). *Ant Encounters: Interaction Networks and Colony Behavior*. en. Google-Books-ID: MabwdXLZ9YMC. Princeton University Press. ISBN: 978-1-4008-3544-7.
- Grant, Robert M. (May 2014). "Toward a knowledge-based theory of the firm". In: *Strategic Management Journal* 17.S2, pp. 109–122. ISSN: 0143-2095. DOI: [10.1002/smj.4250171110](https://doi.org/10.1002/smj.4250171110). URL: <http://onlinelibrary.wiley.com/doi/full/10.1002/smj.4250171110> (visited on 04/25/2018).
- Grek, Sotiria (Sept. 2013). "Expert moves: international comparative testing and the rise of expertocracy". In: *Journal of Education Policy* 28.5. Publisher: Routledge \_eprint: <https://doi.org/10.1080/02680939.2012.758825>, pp. 695–709. ISSN: 0268-0939. DOI: [10.1080/02680939.2012.758825](https://doi.org/10.1080/02680939.2012.758825). URL: <https://doi.org/10.1080/02680939.2012.758825> (visited on 06/17/2020).
- Grofman, Bernard and Scott L. Feld (Dec. 2004). "If you like the alternative vote (a.k.a. the instant runoff), then you ought to know about the Coombs rule". en. In: *Electoral Studies* 23.4, pp. 641–659. ISSN: 0261-3794. DOI: [10.1016/j.electstud.2003.08.001](https://doi.org/10.1016/j.electstud.2003.08.001). URL: <http://www.sciencedirect.com/science/article/pii/S026137940300060X> (visited on 05/27/2020).
- Grofman, Bernard, Guillermo Owen, and Scott L. Feld (Sept. 1983). "Thirteen theorems in search of the truth". en. In: *Theory and Decision* 15.3, pp. 261–278. ISSN: 1573-7187. DOI: [10.1007/BF00125672](https://doi.org/10.1007/BF00125672). URL: <https://doi.org/10.1007/BF00125672> (visited on 05/03/2020).
- Grönlund, Åke (2009). "ICT Is Not Participation Is Not Democracy – eParticipation Development Models Revisited". en. In: *Electronic Participation*. Ed.

- by Ann Macintosh and Efthimios Tambouris. Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 12–23. ISBN: 978-3-642-03781-8.
- Grönlund, Åke and Joachim Åström (2009). “DoIT Right: Measuring Effectiveness of Different eConsultation Designs”. en. In: *Electronic Participation*. Ed. by Ann Macintosh and Efthimios Tambouris. Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 90–100. ISBN: 978-3-642-03781-8.
- Guerineau, Mathias (2018). “Des modèles de diffusion et de transfert de l’innovation à celui de déploiement: une conceptualisation nouvelle de la phase aval des processus d’innovation des firmes multinationale”. fr. In: p. 443.
- Gómez, Vicenç, Hilbert J. Kappen, and Andreas Kaltenbrunner (2011). “Modeling the structure and evolution of discussion cascades”. In: *arXiv:1011.0673 [physics]*. arXiv: 1011.0673, p. 181. DOI: [10.1145/1995966.1995992](https://doi.org/10.1145/1995966.1995992). URL: <http://arxiv.org/abs/1011.0673> (visited on 09/10/2015).
- Habermas, Jürgen (1984). *The theory of communicative action*. engger. Boston: Beacon Press. ISBN: 978-0-8070-1506-3 978-0-8070-1400-4.
- (1996). *Between facts and norms: contributions to a discourse theory of law and democracy*. en. Studies in contemporary German social thought. Cambridge, Mass: MIT Press. ISBN: 978-0-262-08243-3.
- (2002). *On the Pragmatics of Social Interaction: Preliminary Studies in the Theory of Communicative Action*. en. Google-Books-ID: CQtDResLBQ8C. MIT Press. ISBN: 978-0-262-58213-1.
- (Oct. 2015). *The Theory of Communicative Action: Lifeworld and Systems, a Critique of Functionalist Reason, Volume 2*. en. Google-Books-ID: kGSzC-gAAQBAJ. John Wiley & Sons. ISBN: 978-0-7456-9446-7.
- Habermas, Jürgen and Thomas Burger (1998). *The structural transformation of the public sphere: an inquiry into a category of bourgeois society*. en. 9. print. Studies in contemporary German social thought. OCLC: 248073326. Cambridge, Mass: MIT Press. ISBN: 978-0-262-08180-1 978-0-262-58108-0.
- Hagel, John (1999). “Net gain: Expanding markets through virtual communities”. en. In: *Journal of Interactive Marketing* 13.1, pp. 55–65. ISSN: 1520-6653. DOI: [10.1002/\(SICI\)1520-6653\(199924\)13:1<55::AID-DIR5>3.0.CO;2-](https://doi.org/10.1002/(SICI)1520-6653(199924)13:1<55::AID-DIR5>3.0.CO;2-)

- C. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/%28SICI%291520-6653%28199924%2913%3A1%3C55%3A%3AAID-DIR5%3E3.0.CO%3B2-C> (visited on 08/15/2019).
- Hall, Hazel and Dianne Graham (2004). "Creation and recreation: motivating collaboration to generate knowledge capital in online communities". en. In: *International Journal of Information Management*, p. 12.
- Halpern, Joseph Y. and Yoram Moses (July 1990). "Knowledge and common knowledge in a distributed environment". In: *Journal of the ACM* 37.3, pp. 549–587. ISSN: 0004-5411. DOI: [10.1145/79147.79161](https://doi.org/10.1145/79147.79161). URL: <https://doi.org/10.1145/79147.79161> (visited on 05/03/2020).
- Hayek, F. A. (1952). *The Sensory Order: An Inquiry Into the Foundations of Theoretical Psychology*. en. Google-Books-ID: UFazm1Xy\_j4C. University of Chicago Press. ISBN: 978-0-226-32094-6.
- Hayek, F. A. von (1937). "Economics and Knowledge". In: *Economica* 4.13. Publisher: [London School of Economics, Wiley, London School of Economics and Political Science, Suntory and Toyota International Centres for Economics and Related Disciplines], pp. 33–54. ISSN: 0013-0427. DOI: [10.2307/2548786](https://www.jstor.org/stable/2548786). URL: <https://www.jstor.org/stable/2548786> (visited on 06/19/2020).
- Hecht-Nielsen, R. (Mar. 1988). "Neurocomputing: picking the human brain". In: *IEEE Spectrum* 25.3, pp. 36–41. ISSN: 0018-9235. DOI: [10.1109/6.4520](https://doi.org/10.1109/6.4520).
- Holland, Paul W. and Samuel Leinhardt (1970). "A Method for Detecting Structure in Sociometric Data". In: *American Journal of Sociology* 76.3, pp. 492–513. ISSN: 0002-9602. URL: <http://www.jstor.org/stable/2775735> (visited on 09/15/2016).
- Holme, Petter (Sept. 2015). "Modern temporal network theory: a colloquium". en. In: *The European Physical Journal B* 88.9, p. 234. ISSN: 1434-6036. DOI: [10.1140/epjb/e2015-60657-4](https://doi.org/10.1140/epjb/e2015-60657-4). URL: <https://doi.org/10.1140/epjb/e2015-60657-4> (visited on 10/14/2020).
- Hong, Lu and Scott E. Page (Nov. 2004). "Groups of diverse problem solvers can outperform groups of high-ability problem solvers". en. In: *Proceedings of the National Academy of Sciences of the United States of America* 101.46,

