



HAL
open science

Behavior, opinion formation and cultural identity in social networks

Shaden Shabayek

► **To cite this version:**

Shaden Shabayek. Behavior, opinion formation and cultural identity in social networks. Economics and Finance. Université Panthéon-Sorbonne - Paris I, 2021. English. NNT : 2021PA01E005 . tel-03383378

HAL Id: tel-03383378

<https://theses.hal.science/tel-03383378>

Submitted on 18 Oct 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.



UNIVERSITÉ PARIS I PANTHÉON SORBONNE
UFR 02

Paris-Jourdan Sciences Economiques - Paris School of Economics

THÈSE

Pour l'obtention du titre de Docteur en Economie

Présentée et soutenue publiquement

le 25 janvier 2021 par

Shaden Shabayek

**Behavior, Opinion Formation & Cultural Identity
in Social Networks**

Sous la direction de Monsieur : **Francis Bloch**, Professeur à l'Université Paris 1 Panthéon-Sorbonne et Chaire associée à l'Ecole d'Economie de Paris.

Membres du Jury :

Laclau	Marie	Chargée de recherche CNRS Professeur Associée, HEC Paris	Examinatrice
Mauleon	Ana	Directrice de recherche FNRS Professeur, Université Catholique de Louvain	Rapporteuse
Pin	Paolo	Professeur, Università di Siena	Rapporteur
Rusinowska	Agnieszka	Directrice de recherche CNRS Université Paris1 Panthéon Sorbonne et PSE	Présidente du Jury
Vega-Redondo	Fernando	Professeur, Università Bocconi	Examineur

To my father, Dr. Ahmed-Shereen Shabayek, who never saw this adventure.

Remerciements

What you seek is seeking you.

Rumi

First and foremost, I would like to express my sincere gratitude to my supervisor Francis Bloch for his support and patience. I thank him very much for his help during these four and a half years of intense work. I have learned a great deal from him, in terms of the rigor for composing such a dissertation. I have had the chance to learn from his in-depth knowledge of an uncountable number of strands of literatures and his impressive way of thinking of problems and solving them. For all these reasons, I thank him for having accompanied me and for having always left his door open during these years of work. I also thank Francis to have given me the opportunity to share my work with my jury members: Marie Laclau, Ana Mauleon, Paolo Pin, Agnieszka Rusinowska and Fernando Vega-Redondo. I am thankful for all their comments and suggestions.

I am also infinitely grateful for Jean-François Caulier, who was my master I thesis supervisor at University Paris 1. I have a very vivid memory of being keen to find an exciting research topic for my master I thesis back in 2013. After roaming around the library *MIR* at Jussieu, I decide to pick up a recent volume of the International Journal of Game Theory, that was resting quietly on the Newsstand and waiting for me to pick up. One article arrests my attention with many graphs: *Contractually Stable Networks* by Jean-François Caulier, Ana Mauleon and Vincent Vannetelbosch. I start going through it, then notice that Jean-François was actually a professor at University Paris 1. On the very same day I decided to contact him to ask whether he was willing to supervise my master I thesis. From that point onwards, he invested a lot of time to make me discover many seminal papers in Networks, one of which was *The formation of networks with transfers among players* by Matthew Jackson and Francis Bloch - who would in few months become my master II thesis supervisor and one year later my Phd supervisor. Jean-François clearly put me on a path that I would pursue for years

and has made me discover a field of research which is both elegant and very relevant for our society.

I would like to extend my gratitude to four fantastic women in Academia I have met along the way: Agnieszka, Ana, Marie and Margherita. I had the chance to meet Agnieszka Rusinowska during my master I when I was invited to present my very elementary master I thesis at the reading group she co-organizes with Jean-François. She spoke to me after my presentation, gave me precious advice and encouraged me. I am happy to have met such a brilliant and kind researcher. I would also like to thank Margherita Comola for having been present along my Phd journey. I was always impressed by her genuine questions during our Networks reading group at PSE, by how much she masters the empirical Networks literature and finally by her good vibes. I am very happy and thankful that Marie Laclau was also on board during my journey. She always made time for me and I have always appreciated her far reaching and constructive comments, I have learnt a lot from her. Ana Mauleon and Vincent Vannetelbosch kindly hosted me for a pre-doctoral research visit. They have helped not only learn and discover the network formation literature, but were always so attentive to my well-being. I am very thankful to them. I would also like to thank Catherine Germain for all her support and kindness.

Since the beginning of the Phd I had the chance to meet many curious souls at PSE: Elias, Matthew O., Guilio, Philippe. I would also like to thank Ania Le Fur and Véronique Guillotin for all the fantastic work they do. I am very blessed to have developed along the way very original friendships: Narimène and Matthieu, Roghaya and Michela, Cathie, Claire and Aurore. I would like to also thank Guillaume for all his support, help and advise. I am also grateful for having encountered very inspiring spirits: Alexia, Juni and Caroline. Many of our talks give me hope for a better future, where we would all find a place in this world where we feel comfortable, light, fulfilled, valued and mostly useful for the society.

Furthermore I am thankful for my step family, Ellio and Nathalie, who were always so sweet to me. I am very lucky to have a big Egyptian family and deeply indebted to every one of them: Nana, Eman, Manal, Semsem and Aya. You all bring so much joy in my life and make me feel safe. My mom Mai has always believed in me, supported my decisions and loved me unconditionally. Life is inconceivable without her daily presence, support and wisdom. I always look up to her and dream of having someday her strength. Since we were baby girls, Rola has always been my back bone. Our unconditional trust and love to

each other is a bliss. Ellian my lifetime partner has helped me until the very last line of this thesis, thank you for being you.

Thank you all for contributing to this work!

SUMMARY

Understanding human interactions and behavior in modern social contexts is a necessary condition, for the design of public policies to face the challenges induced by digital technology today and understand its functioning and impact on the society. The form that social influence takes and the way it operates rely on the technology or the medium used. Therefore, it is essential to include features of modern social networks in theoretical models and study the relevance of regulations of digital technology. This is precisely the goal pursued in this present dissertation. Chapter 1, joint with Francis Bloch, considers a targeting problem with peer effects where the identity of individuals is anonymized. Chapter 2 studies opinion formation in social networks, by taking into account the inequality of attention towards expressed opinions. Chapter 3, joint with Alexia Lochmann, models the behavior of individuals when cultural identity is at play and when individuals interact in two different contexts with their peers.

Résumé

La compréhension des interactions et des comportements humains dans les contextes sociaux modernes est une condition nécessaire à la conception de politiques publiques permettant de relever les défis induits par la technologie numérique aujourd'hui et de comprendre son fonctionnement et son impact sur la société. La forme que prend l'influence sociale et son mode de fonctionnement dépendent de la technologie ou du support utilisé. Il est donc essentiel d'inclure les caractéristiques des réseaux sociaux modernes dans les modèles théoriques et d'étudier la pertinence des réglementations de la technologie numérique. C'est précisément l'objectif poursuivi dans le cadre de la présente thèse. Le chapitre 1, en collaboration avec Francis Bloch, examine un problème de ciblage avec des effets de pair où l'identité des individus est anonymisée. Le chapitre 2 étudie la formation de l'opinion dans les réseaux sociaux, en prenant en compte l'inégalité d'attention envers les opinions exprimées. Le chapitre 3, en collaboration avec Alexia Lochmann, modélise le comportement des individus lorsque l'identité culturelle est en jeu et lorsque les individus interagissent dans deux contextes différents avec leurs pairs.

Contents

Acknowledgments	iii
Summary	v
Résumé	vii
Introduction (English)	1
Introduction (Français)	11
1 Targeting in Social Networks with Anonymized Information	21
1.1 Introduction	22
1.2 The model	28
1.2.1 Agents, utilities and network effects	28
1.2.2 The planner's first-best	29
1.2.3 Network architectures and information structure	31
1.2.4 A menu of contracts	32
1.2.5 Efficient incentive compatible contracts	35
1.2.6 Group incentive compatible contracts	35
1.3 Undirected Networks	37
1.4 Hierarchical Networks	37
1.4.1 Single root	38
1.4.2 Nested neighborhoods	39
1.4.3 Regular Oriented Trees	42
1.4.4 Connected agents at the same tier of the hierarchy	43
1.5 Robustness and Extensions	44

1.5.1	Nonadjacent agents	44
1.5.2	Richer information structures and mechanisms	46
1.5.3	Partial information	47
1.5.4	Heterogenous marginal returns to own efforts	48
1.6	Conclusion	49
1.7	Appendix	51
1.7.1	Proof of proposition 1.4	51
1.7.2	Proof of proposition 1.5	60
1.7.3	Proof of proposition 1.7	63
1.7.4	Proof of proposition 1.9	64
2	Hidden Opinions	67
2.1	Introduction	68
2.2	Related work in social psychology	74
2.2.1	Hidden Profiles	74
2.2.2	Dynamic social Impact theory	75
2.3	The model	76
2.3.1	Set-up	76
2.3.2	Expression heuristic	76
2.3.3	Micro-foundation	78
2.4	Opinion Dynamics	79
2.4.1	Long-run opinions of expressers	79
2.4.2	The process of interpersonal influence	84
2.4.3	Patterns of long-run opinions: consensus and bi-polarization	87
2.5	Network topology and opinion patterns	91
2.5.1	Example	92
2.5.2	Agregate statistics	93
2.6	Conclusion and the way forward	94
2.7	Appendix	100
2.7.1	Proof of Proposition 2.1	100
2.7.2	Proof of Proposition 2.2	101
2.7.3	Proof of Lemma 2.1	107
2.7.4	Proof of Theorem 2.1	109

2.7.5	Proof of Proposition 2.3	111
2.7.6	Proof of Lemma 2.2	113
2.7.7	Network Statistics	114
2.7.8	Eigenvector centrality	115
3	Strategic cultural migration with peer effects	117
3.1	Introduction	118
3.2	Historical context	124
3.2.1	Migration decision: The procedure	125
3.3	The model	126
3.3.1	Set-up	126
3.3.2	The game	127
3.3.3	Equilibrium outcomes	129
3.4	Empirical evidence	131
3.4.1	Data	132
3.4.2	Definition of proxy variables	133
3.4.3	Empirical Strategy	137
3.4.4	Results	137
3.5	Conclusion	140
3.6	Appendix	152
3.6.1	Proof of Lemma 3.1	152
3.6.2	Proof of proposition 3.1	152
3.6.3	Beyond the Italian Border	154
	Bibliography	157