- pp. 16385–16389. ISSN: 0027-8424, 1091-6490. DOI: [10.1073/pnas.0403723101](https://doi.org/10.1073/pnas.0403723101).  
URL: <http://www.pnas.org/content/101/46/16385> (visited on 01/08/2018).
- Howe, Jeff (June 2006). *Crowdsourcing: A Definition*. URL: [http://www.crowdsourcing.com/cs/2006/06/crowdsourcing\\_a.html](http://www.crowdsourcing.com/cs/2006/06/crowdsourcing_a.html) (visited on 04/24/2018).
- Howison, James and Kevin Crowston (2014). "Collaboration Through Open Superposition: A Theory of the Open Source Way." In: *Mis Quarterly* 38.1, pp. 29–50. URL: <http://aisel.aisnet.org/cgi/viewcontent.cgi?article=3156&context=misq> (visited on 01/16/2017).
- Ilgen, Daniel R. (1999). "Teams embedded in organizations: Some implications". In: *American Psychologist* 54.2. Place: US Publisher: American Psychological Association, pp. 129–139. ISSN: 1935-990X(Electronic),0003-066X(Print). DOI: [10.1037/0003-066X.54.2.129](https://doi.org/10.1037/0003-066X.54.2.129).
- Iriberry, Alicia and Gonyea Leroy (Feb. 2009). "A life-cycle perspective on on-line community success". en. In: *ACM Computing Surveys* 41.2, pp. 1–29. ISSN: 03600300. DOI: [10.1145/1459352.1459356](https://doi.org/10.1145/1459352.1459356). URL: <http://portal.acm.org/citation.cfm?doid=1459352.1459356> (visited on 08/15/2019).
- Janis, Irving L. (1972). *Victims of groupthink: A psychological study of foreign-policy decisions and fiascoes*. Victims of groupthink: A psychological study of foreign-policy decisions and fiascoes. Oxford, England: Houghton Mifflin.
- (1991). "GroupThink". English. In: *A first look at communication theory*. Ed. by Emory A Griffin. OCLC: 21973427. New York: McGraw-Hill. ISBN: 978-0-07-022778-1.
- Janis, Irving Lester (June 1982). *Groupthink: Psychological Studies of Policy Decisions and Fiascoes*. Anglais. Boston: Houghton Mifflin. ISBN: 978-0-395-31704-4.
- Janis, Irving Lester and Leon Mann (1977). *Decision making: a psychological analysis of conflict, choice, and commitment*. en. Google-Books-ID: YgRHAAAA-MAAJ. Free Press. ISBN: 978-0-02-916160-9.
- Jarvenpaa, Sirkka L., Kathleen Knoll, and Dorothy E. Leidner (1998). "Is Anybody Out There? Antecedents of Trust in Global Virtual Teams". In: *J. of*

- Management Information Systems* 14, pp. 29–64. DOI: [10.1080/07421222.1998.11518185](https://doi.org/10.1080/07421222.1998.11518185).
- Johnson, Christopher M (Jan. 2001). “A survey of current research on online communities of practice”. In: *The Internet and Higher Education* 4.1, pp. 45–60. ISSN: 1096-7516. DOI: [10.1016/S1096-7516\(01\)00047-1](https://doi.org/10.1016/S1096-7516(01)00047-1). URL: <http://www.sciencedirect.com/science/article/pii/S1096751601000471>.
- Kaltenbrunner, Andreas et al. (Aug. 2007). “Homogeneous temporal activity patterns in a large online communication space”. In: *arXiv:0708.1579 [cs]*. arXiv: 0708.1579. URL: <http://arxiv.org/abs/0708.1579> (visited on 08/08/2016).
- Kelley, Harold H. and John W. Thibaut (1978). *Interpersonal relations: a theory of interdependence*. en. Google-Books-ID: NyVHAAAAMAAJ. Wiley. ISBN: 978-0-471-03473-5.
- Kephart, Jeffrey O. and Steve R. White (Sept. 1992). “Directed-graph epidemiological models of computer viruses”. In: *Computation: The Micro and the Macro View*. WORLD SCIENTIFIC, pp. 71–102. ISBN: 978-981-02-0981-0. DOI: [10.1142/9789812812438\\_0004](https://doi.org/10.1142/9789812812438_0004). URL: [https://www.worldscientific.com/doi/abs/10.1142/9789812812438\\_0004](https://www.worldscientific.com/doi/abs/10.1142/9789812812438_0004) (visited on 03/08/2019).
- Khatib, Firas et al. (Oct. 2011). “Crystal structure of a monomeric retroviral protease solved by protein folding game players”. en. In: *Nature Structural & Molecular Biology* 18.10, pp. 1175–1177. ISSN: 1545-9985. DOI: [10.1038/nsmb.2119](https://doi.org/10.1038/nsmb.2119). URL: <http://www.nature.com/articles/nsmb.2119> (visited on 04/30/2018).
- Khatua, A. and A. Khatua (Dec. 2016). “Leave or Remain? Deciphering Brexit Deliberations on Twitter”. In: *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)*, pp. 428–433. DOI: [10.1109/ICDMW.2016.0067](https://doi.org/10.1109/ICDMW.2016.0067).
- Kittur, Aniket et al. (2007). “He says, she says: conflict and coordination in Wikipedia”. en. In: ACM Press, p. 453. ISBN: 978-1-59593-593-9. DOI: [10.1145/1240624.1240698](https://doi.org/10.1145/1240624.1240698). URL: <http://portal.acm.org/citation.cfm?doid=1240624.1240698> (visited on 01/10/2017).