INTRODUCTION (English)

There is truth in Marshall McLuhan's assertion that the one thing of which the fish is unaware is water, the very medium that forms its ambience and supports its existence. Similarly, communication (...) comprises the ambience of human existence. The activities we collectively call communication - having conversations, giving instructions, imparting knowledge, sharing significant ideas, seeking information, entertaining and being entertained - are so ordinary and mundane that it is difficult for them to arrest our attention.

Communication as Culture: essays on media and society, James Carey (1989)

Humans were always far better at inventing tools than using them wisely.

21 lessons for the 21st century, Yuval Harari (2018)

Social networks play a substantial role in shaping and conditioning our behavior, decisions and opinions. Technology adoption, consumption attitudes, job search, political opinion formation, occupational and geographic mobility are all examples of how influence and communication via human interactions affect our lives. Hunter-gatherers have used story-telling via word-of-mouth communication to sustain cooperative behavior and expand their community.¹ Few decades ago, hippies, hackers and scholars among many others have come together to develop the internet, aiming to connect people, promote freedom and suppress geographic barriers to communication.² Today, nearly one third of the world population uses social media.³

¹Smith et al. (2017) [92] show that *the presence of good storytellers is associated with increased cooperation.*

²See the first section of the book *Digital Culutre* by Dominique Cardon (2019) [34].

³See for example, Facebook Reports Third Quarter 2020 Results: <https://investor.fb.com/investor-news/press-release-details/2020/Facebook-Reports-Third-Quarter-2020-Results/default.aspx>.

Cardon (2019) [34] compares the impact that the digital technology has had so far on our society, to the invention of printing press in the fifteenth century: (...) *“the digital revolution is above all a break in the way our societies produce, share and use knowledge. Printing certainly had very immediate effects from its earliest beginnings, starting with the multiplication and speed of reproduction and distribution of texts. But it has also been the starting point for a much more subtle set of mutations in ways of thinking, of challenging authority, of storing and circulating information. (...) The changes are intellectual, religious, psychological as well as economic or political. That is why it is useful to say that digital technology is a culture.”*

The digital revolution has not only transformed our societies, but it has also brought with it new challenges. To cite a few, private entities like Facebook or Twitter store ever-growing personal and social data measured in petabytes. This personal and social data is used to improve consumer experience and services provided to third-parties in order to sustain their business models.⁴ In spite of existing regulation and laws for privacy protection, multiple breaches have been recorded (e.g. HIV status of users via the dating app Grindr, Cambridge Analytica, etc.) suggesting the possibility of unethical use of confidential personal and social data. Hence, privacy has become a common concern and the distinction between the public and private sphere has become blurry. Another challenge faced by the society is the inequality of voices in online social networks. Social media have become a new networked public sphere where influencers and politicians express themselves, interact among each other and supposedly with the rest of the population. In particular, opinion polarization is empirically well documented and online social networking platforms are often accused of weakening democratic institutions.

With the abundance of data about social ties and increasingly powerful computers, a myriad of scholars such as social psychologists, sociologists and computer scientists have taken up the study of social networks. Economists as well are acknowledging the growing relevance of social interactions for the decision making process of agents in many situations. In particular since a handful of seminal papers in the 1990s, namely Jackson and Wolinsky (1996) [62], Bala and Goyal (2000) [9] and Kranton and Minehart (2001) [71], theoretical models in economics have developed a rational appetite for including social ties as recurrent modeling ingredients.

That being so, understanding human interactions and behavior in modern social contexts

⁴For more on this topic, see *The Age of Surveillance Capitalism* (2019) [102] by Shoshana Zuboff.

is a necessary condition, for the design of public policies to face the challenges induced by digital technology today and understand its functioning and impact on the society. The form that social influence takes and the way it operates rely on the technology or the medium used. Therefore, it is essential to include features of *modern* social networks in theoretical models and study the relevance of regulations of digital technology. This is precisely the goal pursued in this present dissertation.

(i) *Contribution of the thesis and Outline*

Chapter 1, joint with Francis Bloch, considers a targeting problem with peer effects where the identity of individuals is anonymized. There is a third party, a planner or an external entity that would like to affect the behavior of a set of agents embedded in a network in order to attain a predefined goal. Examples of such situations include:

- A firm that attempts to exploit consumption externalities among agents in order to maximize profit.
- A local government that would like to affect the crime level within a network of criminals in order to reduce the overall level of crime in a given city.
- A local government that would like to affect the effort level of pupils in the same classroom in order to increase the overall level of achievement.

While targeting problems belong to a well established literature, the novelty of the present research lies in incorporating anonymized social data rather than complete social data. Such exercise is pertinent because anonymizing data is often presented as the solution to privacy breaches. In our model, the planner can observe the network topology but cannot learn the identity of each agent. Going back to the above three examples, this would correspond to a European firm that buys data about social ties of users of a given platform from data brokers but due to regulations (e.g. GDPR) it cannot observe the identity. In the case of the criminal network, knowing the network topology would come down to observing communication flows between individuals without being able to unravel the identity of the senders and the receivers.

The main contribution of Chapter 1 is to show that depending on the interaction pattern between agents, the knowledge of their identity by the planner may or may not matter. In

particular, when influence is reciprocal, meaning that any pair of directly connected agents influence each other mutually, the knowledge of the identity of agents does not matter and the planner can attain her goal. However, for networks where there is a discrepancy between the influence received and the influence exerted, we show that the planner cannot attain her goal and ends up offering uniform quantities in the case of a monopolist.

Chapter 2 studies opinion formation in social networks, by taking into account the inequality of attention towards expressed opinions. The access to internet in hundreds of countries⁵ and the availability of multiple social networking platforms have given to billions of people the *possibility* to have a voice, express an opinion, listen to different viewpoints and be part of a global tribe. Yet, only a happy few are heard, as Cardon (2019) [34] argues that visibility on the internet follows a power law where 1% of the available content attracts over 90% of users' attention.⁶ A handful of politicians, artists and scholars, act as opinion leaders who create content, interact with each other, influence public opinion and receive attention from masses of molecular users.

Again like problems of Targeting studied in Chapter 1, opinion formation is a topic extensively studied. Furthermore, it is embedded in several well established literatures in Sociology, Social Psychology, Computer Science, Physics and Economics. The novelty in Chapter 2 lies in modeling precisely the inequality of attention, observed on social media. To that end, I formalize a model where an individual chooses to express their opinion or hide it based on their popularity. Individuals who hide their opinion can be interpreted as individuals who have a low popularity such that even if they speak-up they will not be heard or considered. One can think of an individual who has very few followers on Twitter. Alternatively, a second interpretation can be that hiding one's opinion is less costly than expressing it. The cost of expression can be the time spent arguing with more eloquent and persuasive peers or the cost of social isolation when one's opinion drifts from the average group viewpoint. In addition, there are a number of empirical papers that support the idea that ideologically-opposed individuals interact together often (e.g. Conover et al. (2011) [39]). Henceforth, I include this feature in my model. The study of the interactions between locally popular individuals who interact with like-minded or ideologically-opposed peers can explain whether consensus or polarization prevails. Since influence is stronger locally, clusters can form. But some members within a given cluster who are popular enough and inter-

⁵According to [Wikipedia](#), 53.5% of the world population in 2019 were internet users.

⁶See the Chapter *The digital transformation of the public sphere*, on page 147.

act with ideologically-opposed peers can fall into disagreement, which causes opinions to polarize across clusters. I define a local popularity measure and use it for simulations to illustrate the findings of my model.

The main contribution of chapter 2 is to provide a unifying framework which explains the emergence of several opinion patterns, such as consensus, polarization and total disagreement. The key element is to study the interaction pattern between expressers, as it determines the outcome of opinion dynamics. In particular I show that, if long-run opinions are bi-polarized then there do not exist individuals who receive an equal amount of influence from two ideologically-opposed extreme opinion groups. Those individuals occupy a very particular location in the network, because they are not locally popular enough to express and they are equally influenced by two ideologically-opposed groups of expressers. Those individuals could be interpreted as neutral TV hosts, or non-biased journalists or intermediaries in general. Such individuals typically do not exist in modern communication networks such as Twitter where politicians can *mention* each other and interact directly. While in past decades, debates between politicians, open to the public were usually moderated by TV hosts.

Chapter 3, joint with Alexia Lochmann, models the behavior of individuals when cultural identity is at play and when individuals interact in two different contexts with their peers. We provide novel empirical evidence based on historical data. Akerlof and Kranton (2000) [4] in their seminal book *Identity Economics* have made clear that identity and norms are an important determinants for decision making. They argue that identity can be a strong source of motivation. Furthermore, individual behavior is affected and shaped by many simultaneous interactions, simply because individuals are embedded in several networks. For example, exchange relations or friendship ties may influence behavior in different ways. Wasserman and Faust (1994) [97] in the last section of their widely used standard textbook *Social Network Analysis* point out that *one area where there is clear need for continued work is developing methods to study multiple relations (page 730)*. In spite of the existence of some academic work on this question, there is no consensus about how to model multiple relations and how to think about centrality when multiple networks including the same actors are studied simultaneously. Hence, the vast majority of papers in Economics dealing with social and economic networks, tend to focus on one single type of relationships and model it with a simple graph.

In our work, we provide a simple model where each individual has a cultural type and is embedded in two networks. Our model is framed in terms of selection into migration. The novel evidence we provide is based on digitizing historical household level data from South Tyrol, the northernmost province of Italy, following the Option Agreement between Hitler and Mussolini in 1939 that gave South Tyroleans the option to leave their homeland and keep their culture (by that time they were all german speaking) or stay in Italy and abandon their culture. Although it seems specific to the migration literature, the modeling approach is pertinent for the study of cultural identity in multiplex social networks, people interact together using several media tools entailing correlation of behavior and influence across different platforms.