- Klarreich, Erica (Feb. 2018). "The Controversial Theory That Explains the Structure of the Internet". In: *The Atlantic*. ISSN: 1072-7825. URL: <https://www.theatlantic.com/science/archive/2018/02/power-laws-networks/553562/> (visited on 02/26/2018).
- Knight, Jack and James Johnson (1994). "Aggregation and Deliberation: On the Possibility of Democratic Legitimacy". In: *Political Theory* 22.2. Publisher: Sage Publications, Inc., pp. 277–296. ISSN: 0090-5917. URL: <https://www.jstor.org/stable/192147> (visited on 05/27/2020).
- Krause, Jens, Graeme D. Ruxton, and Stefan Krause (Jan. 2010). "Swarm intelligence in animals and humans". en. In: *Trends in Ecology & Evolution* 25.1, pp. 28–34. ISSN: 0169-5347. DOI: 10.1016/j.tree.2009.06.016. URL: <http://www.sciencedirect.com/science/article/pii/S0169534709002298> (visited on 04/29/2020).
- Kraut, Robert E. et al. (2011). *Building Successful Online Communities: Evidence-Based Social Design*. MIT Press. ISBN: 978-0-262-01657-5. URL: <http://www.jstor.org/stable/j.ctt5hhgvw> (visited on 08/09/2018).
- Kriplean, Travis et al. (2011). "ConsiderIt: improving structured public deliberation". en. In: ACM Press, p. 1831. ISBN: 978-1-4503-0268-5. DOI: 10.1145/1979742.1979869. URL: <http://portal.acm.org/citation.cfm?doid=1979742.1979869> (visited on 03/29/2018).
- Kuhn, Thomas S. (Apr. 2012). *The Structure of Scientific Revolutions: 50th Anniversary Edition*. en. Google-Books-ID: 3eP5Y\_OOuzwC. University of Chicago Press. ISBN: 978-0-226-45814-4.
- Kumar, Ravi and Mohammad Mahdian (2010). "Dynamics of conversations". In: pp. 553–562. DOI: 10.1145/1835804.1835875.
- Kurvers, Ralf H. J. M. et al. (2015). "Self-organized flexible leadership promotes collective intelligence in human groups". In: *Royal Society Open Science* 2.12. Publisher: Royal Society, p. 150222. DOI: 10.1098/rsos.150222. URL: <https://royalsocietypublishing.org/doi/full/10.1098/rsos.150222> (visited on 04/29/2020).
- Ladha, Krishna K. (1992). "The Condorcet Jury Theorem, Free Speech, and Correlated Votes". In: *American Journal of Political Science* 36.3. Publisher:

- [Midwest Political Science Association, Wiley], pp. 617–634. ISSN: 0092-5853. DOI: [10.2307/2111584](https://doi.org/10.2307/2111584). URL: <https://www.jstor.org/stable/2111584> (visited on 05/03/2020).
- Landemore, Helene (Feb. 2017). *Democratic Reason: Politics, Collective Intelligence, and the Rule of the Many*. Anglais. Princeton, NJ Oxford: Princeton University Press. ISBN: 978-0-691-17639-0.
- Lave, Jean and Etienne Wenger (Sept. 1991). *Situated Learning: Legitimate Peripheral Participation*. en. Google-Books-ID: CAVIOrW3vYAC. Cambridge University Press. ISBN: 978-0-521-42374-8.
- Lazer, David et al. (Feb. 2009). “Computational Social Science”. en. In: *Science* 323.5915, pp. 721–723. ISSN: 0036-8075, 1095-9203. DOI: [10.1126/science.1167742](https://doi.org/10.1126/science.1167742). URL: <http://science.sciencemag.org/content/323/5915/721> (visited on 08/10/2018).
- Le Bon, Gustave (1895). *Psychologie des foules*. French. OCLC: 858217431. Paris. ISBN: 978-2-13-062062-4.
- Lee, Fion S L, Douglas Vogel, and Moez Limayem (2003). “Virtual Community Informatics: A Review and Research Agenda”. en. In: p. 15.
- Leimeister, Jan Marco (Aug. 2010). “Collective Intelligence”. en. In: *Business & Information Systems Engineering* 2.4, pp. 245–248. ISSN: 1867-0202. DOI: [10.1007/s12599-010-0114-8](https://doi.org/10.1007/s12599-010-0114-8). URL: <http://link.springer.com/10.1007/s12599-010-0114-8> (visited on 02/20/2018).
- Lenhart, Amanda, John B. Horrigan, and Deborah Fallows (Feb. 2004). *Content Creation Online*. en-US. URL: <http://www.pewinternet.org/2004/02/29/content-creation-online/> (visited on 08/16/2018).
- Liang, Hai (Oct. 2014). “The Organizational Principles of Online Political Discussion: A Relational Event Stream Model for Analysis of Web Forum Deliberation”. en. In: *Human Communication Research* 40.4, pp. 483–507. ISSN: 1468-2958. DOI: [10.1111/hcre.12034](https://doi.org/10.1111/hcre.12034). URL: <http://onlinelibrary.wiley.com/doi/10.1111/hcre.12034/abstract> (visited on 03/03/2017).
- Lindley, Dennis V. (1985). “Reconciliation of Discrete Probability Distributions.” In: *Bayesian Statistics*. Amsterdam: North-Holland. Vol. 2.

- List, Christian and Robert E. Goodin (Sept. 2001). "Epistemic Democracy: Generalizing the Condorcet Jury Theorem". en. In: *Journal of Political Philosophy* 9.3, pp. 277–306. ISSN: 09638016. DOI: [10.1111/1467-9760.00128](https://doi.org/10.1111/1467-9760.00128). URL: <http://doi.wiley.com/10.1111/1467-9760.00128> (visited on 05/03/2020).
- List, Christian and Philip Pettit (2004). "An epistemic free-riding problem?" eng. In: *Karl Popper: Critical Appraisals*. Ed. by Philip Catton and Graham Macdonald. Num Pages: 240. Abingdon, UK: Routledge, pp. 128–158. ISBN: 978-0-415-31971-3. URL: <http://eprints.lse.ac.uk/19624/> (visited on 05/03/2020).
- Lomi, Alessandro et al. (June 2014). "Participation motifs and the emergence of organization in open productions". en. In: *Structural Change and Economic Dynamics* 29, pp. 40–57. ISSN: 0954349X. DOI: [10.1016/j.strueco.2013.01.002](https://doi.org/10.1016/j.strueco.2013.01.002). URL: <http://linkinghub.elsevier.com/retrieve/pii/S0954349X13000039> (visited on 01/16/2017).
- Lorenz, Jan et al. (May 2011). *How social influence can undermine the wisdom of crowd effect*. Vol. 108. DOI: [10.1073/pnas.1008636108](https://doi.org/10.1073/pnas.1008636108).
- Lumbreras, Alberto (Nov. 2016). "Automatic role detection in online forums". PhD thesis. URL: <http://www.theses.fr/s148461> (visited on 01/12/2017).
- MacGilvray, Eric (2014). "Democratic Doubts: Pragmatism and the Epistemic Defense of Democracy". en. In: *Journal of Political Philosophy* 22.1. \_eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/jopp.12017>, pp. 105–123. ISSN: 1467-9760. DOI: [10.1111/jopp.12017](https://doi.org/10.1111/jopp.12017). URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/jopp.12017> (visited on 05/27/2020).
- Macintosh, Ann (2008). "E-Democracy and E-Participation Research in Europe". en. In: *Digital Government: E-Government Research, Case Studies, and Implementation*. Ed. by Hsinchun Chen et al. Integrated Series In Information Systems. Boston, MA: Springer US, pp. 85–102. ISBN: 978-0-387-71611-4. DOI: [10.1007/978-0-387-71611-4\\_5](https://doi.org/10.1007/978-0-387-71611-4_5). URL: [https://doi.org/10.1007/978-0-387-71611-4\\_5](https://doi.org/10.1007/978-0-387-71611-4_5) (visited on 03/01/2019).
- Madirolas, Gabriel and Gonzalo De Polavieja (June 2014). "Wisdom of the Confident: Using Social Interactions to Eliminate the Bias in Wisdom of

- the Crowds". In: *arXiv:1406.7578 [physics]*. arXiv: 1406.7578. URL: <http://arxiv.org/abs/1406.7578> (visited on 06/25/2018).
- Malhotra, Arvind, Sanjay Gosain, and Alexander Hars (1997). "Evolution of a Virtual Community: Understanding Design Issues Through a Longitudinal Study". en. In: p. 16.
- Malone, Thomas W. and Michael S. Bernstein (Dec. 2015). *Handbook of Collective Intelligence*. Anglais. 1st ed. Cambridge, Massachusetts: MIT Press. ISBN: 978-0-262-02981-0.
- Malone, Thomas W., Robert Laubacher, and Chrysanthos Dellarocas (2009). "Harnessing crowds: Mapping the genome of collective intelligence". In: Mano, Mattias, Jean-Michel Dalle, and Joanna Tomasik (2017). "An Author Network to Classify Open Online Discussions". en. In: ACM Press, pp. 1–4. ISBN: 978-1-4503-5187-4. DOI: [10.1145/3125433.3125450](https://doi.org/10.1145/3125433.3125450). URL: <http://dl.acm.org/citation.cfm?doid=3125433.3125450> (visited on 02/14/2018).
- (2018). "The Consent of the Crowd Detected in an Open Forum". en. In: *Proceedings of the 14th International Symposium on Open Collaboration - OpenSym '18*. Paris, France: ACM Press, pp. 1–9. ISBN: 978-1-4503-5936-8. DOI: [10.1145/3233391.3233538](https://doi.org/10.1145/3233391.3233538). URL: <http://dl.acm.org/citation.cfm?doid=3233391.3233538> (visited on 09/21/2018).
- Mathieu, John E. and Dennis M. Zajac (1990). "A review and meta-analysis of the antecedents, correlates, and consequences of organizational commitment". In: *Psychological Bulletin* 108.2, pp. 171–194. ISSN: 1939-1455(Electronic),0033-2909(Print). DOI: [10.1037/0033-2909.108.2.171](https://doi.org/10.1037/0033-2909.108.2.171).
- McFadden, D. L. (1973). "Conditional Logit Analysis of Qualitative Choice Behavior". In: *Frontiers in Econometrics*. New York: Wiley.
- Medvedev, Alexey N, Renaud Lambiotte, and Jean-Charles Delvenne (2018). "The anatomy of Reddit: An overview of academic research". en. In: p. 19.
- Mendelberg, Tali (2002). "The deliberative citizen: Theory and evidence". In: *In*. Jawe Press.
- Merchant, Arpit et al. (2019). "Signals Matter: Understanding Popularity and Impact of Users on Stack Overflow". en. In: *The World Wide Web Conference*

- on - WWW '19. San Francisco, CA, USA: ACM Press, pp. 3086–3092. ISBN: 978-1-4503-6674-8. DOI: [10.1145/3308558.3313583](https://doi.org/10.1145/3308558.3313583). URL: <http://dl.acm.org/citation.cfm?doid=3308558.3313583> (visited on 04/17/2020).
- Milgram, Stanley, Leonard Bickman, and Lawrence Berkowitz (1969). “Note on the drawing power of crowds of different size.” en. In: *Journal of Personality and Social Psychology* 13.2, pp. 79–82. ISSN: 0022-3514. DOI: [10.1037/h0028070](https://doi.org/10.1037/h0028070). URL: <http://content.apa.org/journals/psp/13/2/79> (visited on 04/24/2018).
- Mill, John Stuart (1998). *On Liberty and Other Essays*. en. Google-Books-ID: m6cQfXCETnIC. Oxford University Press. ISBN: 978-0-19-283384-6.
- Millen, David R., Michael A. Fontaine, and Michael J. Muller (Apr. 2002). “Understanding the benefit and costs of communities of practice”. en. In: *Communications of the ACM* 45.4, p. 69. ISSN: 00010782. DOI: [10.1145/505248.505276](https://doi.org/10.1145/505248.505276). URL: <http://portal.acm.org/citation.cfm?doid=505248.505276> (visited on 08/15/2019).
- Miller, Nicholas R. (1986). “Information, Electorates, and Democracy: Some Extensions and Interpretations of the Condorcet Jury Theorem”. In: *Information Pooling and Group Decision Making*. JAI Press.
- Milo, Ron et al. (Oct. 2002). “Network Motifs: Simple Building Blocks of Complex Networks”. en. In: *Science* 298.5594, pp. 824–827. ISSN: 0036-8075, 1095-9203. DOI: [10.1126/science.298.5594.824](https://doi.org/10.1126/science.298.5594.824). URL: <http://www.sciencemag.org/content/298/5594/824> (visited on 09/17/2015).
- Moussaïd, Mehdi (Jan. 2019). *Fouloscopie. Ce que la foule dit de nous*. fr. Google-Books-ID: NdmEDwAAQBAJ. Humensis. ISBN: 978-2-37931-014-0.
- Musi, Elena (Apr. 2018). “How did you change my view? A corpus-based study of concessions’ argumentative role”. en. In: *Discourse Studies* 20.2, pp. 270–288. ISSN: 1461-4456, 1461-7080. DOI: [10.1177/1461445617734955](https://doi.org/10.1177/1461445617734955). URL: <http://journals.sagepub.com/doi/10.1177/1461445617734955> (visited on 10/14/2020).
- Musi, Elena, Debanjan Ghosh, and Smaranda Muresan (June 2018). “Change-MyView Through Concessions: Do Concessions Increase Persuasion?” en.