To study the cultural, social and economic components of self-selection into migration, we formalize a game with strategic households that report their intentions-to-emigrate. Each household has a cultural type, which is their degree of attachment to the germanic culture. Households interact with other households within their social circles, including friends, work colleagues and neighbors. The social circle of each household represents their social capital. Furthermore, to account for economic factors we consider property ownership; which is the most relevant economic component in an agricultural society. The properties of neighbors who decide to migrate are potential economic opportunities that a given household can seize. We show that when only considering interactions in the social network, if most of a given household's social contacts intend to emigrate, then the household is also more likely to emigrate otherwise they lose their social capital. Furthermore, independently of interactions in the social network, households have a lower incentive to emigrate, when all their neighbors in the geographic network have the intention to emigrate. This is because they can purchase the neighbor's house and enhance their economic well-being. Two individuals who are linked in both the social and the geographic network (e.g. a geographic neighbor that is also a friendly work colleague), affect positively each other's migration decisions, but the higher the value of the neighbor's property, the lower the effect. We study an exhaustive example to show how network positions play out in the net migration gains. A given individual who is central in the social network but peripheral in the geographic network, has negative migration gains whenever her social contacts have low cultural types entailing relatively low intentions-to-emigrate, even if the individual herself has relatively strong germanic culture.

(ii) *General Literature review*

The growing presence of networks since the 1990s in Economic models has nearly followed the evolution of digital technology and the growing interconnectedness of world. Including patterns of interaction has become of paramount importance for the study of several contemporary phenomena with a direct consequences on the economy such as: financial contagion, resilience of economic systems, international trade, supply chains, occupational mobility, human mobility, public opinions, social movements, climate change agreements, transnational cooperation, conflict, disease diffusion, technology adoption, etc. The literature can be broadly divided into two big strands. The first strand deals with games on Networks (see Bramoullé and Kranton (2016) [32] and Zenou (2016) [101]). It studies the behavior of interacting agents in a fixed network. It typically relates the position of agents described by a network centrality measure to the chosen action and the magnitude of externalities generated. The second strand deals with network formation and studies how those fixed networks to start with were formed (see Mauleon and Vannetelbosch (2016) [75]). This literature typically deals with notions of stability of the formed links and the efficiency of the networks formed from a societal perspective. Finally, there is a recent literature that mixes both strands where the actions of agents in a network and the choice of whom to interact with are studied simultaneously (see Vega-Redondo (2016) [96]).

The present dissertation is embedded in the first strand, as in the three chapters the network is fixed. Chapter 1 builds on a ramification of this literature, which studies targeting problems in fixed networks - such as price discrimination with price discounts to certain consumers in the presence of consumption externalities (see Bloch and Quérou (2013) [23] and Bimpikis et al. (2012) [33]) or unequal rewards for efforts in teams with synergies (see Galeotti et al. (2019) [50]) or injecting funds to in financial network to specific agents (Demange (2016) [43]). This literature makes the assumption that both the planner and the agents have full information regarding the interaction structure. While this assumption might be pertinent in small networks (e.g. banks), it can be naturally challenged for larger networks. The complete information assumption can be relaxed in several ways. In Chapter 1, we assume that the network topology is known but not the identity of agents. Other approaches include providing aggregate network statistics to the planner such as the degree distribution of the network (Fainmesser and Galeotti (2015) [45]) or the network generating process (Wilder et al. (2018) [60]).

Chapter 2 is embedded in a literature that studies information sharing and opinion formation. Economic models of learning in networks contain two big families. The first one is Bayesian learning, where rational agents can process the information they receive from their neighbors. This approach requires an extraordinary mental capacity due to the inherent complexity of networks (the need to consider the beliefs of neighbors about the beliefs of their neighbors and so on). The second family is Naive learning and it is often presented as a credible alternative to Bayesian learning (see Banerjee et al. (2018) [11]). Agents update their opinion by using simple heuristics. Prominent models of Naive learning are models where agents update their opinion at each period, by taking an average of the opinions of their friends at the previous period (see DeGroot (1974) [42], French (1956) [67] and Harary (1959) [57]). Both families of learning models provide results which answer the following questions: under what conditions do agents reach consensus of opinion in the long-run? If they do reach consensus, how long does it take to converge to those beliefs? Can agents aggregate the available information in the network? Does the consensus belief correspond to the correct belief, when agents' objective is to learn the true value of a parameter? The main theorem of this literature (*The Wisdom of the Crowds*) is that agents converge to the correct belief in the absence of a royal family or influencers (see Bala and Goyal (1998) [9], Acemoglu et al. (2008) [2], Golub and Jackson (2012) [53]). A growing body of research is enriching this literature to encompass many aspects that have received little attention, such as the strategic behavior (see Grabish et al. (2018) [55]). Chapter 2 refines these contributions by relating the opinion updating rule to the topology of the network. Namely, I introduce two types of individuals with different opinion updating rules. The type of individuals is determined by their network position (popularity). The first updating rule of the first type of agents is averaging á la DeGroot, while the second updating rule is a law of motion which incorporates an attractive effect and a repulsive effect. In doing so, depending on the network structure and initial opinions, long-run opinions can form different patterns, namely consensus or disagreement.