- In: *arXiv:1806.03223 [cs]*. arXiv: 1806.03223. URL: <http://arxiv.org/abs/1806.03223> (visited on 10/14/2020).
- Newman, Mark EJ (2005). "Power laws, Pareto distributions and Zipf's law". In: *Contemporary physics* 46.5, pp. 323–351. URL: <http://www.tandfonline.com/doi/abs/10.1080/00107510500052444> (visited on 06/03/2016).
- Ng, Andrew Y., Michael I. Jordan, and Yair Weiss (2001). "On Spectral Clustering: Analysis and an algorithm". In: *Advances in Neural Information Processing Systems*. MIT Press, pp. 849–856.
- Nitzan, Shmuel and Jacob Paroush (1982). "Optimal Decision Rules in Uncertain Dichotomous Choice Situations". In: *International Economic Review* 23.2. Publisher: [Economics Department of the University of Pennsylvania, Wiley, Institute of Social and Economic Research, Osaka University], pp. 289–297. ISSN: 0020-6598. DOI: [10.2307/2526438](https://www.jstor.org/stable/2526438). URL: <https://www.jstor.org/stable/2526438> (visited on 05/03/2020).
- Ober, Josiah (2009). "An Aristotelian middle way between deliberation and independent-guess aggregation". en. In: p. 11.
- (Apr. 2010). *Democracy and Knowledge – Innovation and Learning in Classical Athens*. Anglais. Princeton, N.J.: Princeton University Press. ISBN: 978-0-691-14624-9.
- (Feb. 2013). "Democracy's Wisdom: An Aristotelian Middle Way for Collective Judgment". en. In: *American Political Science Review* 107.1. Publisher: Cambridge University Press, pp. 104–122. ISSN: 0003-0554, 1537-5943. DOI: [10.1017/S0003055412000627](https://www.cambridge.org/core/journals/american-political-science-review/article/democracys-wisdom-an-aristotelian-middle-way-for-collective-judgment/BB1443001C1FE0B7F2B3129A61C58970). URL: <https://www.cambridge.org/core/journals/american-political-science-review/article/democracys-wisdom-an-aristotelian-middle-way-for-collective-judgment/BB1443001C1FE0B7F2B3129A61C58970> (visited on 05/03/2020).
- Oldenburg, Ray (Aug. 1999). *The Great Good Place: Cafes, Coffee Shops, Bookstores, Bars, Hair Salons, and Other Hangouts at the Heart of a Community*. Anglais. 3rd ed. New York : Berkeley, Calif.: Da Capo Press. ISBN: 978-1-56924-681-8.
- Oliver, Richard L. (1980). "A Cognitive Model of the Antecedents and Consequences of Satisfaction Decisions". In: *Journal of Marketing Research* 17.4,

- pp. 460–469. ISSN: 0022-2437. DOI: [10.2307/3150499](https://doi.org/10.2307/3150499). URL: <http://www.jstor.org/stable/3150499> (visited on 08/17/2018).
- O'Bryan, Lisa, Margaret Beier, and Eduardo Salas (Mar. 2020). "How Approaches to Animal Swarm Intelligence Can Improve the Study of Collective Intelligence in Human Teams". en. In: *Journal of Intelligence* 8.1, p. 9. ISSN: 2079-3200. DOI: [10.3390/jintelligence8010009](https://doi.org/10.3390/jintelligence8010009). URL: <https://www.mdpi.com/2079-3200/8/1/9> (visited on 04/24/2020).
- Padgett, John F. and Christopher K. Ansell (1993). "Robust Action and the Rise of the Medici, 1400-1434". In: *American journal of sociology*, pp. 1259–1319. URL: <http://www.jstor.org/stable/2781822> (visited on 09/20/2016).
- Page, Scott E. (2008). *The difference: How the power of diversity creates better groups, firms, schools, and societies*. Princeton University Press.
- Pardes, Arielle (Apr. 2019). "'Change My View' Reddit Community Launches Its Own Website". en. In: *Wired*. ISSN: 1059-1028. URL: <https://www.wired.com/story/change-my-view-gets-its-own-website/> (visited on 04/08/2020).
- Park, Juyong and Albert-László Barabási (2007). "Distribution of node characteristics in complex networks". In: *Proceedings of the National Academy of Sciences* 104.46, pp. 17916–17920. URL: <http://www.pnas.org/content/104/46/17916.short> (visited on 04/05/2016).
- Perkmann, Markus and Joel West (2014). "Open Science and Open Innovation". en. In: p. 35.
- Platon (Feb. 2002). *La République*. Français. Nouvelle. Paris: Flammarion. ISBN: 978-2-08-070653-9.
- Rand, William M. (1971). "Objective Criteria for the Evaluation of Clustering Methods". In: *Journal of the American Statistical Association* 66.336, pp. 846–850. ISSN: 0162-1459. DOI: [10.2307/2284239](https://doi.org/10.2307/2284239). URL: [www.jstor.org/stable/2284239](http://www.jstor.org/stable/2284239) (visited on 12/17/2019).
- Rawls, John (1997). "The Idea of Public Reason Revisited". en. In: *The University of Chicago Law Review*, p. 44.

- Raymond, Eric S. (Jan. 2001). *The Cathedral & the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary*. English. 1 edition. Beijing ; Cambridge, Mass: O'Reilly Media. ISBN: 978-0-596-00108-7.
- Rheingold, Howard (1993). *The Virtual Community: Finding Connection in a Computerized World*. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc. ISBN: 978-0-201-60870-0.
- Riker, William H. (1988). *Liberalism against populism: a confrontation between the theory of democracy and the theory of social choice*. eng. OCLC: ocm18869468. Prospect Heights, Ill: Waveland Press. ISBN: 978-0-88133-367-1.
- Rittenberg, Libby, Timothy Tregarthen, and Institute for the Study of Knowledge Management in Education (ISKME) (2009). *Principles of Microeconomics*. English. OCLC: 1066384768. Place of publication not identified: publisher not identified. URL: [http://VH7QX3XE2P.search.serialssolutions.com/?V=1.0&L=VH7QX3XE2P&S=AC\\_T\\_B&C=Principles%20of%20Microeconomics&T=marc&tab=BOOKS](http://VH7QX3XE2P.search.serialssolutions.com/?V=1.0&L=VH7QX3XE2P&S=AC_T_B&C=Principles%20of%20Microeconomics&T=marc&tab=BOOKS) (visited on 06/19/2020).
- Rokach, Lior and Oded Maimon (2005). "Clustering Methods". en. In: *Data Mining and Knowledge Discovery Handbook*. Ed. by Oded Maimon and Lior Rokach. Boston, MA: Springer US, pp. 321–352. ISBN: 978-0-387-25465-4. DOI: [10.1007/0-387-25465-X\\_15](https://doi.org/10.1007/0-387-25465-X_15). URL: [https://doi.org/10.1007/0-387-25465-X\\_15](https://doi.org/10.1007/0-387-25465-X_15) (visited on 12/17/2019).
- Rousseau, Jean-Jacques (1782). *Du contrat social: ou Principes du droit politique*. fr. Google-Books-ID: nEk8AAAACAAJ. A. Londres.
- Rousseeuw, Peter J. (Nov. 1987). "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis". en. In: *Journal of Computational and Applied Mathematics* 20, pp. 53–65. ISSN: 0377-0427. DOI: [10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7). URL: <http://www.sciencedirect.com/science/article/pii/0377042787901257> (visited on 12/17/2019).
- Sackin, M. J. (July 1972). "'Good' and 'Bad' Phenograms". en. In: *Systematic Biology* 21.2, pp. 225–226. ISSN: 1063-5157. DOI: [10.1093/sysbio/21.2.225](https://doi.org/10.1093/sysbio/21.2.225). URL: <https://academic.oup.com/sysbio/article/21/2/225/1713481> (visited on 09/26/2019).

- Salminen, Juho (Apr. 2012). "Collective Intelligence in Humans: A Literature Review". In: *arXiv:1204.3401 [cs]*. arXiv: 1204.3401. URL: <http://arxiv.org/abs/1204.3401> (visited on 04/29/2020).
- Sanders, Lynn M. (June 1997). "Against deliberation". English. In: *Political Theory* 25.3. Publisher: Sage Publications, Inc., pp. 347–377. ISSN: 00905917. URL: <https://go.gale.com/ps/i.do?p=AONE&sw=w&issn=00905917&v=2.1&it=r&id=GALE%7CA19589230&sid=googleScholar&linkaccess=abs> (visited on 05/27/2020).
- Schwartzberg, Melissa (2015). "Epistemic Democracy and Its Challenges". In: *Annual Review of Political Science* 18.1. \_eprint: <https://doi.org/10.1146/annurev-polisci-110113-121908>, pp. 187–203. DOI: [10.1146/annurev-polisci-110113-121908](https://doi.org/10.1146/annurev-polisci-110113-121908). URL: <https://doi.org/10.1146/annurev-polisci-110113-121908> (visited on 05/03/2020).
- Sen, Amartya (Jan. 1986). "Chapter 22 Social choice theory". en. In: *Handbook of Mathematical Economics*. Vol. 3. Elsevier, pp. 1073–1181. DOI: [10.1016/S1573-4382\(86\)03004-7](https://doi.org/10.1016/S1573-4382(86)03004-7). URL: <http://www.sciencedirect.com/science/article/pii/S1573438286030047> (visited on 05/03/2020).
- Shaw, Marvin E. (Jan. 1964). "Communication Networks". In: *Advances in Experimental Social Psychology*. Ed. by Leonard Berkowitz. Vol. 1. Academic Press, pp. 111–147. DOI: [10.1016/S0065-2601\(08\)60050-7](https://doi.org/10.1016/S0065-2601(08)60050-7). URL: <http://www.sciencedirect.com/science/article/pii/S0065260108600507> (visited on 04/25/2018).
- Shimazaki, Hiroto et al. (Nov. 2004). "Network analysis of potential migration routes for Oriental White Storks (*Ciconia boyciana*)". en. In: *Ecological Research* 19.6, pp. 683–698. ISSN: 0912-3814, 1440-1703. DOI: [10.1111/j.1440-1703.2004.00684.x](https://doi.org/10.1111/j.1440-1703.2004.00684.x). URL: <http://link.springer.com/10.1111/j.1440-1703.2004.00684.x> (visited on 03/08/2019).
- Shirazi, Farid (2008). "The Contribution of ICT to Freedom and Democracy: An Empirical Analysis of Archival Data on the Middle East". en. In: *The Electronic Journal of Information Systems in Developing Countries* 35.1, pp. 1–24. ISSN: 1681-4835. DOI: [10.1002/j.1681-4835.2008.tb00243.x](https://doi.org/10.1002/j.1681-4835.2008.tb00243.x). URL:

- <https://onlinelibrary.wiley.com/doi/abs/10.1002/j.1681-4835.2008.tb00243.x> (visited on 03/01/2019).
- Shirazi, Farid, Ojelanki Ngwenyama, and Olga Morawczynski (Feb. 2010). "ICT expansion and the digital divide in democratic freedoms: An analysis of the impact of ICT expansion, education and ICT filtering on democracy". In: *Telematics and Informatics* 27.1, pp. 21–31. ISSN: 0736-5853. DOI: [10.1016/j.tele.2009.05.001](https://doi.org/10.1016/j.tele.2009.05.001). URL: <http://www.sciencedirect.com/science/article/pii/S0736585309000203> (visited on 03/01/2019).
- Smith, Adam et al. (1859). *Recherches sur la nature et les causes de la richesse des nations*. fr. Google-Books-ID: NfEeAQAAMAAJ. Guillaumin et Cie.
- Smith, Peter J. (2008). "E-Democracy and Local Government - Dashed Expectations". en. In: *Electronic Government: Concepts, Methodologies, Tools, and Applications*, pp. 1708–1717. DOI: [10.4018/978-1-59904-947-2.ch129](https://doi.org/10.4018/978-1-59904-947-2.ch129). URL: <https://www.igi-global.com/chapter/democracy-local-government-dashed-expectations/9816> (visited on 03/01/2019).
- Sowe, Sulayman K., Ioannis Stamelos, and Lefteris Angelis (Mar. 2008). "Understanding knowledge sharing activities in free/open source software projects: An empirical study". In: *Journal of Systems and Software*. Selected Papers from the 2006 Brazilian Symposia on Databases and on Software Engineering 81.3, pp. 431–446. ISSN: 0164-1212. DOI: [10.1016/j.jss.2007.03.086](https://doi.org/10.1016/j.jss.2007.03.086). URL: <http://www.sciencedirect.com/science/article/pii/S0164121207000842>.
- Srba, Ivan and Maria Bielikova (2015). "Askalot: Community Question Answering as a Means for Knowledge Sharing in an Educational Organization". en. In: ACM Press, pp. 179–182. ISBN: 978-1-4503-2946-0. DOI: [10.1145/2685553.2699001](https://doi.org/10.1145/2685553.2699001). URL: <http://dl.acm.org/citation.cfm?doid=2685553.2699001> (visited on 07/04/2017).
- Sunstein, Cass R (2000). "Deliberative Trouble? Why Groups Go to Extremes". en. In: *The Yale Law Journal* 110, p. 49.
- Sunstein, Cass R. (Dec. 2002a). *Designing Democracy: What Constitutions Do*. Anglais. New Ed. Oxford: OUP USA. ISBN: 978-0-19-515840-3.