Finally, this economic literature addresses specific questions but also builds on research lead by social scientists over the past two centuries with the objective of understanding the formation of public opinion (and its manipulation). Early seminal contributions include *The Crowd: A Study of the Popular Mind* by Gustave Le Bon (1895) [26], *Propaganda* by Edward L. Bernays (1928) [17] and *Personal Influence* by Elihu Katz & Paul Lazarsfeld (1955) [69].

Chapter 3, besides introducing a multiplex network as mentioned earlier, belongs to the emerging literature which incorporates origin and destination networks in the study of migration. Borjas (1987) [31] introduced the *Roy Model* for the study of self-selection into migration. This model is a workhorse in the migration literature and it analyzes the economic drivers of migration: under which conditions does an individual decide to move? The main finding is that individuals decide to move when the earnings for their skill level are higher abroad than at home. Munshi (2020) [80] argues that this model cannot explain a number of stylized facts that include higher mobility of individuals towards destinations where others from their origin have moved before. Blumenstock, Chi, Tan (2019) [24] further explore the determinants of migration decision by focusing on the specific topology of the network. Their work confirms the hypothesis according to which people are more likely to migrate to destinations where they have more social contacts and are less likely to leave their original place of residence when they belong to larger social networks. By considering in chapter 3 household ties with respect to a geographic network and a social network, we extend the growing literature on migration networks. This literature typically considers one single type of interaction modeled by one network. Yet, influence between households' migration decisions can depend on the particular type of a tie they have among each other.

Finally, the three chapters aim to give the reader new perspectives. In particular, the inclusion of features of modern social networks can broaden the horizons of traditional models in network economics and provide innovative solutions to social and economic challenges.

INTRODUCTION (Français)

Il y a du vrai dans l'affirmation de Marshall McLuhan : la seule chose dont le poisson n'a pas conscience, c'est l'eau. Le milieu même qui forme son ambiance et soutient son existence. De même, la communication (...) correspond à l'atmosphère de l'existence humaine. Les activités que nous appelons collectivement communication - avoir des conversations, donner des instructions, transmettre des connaissances, partager des idées importantes, rechercher des informations, se divertir et être diverti - sont si ordinaires et banales qu'il est difficile pour elles d'attirer notre attention.

Communication comme culture: essais sur les *media* et la société,
James Carey (1989) - Traduction personnelle.

Les humains ont toujours excellé à inventer des outils, beaucoup moins à en faire un usage avisé.

21 leçons pour le XXI^e siècle, Yuval Harari (2018)

Les réseaux sociaux jouent un rôle important dans la formation et le conditionnement de nos comportements, décisions et opinions : l'adoption des technologies, les comportements consuméristes, la recherche d'emploi, la formation d'opinions politiques ou encore la mobilité professionnelle et géographique sont autant d'exemples de la manière dont la persuasion et la communication, via les interactions humaines, affectent notre vie. Les chasseurs-cueilleurs ont utilisé le récit d'histoires par le biais de la communication de bouche à oreille pour maintenir un comportement coopératif et élargir leur communauté.⁷ Il y a quelques décennies, les hippies, les hackers et les universitaires parmi tant d'autres se sont réunis pour développer l'internet, dans le but de connecter les individus, de promouvoir la liberté

⁷Smith et al. (2017) [92] montrent que *la présence de bons conteurs est associée à une coopération accrue.*

et de supprimer les barrières géographiques à la communication.⁸ Aujourd'hui, près d'un tiers de la population mondiale utilise les médias sociaux.⁹

Cardon (2019) [34] compare l'impact que la technologie numérique a eu jusqu'à présent sur notre société à l'invention de la presse à imprimer au XVe siècle : *"(...) la révolution numérique est avant tout une rupture dans la manière dont nos sociétés produisent, partagent et utilisent les connaissances. L'imprimerie a certes eu des effets très immédiats dès ses premiers balbutiements, à commencer par la multiplication et la rapidité de reproduction et de diffusion des textes. Mais elle a aussi été le point de départ d'un ensemble de mutations beaucoup plus subtiles dans les façons de penser, de contester l'autorité, de mettre l'information en mémoire ou en circulation. (...) Les changements sont intellectuels, religieux, psychologiques autant qu'économiques ou politiques. Voilà pourquoi il est utile de dire que le numérique est une culture."*

La révolution numérique a non seulement transformé nos sociétés, mais elle a également apporté avec elle de nouveaux défis. Pour n'en citer que quelques-uns, des entités privées comme Facebook ou Twitter stockent des données personnelles et sociales toujours plus nombreuses, mesurées en pétaoctets. Ces données personnelles et sociales sont utilisées pour améliorer l'expérience des consommateurs et les services fournis à des tiers afin de soutenir leurs modèles commerciaux.¹⁰ Malgré la réglementation et les lois existantes en matière de protection de la vie privée, de multiples violations ont été enregistrées (par exemple, la séropositivité des utilisateurs via l'application Grindr, Cambridge Analytica, etc.), ce qui laisse entrevoir la possibilité d'une utilisation non éthique des données personnelles et sociales confidentielles. La vie privée est donc devenue une préoccupation commune et la distinction entre la sphère publique et la sphère privée est devenue floue. L'inégalité des voix dans les réseaux sociaux en ligne est un autre défi auquel la société est confrontée. Les médias sociaux sont devenus une nouvelle sphère publique en réseau où les personnes d'influence et les hommes politiques s'expriment, interagissent entre eux et soi-disant avec le reste de la population. En particulier, la polarisation des opinions est empiriquement bien documentée et les plateformes de réseaux sociaux en ligne sont souvent accusées d'affaiblir les institutions démocratiques.