- (2002b). “The Law of Group Polarization”. en. In: *Journal of Political Philosophy* 10.2. \_eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/1467-9760.00148>, pp. 175–195. ISSN: 1467-9760. DOI: [10.1111/1467-9760.00148](https://doi.org/10.1111/1467-9760.00148). URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-9760.00148> (visited on 05/26/2020).
- Surowiecki, James (Aug. 2005). *The Wisdom of Crowds*. en. Knopf Doubleday Publishing Group. ISBN: 978-0-307-27505-9.
- Tan, Chenhao et al. (Feb. 2016). “Winning Arguments: Interaction Dynamics and Persuasion Strategies in Good-faith Online Discussions”. In: *arXiv:1602.01103 [physics]*. arXiv: 1602.01103. DOI: [10.1145/2872427.2883081](https://doi.org/10.1145/2872427.2883081). URL: <http://arxiv.org/abs/1602.01103> (visited on 04/25/2016).
- Tang, John et al. (Aug. 2009). “Temporal distance metrics for social network analysis”. In: *Proceedings of the 2nd ACM workshop on Online social networks*. WOSN '09. New York, NY, USA: Association for Computing Machinery, pp. 31–36. ISBN: 978-1-60558-445-4. DOI: [10.1145/1592665.1592674](https://doi.org/10.1145/1592665.1592674). URL: <https://doi.org/10.1145/1592665.1592674> (visited on 10/14/2020).
- Tarde, Gabriel (1901). *L'opinion et la foule*. fr.
- Tedjamulia, S.J.J. et al. (Jan. 2005). “Motivating Content Contributions to Online Communities: Toward a More Comprehensive Theory”. In: *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*. ISSN: 1530-1605, 193b–193b. DOI: [10.1109/HICSS.2005.444](https://doi.org/10.1109/HICSS.2005.444).
- Terwiesch, Christian and Karl Ulrich (June 2009). *Innovation Tournaments: Creating and Selecting Exceptional Opportunities*. Anglais. Boston, Mass: Harvard Business Review Press. ISBN: 978-1-4221-5222-5.
- Trockman, Asher et al. (May 2018). “Adding sparkle to social coding: an empirical study of repository badges in the *npm* ecosystem”. en. In: *Proceedings of the 40th International Conference on Software Engineering*. Gothenburg Sweden: ACM, pp. 511–522. ISBN: 978-1-4503-5638-1. DOI: [10.1145/3180155.3180209](https://doi.org/10.1145/3180155.3180209). URL: <https://dl.acm.org/doi/10.1145/3180155.3180209> (visited on 04/17/2020).

- Velkovska, Julia (Dec. 2015). "La « communauté de marque » comme accomplissement pratique". fr. In: *Communication. Information médias théories pratiques* vol. 33/2. ISSN: 1189-3788. DOI: [10.4000/communication.5906](https://doi.org/10.4000/communication.5906). URL: <https://communication.revues.org/5906> (visited on 07/29/2016).
- Von Hippel, Eric A. and Karim R. Lakhani (2000). "How Open Source Software Works: 'Free' User-to-User Assistance?" In: URL: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=290305](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=290305) (visited on 06/20/2017).
- Waldron, Jeremy (Apr. 1999). *Law and Disagreement*. Anglais. New Ed. Oxford: OUP Oxford. ISBN: 978-0-19-924303-7.
- Ward, Joe H. (1963). "Hierarchical Grouping to Optimize an Objective Function". In: *Journal of the American Statistical Association* 58.301, pp. 236–244. ISSN: 0162-1459. DOI: [10.2307/2282967](https://doi.org/10.2307/2282967). URL: [www.jstor.org/stable/2282967](http://www.jstor.org/stable/2282967) (visited on 12/17/2019).
- Wasserman, Stanley and Katherine Faust (Nov. 1994). *Social Network Analysis: Methods and Applications*. en. Google-Books-ID: CAm2DpIqRUIC. Cambridge University Press. ISBN: 978-0-521-38707-1.
- Watts, Duncan J. and Steven H. Strogatz (June 1998). "Collective dynamics of 'small-world' networks". en. In: *Nature* 393.6684, pp. 440–442. ISSN: 0028-0836. DOI: [10.1038/30918](https://doi.org/10.1038/30918). URL: <http://www.nature.com/nature/journal/v393/n6684/abs/393440a0.html> (visited on 09/10/2015).
- Wechsler, David (1939). *The measurement of adult intelligence*. en. Baltimore: Williams & Wilkins Co. DOI: [10.1037/10020-000](https://doi.org/10.1037/10020-000). URL: <http://content.apa.org/books/10020-000> (visited on 04/23/2018).
- Weick, Karl E. (1991). "The Nontraditional Quality of Organizational Learning". In: *Organization Science* 2.1. Publisher: INFORMS, pp. 116–124. ISSN: 1047-7039. URL: <https://www.jstor.org/stable/2634942> (visited on 06/19/2020).
- Wenger, Etienne (1998). *Communities of Practice: Learning, Meaning, and Identity*. en. Google-Books-ID: heBZpgYUKdAC. Cambridge University Press. ISBN: 978-0-521-66363-2.