Avec l'abondance de données sur les liens sociaux et des ordinateurs de plus en plus puissants, une myriade de chercheurs tels que des psychologues sociaux, des sociologues et

⁸Voir la première section du livre *Digital Culutre* de Dominique Cardon (2019) [34].

⁹Facebook Reports Third Quarter 2020 Results: <https://investor.fb.com/investor-news/press-release-details/2020/Facebook-Reports-Third-Quarter-2020-Results/default.aspx>.

¹⁰Pour en savoir plus sur ce sujet, voir *L'ère du capitalisme de surveillance* (2019) [102] de Shoshana Zuboff.

des informaticiens se sont lancés dans l'étude des réseaux sociaux. Les économistes aussi reconnaissent la pertinence croissante des interactions sociales pour le processus de prise de décision des agents dans de nombreuses situations. En particulier, depuis une poignée d'articles fondateurs dans les années 1990, à savoir Jackson et Wolinsky (1996) [62], Bala et Goyal (2000) [9] et Kranton et Minehart (2001) [71], les modèles théoriques en économie ont développé un appétit rationnel pour l'inclusion des liens sociaux comme ingrédients récurrents de modélisation.

Cela étant, la compréhension des interactions et des comportements humains dans les contextes sociaux modernes est une condition nécessaire à la conception de politiques publiques permettant de relever les défis induits par la technologie numérique aujourd'hui et de comprendre son fonctionnement et son impact sur la société. La forme que prend l'influence sociale et son mode de fonctionnement dépendent de la technologie ou du support utilisé. Il est donc essentiel d'inclure les caractéristiques des réseaux sociaux *modern* dans les modèles théoriques et d'étudier la pertinence des réglementations de la technologie numérique. C'est précisément l'objectif poursuivi dans le cadre de la présente thèse.

(i) *Contribution de la thèse et axes de recherches*

Le chapitre 1, en collaboration avec Francis Bloch, examine un problème de ciblage avec effets de pair où l'identité des individus est anonymisée. Un tiers, un planificateur ou une entité externe souhaite influencer le comportement d'un ensemble d'agents intégrés dans un réseau afin d'atteindre un objectif prédéfini. Voici quelques exemples de ces situations :

- Une entreprise qui tente d'exploiter les externalités de consommation entre les agents afin de maximiser le profit.
- Une administration locale qui souhaite influencer sur le niveau de criminalité au sein d'un réseau de criminels afin de réduire le niveau global de criminalité dans une ville donnée.
- Une administration locale qui souhaite influencer sur le niveau d'effort des élèves d'une même classe afin d'augmenter le niveau global de réussite.

Bien que les problèmes de ciblage appartiennent à une littérature bien établie, la nouveauté de la présente recherche réside dans l'intégration de données sociales anonymes plutôt que

de données sociales complètes. Un tel exercice est pertinent car l’anonymisation des données est souvent présentée comme la solution aux atteintes à la vie privée. Dans notre modèle, le planificateur peut observer la topologie du réseau mais ne peut pas connaître l’identité de chaque agent. Pour revenir aux trois exemples ci-dessus, cela correspondrait à une entreprise européenne qui achète des données sur les liens sociaux des utilisateurs d’une plateforme donnée à des courtiers mais qui, en raison de la réglementation (par exemple, RGPD), ne peut pas observer l’identité. Dans le cas du réseau criminel, connaître la topologie du réseau reviendrait à observer les flux de communication entre les individus sans pouvoir dévoiler l’identité des destinataires et des destinateurs.

La contribution principale du chapitre 1 est de montrer qu’en fonction du schéma d’interaction entre les agents, la connaissance de leur identité par le planificateur peut ou non avoir de l’importance. En particulier, lorsque l’influence est réciproque, c’est-à-dire lorsque deux agents directement liés s’influencent mutuellement, la connaissance de l’identité des agents n’a pas d’importance et le planificateur peut atteindre son objectif. Cependant, pour les réseaux où il existe un écart entre l’influence reçue et l’influence exercée, nous montrons que le planificateur ne peut pas atteindre son objectif et finit par offrir des quantités uniformes dans le cas d’un monopole.