- Wenger, Etienne, Richard A. McDermott, and William Snyder (2002). *Cultivating communities of practice: a guide to managing knowledge*. en. Boston, Mass: Harvard Business School Press. ISBN: 978-1-57851-330-7.
- Weninger, Tim, Xihao Avi Zhu, and Jiawei Han (2013). "An exploration of discussion threads in social news sites: A case study of the Reddit community". In: *Advances in Social Networks Analysis and Mining (ASONAM), 2013 IEEE/ACM International Conference on*. IEEE, pp. 579–583. URL: [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=6785761](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6785761) (visited on 03/09/2016).
- Weschler, David (1971). "Concept of collective intelligence". In: *American Psychologist* 26.10. Place: US Publisher: American Psychological Association, pp. 904–907. ISSN: 1935-990X(Electronic),0003-066X(Print). DOI: [10.1037/h0032223](https://doi.org/10.1037/h0032223).
- West, Joel and Jason Dedrick (June 2001). "Open source standardization: The rise of linux in the network era". en. In: *Knowledge, Technology & Policy* 14.2, pp. 88–112. ISSN: 0897-1986, 1874-6314. DOI: [10.1007/PL00022278](https://doi.org/10.1007/PL00022278). URL: <https://link.springer.com/article/10.1007/PL00022278> (visited on 06/04/2018).
- West, Joel and Scott Gallagher (2004). "Key Challenges of Open Innovation: Lessons from Open Source Software". en. In: p. 18.
- West, Joel and Karim R. Lakhani (Apr. 2008). "Getting Clear About Communities in Open Innovation". In: *Industry and Innovation* 15.2, pp. 223–231. ISSN: 1366-2716. DOI: [10.1080/13662710802033734](https://doi.org/10.1080/13662710802033734). URL: <https://doi.org/10.1080/13662710802033734> (visited on 06/04/2018).
- West, Joel and Siobhán O'mahony (Apr. 2008). "The Role of Participation Architecture in Growing Sponsored Open Source Communities". en. In: *Industry and Innovation* 15.2, pp. 145–168. ISSN: 1366-2716, 1469-8390. DOI: [10.1080/13662710801970142](https://doi.org/10.1080/13662710801970142). URL: <https://www.tandfonline.com/doi/full/10.1080/13662710801970142> (visited on 06/04/2018).
- West, Joel and Jonathan Sims (May 2018). *How Firms Leverage Crowds and Communities for Open Innovation*. en. Vol. 1. Oxford University Press. DOI: [10.1093/oso/9780198816225.003.0004](https://doi.org/10.1093/oso/9780198816225.003.0004). URL: <http://www.oxfordscholarship>.

- [com/view/10.1093/oso/9780198816225.001.0001/oso-9780198816225-chapter-4](https://doi.org/10.1093/oso/9780198816225.001.0001/oso-9780198816225-chapter-4) (visited on 03/27/2019).
- Winsor, Dorothy A. (Jan. 2001). "Learning to Do Knowledge Work in Systems of Distributed Cognition". en. In: *Journal of Business and Technical Communication* 15.1, pp. 5–28. ISSN: 1050-6519. DOI: [10.1177/105065190101500101](https://doi.org/10.1177/105065190101500101). URL: <https://doi.org/10.1177/105065190101500101> (visited on 02/15/2019).
- Woolley, Anita W. et al. (Oct. 2010). "Evidence for a Collective Intelligence Factor in the Performance of Human Groups". en. In: *Science* 330.6004, pp. 686–688. ISSN: 0036-8075, 1095-9203. DOI: [10.1126/science.1193147](https://doi.org/10.1126/science.1193147). URL: <http://www.sciencemag.org/cgi/doi/10.1126/science.1193147> (visited on 02/02/2018).
- Wright, Scott (Jan. 2012). "From "Third Place" to "Third Space": Everyday Political Talk in Non-Political Online Spaces". en. In: *Javnost - The Public* 19.3, pp. 5–20. ISSN: 1318-3222, 1854-8377. DOI: [10.1080/13183222.2012.11009088](https://doi.org/10.1080/13183222.2012.11009088). URL: <http://www.tandfonline.com/doi/abs/10.1080/13183222.2012.11009088> (visited on 08/13/2018).
- Yla R. Tausczik, Aniket Kittur (2014). "Collaborative problem solving: a study of MathOverflow". In: pp. 355–367. DOI: [10.1145/2531602.2531690](https://doi.org/10.1145/2531602.2531690).
- Zucchermaglio, Cristina and Alessandra Talamo (July 2003). "The Development of a Virtual Community of Practices Using Electronic Mail and Communicative Genres". en. In: *Journal of Business and Technical Communication* 17.3, pp. 259–284. ISSN: 1050-6519. DOI: [10.1177/1050651903017003001](https://doi.org/10.1177/1050651903017003001). URL: <https://doi.org/10.1177/1050651903017003001> (visited on 03/08/2019).

**Titre :** Comment la foule débat et consent dans un forum en ligne ouvert ? Étude de cas : Reddit – Change My View

**Mots clés :** Communauté en ligne, Démocratie, Forum, Intelligence collective, Réseaux, Sagesse des foules.

**Résumé :** L'émergence d'Internet a révolutionné une grande partie de la vie des individus, en touchant tous les aspects, notamment sur la façon de communiquer. Depuis les premiers newsgroups des années 1980 aux réseaux sociaux, les individus échangent quotidiennement des informations en ligne. Par ailleurs, différents événements politiques de part le monde, comme le mouvement des Gilets Jaunes en 2018, ou le scandale dit de Cambridge Analytica la même année, ont mis en exergue la place de l'espace Internet comme moyen d'expression de la sphère politique de nos sociétés. En dehors de la sphère politique, nous observons d'autres impact d'Internet sur la collaboration entre individu. Wikipedia, encyclopédie en ligne, est le résultat d'actions individuelles. Elle rassemblait déjà à la fin de 2018 plus de 35 millions de contributeurs.

Tous ces phénomènes permettent d'étudier la notion d'intelligence collective. C'est l'objectif de ce travail doctoral, qui porte sur l'étude de l'émergence de l'intelligence collective au sein des fora ouvert en ligne. Le cas présenté dans cette thèse est celui du forum Reddit – Change My view, où un individu présente son opinion sur un sujet et demande à la communauté de lui apporter les arguments nécessaires à ce qu'il change d'avis.

Ce cas d'étude permet d'analyser d'abord le cycle de vie des communautés en ligne. Mais également les différentes façons de débattre sur ce forum. Et enfin, nous y étudions le processus de cette communauté à établir un consensus entre ces membres. Ce dernier point nous a permis de faire émerger un nouveau concept : le Consentement de la Foule.

**Title:** How does the crowd debate and consent in an open online context? A case study: Reddit – Change My View.

**Keywords:** Democracy, Collective intelligence, Network, Online community, Wisdom of crowd.

**Abstract:** The development of the Internet revolutionized multiple aspects of daily life, and particularly how one communicates with another. From the 1980s and the first newsgroup to today's online social networks, people commonly exchange message online. Different political events over the world, such as the movement of the « Gilets Jaunes » in France, 2018, or the Cambridge Analytica scandal, the same year, highlight the place of Internet such as a powerful expression vector on the political sphere. Besides, we observe other impacts on the collaboration among individuals, through Internet. Wikipedia, the most popular online encyclopedia, is mainly the result of individual production. It counted at the end of 2018 more

than 35 millions of registered users. All these phenomena highlight the notion of collective intelligence. The purpose of this doctoral work is to study the emergence of collective intelligence within open online fora. The case study Reddit – Change My View, is a forum where an individual exposes his opinion on a subject and asks to the community to bring him arguments to change his opinion.

This case study allows to analyze the life cycle of online communities. But also the different processes leading to a constructive debate. Third, we study the process of this community to establish a consensus among his members. This last point highlighted a new concept: the Consent of the Crowd.