Le chapitre 2 étudie la formation de l’opinion dans les réseaux sociaux, en prenant en compte l’égalité d’attention envers les opinions exprimées. L’accès à l’internet dans des centaines de pays¹¹ et la disponibilité de multiples plateformes de réseaux sociaux ont donné à des milliards de personnes la possibilité de s’exprimer, d’exprimer une opinion, d’écouter différents points de vue et de faire partie d’une communauté mondiale. Pourtant, seuls quelques heureux élus sont entendus, comme l’affirme Cardon (2019) [34], qui avance que la visibilité sur l’internet suit une loi de puissance où 1% du contenu disponible attire plus de 90% de l’attention des utilisateurs.¹² Une poignée d’hommes politiques, d’artistes et d’universitaires, agit comme des leaders d’opinion qui créent du contenu, interagissent entre eux, influencent l’opinion publique et attirent l’attention des masses d’utilisateurs atomisés.

Tout comme les problèmes de ciblage étudiés au chapitre 1, la formation des opinions est un sujet étudié intensivement. En outre, ce sujet est ancré dans plusieurs littératures bien établies en sociologie, psychologie sociale, informatique, physique et économie. La

¹¹Selon [Wikipedia](#), 53.5% de la population mondiale en 2019 utilisait internet.

¹²Voir le chapitre *La transformation numérique de l’espace public*, à la 147.

nouveauté du chapitre 2 réside dans la modélisation précise de l'inégalité d'attention, observée sur les médias sociaux. A cette fin, je formalise un modèle où un individu choisit d'exprimer son opinion ou de la cacher en fonction de sa popularité. Les individus qui cachent leur opinion peuvent être interprétés comme des individus qui ont une faible popularité, de sorte que même s'ils s'expriment, ils ne seront pas entendus ou pris en considération. On peut penser à un individu qui a très peu de followers sur Twitter. Une autre interprétation peut être que cacher son opinion est moins coûteux que de l'exprimer. Le coût de l'expression peut être le temps passé à se disputer avec des pairs plus éloquentes et persuasifs ou le coût de l'isolement social lorsque l'opinion d'un individu s'écarte du point de vue du groupe moyen. En outre, il existe un certain nombre d'études empiriques qui soutiennent l'idée que les individus idéologiquement opposés interagissent souvent ensemble (par exemple, Conover et al. (2011) [39]). J'inclus désormais cette caractéristique dans mon modèle. L'étude des interactions entre des individus populaires au niveau local qui interagissent avec des pairs avec pensée similaire ou idéologiquement opposés peut indiquer clairement si le consensus ou la polarisation prévaut. Comme l'influence est plus forte localement, des clusters peuvent se former. Mais certains membres d'un groupe donné qui sont assez populaires et qui interagissent avec des pairs idéologiquement opposés peuvent tomber en désaccord, ce qui entraîne une polarisation des opinions entre les groupes. Je définis une mesure de popularité locale et je l'utilise pour des simulations afin d'illustrer les résultats de mon modèle.

La principale contribution du chapitre 2 est de fournir un cadre unificateur qui explique l'émergence de plusieurs catégories d'opinion, telles que le consensus, la polarisation et le désaccord total. L'élément clé est l'étude des schémas d'interaction entre les personnes qui s'expriment car ils définissent les résultats de la dynamique des opinions. Je montre en particulier que, si les opinions à long terme sont bi-polarisées, alors il n'existe pas d'individus qui reçoivent une influence égale de deux groupes d'opinion extrêmes idéologiquement opposés. Ces individus occupent une place très particulière dans le réseau car ils ne sont pas assez populaires localement pour s'exprimer et ils sont également influencés par deux groupes d'expression idéologiquement opposés. Ces personnes pourraient être interprétées comme des animateurs de télévision neutres, des journalistes ou des intermédiaires non biaisés en général. Ces personnes n'existent généralement pas dans les réseaux de communication modernes tels que Twitter où les hommes politiques peuvent se mentionner mutuellement et interagir directement. Alors que dans les décennies passées, les débats en-

tre hommes politiques, ouverts au public, étaient généralement modérés par des animateurs de télévision.

Le chapitre 3, co-écrit avec Alexia Lochmann, modélise le comportement des individus lorsque l'identité culturelle est en jeu et lorsque les individus interagissent dans deux contextes différents avec leurs pairs. Nous fournissons de nouvelles preuves empiriques basées sur des données historiques. Akerlof et Kranton (2000) [4] dans leur ouvrage de référence *Identity Economics* ont clairement indiqué que l'identité et les normes sont des déterminants importants pour la prise de décision. Ils affirment que l'identité peut être une source de motivation importante. En outre, le comportement individuel est affecté et façonné par de nombreuses interactions simultanées, simplement parce que les individus sont intégrés dans plusieurs réseaux. Par exemple, les relations d'échange ou les liens d'amitié peuvent influencer le comportement de différentes manières. Wasserman et Faust (1994) [97] dans la dernière section de leur manuel standard largement utilisé *Social Network Analysis* soulignent "qu'un domaine où il est clairement nécessaire de poursuivre les travaux est l'élaboration de méthodes pour étudier les relations multiples" (traduction personnelle, page 730). Malgré l'existence de certains travaux universitaires sur cette question, il n'y a pas de consensus sur la façon de modéliser les relations multiples et de penser la centralité lorsque plusieurs réseaux incluant les mêmes acteurs sont étudiés simultanément. Par conséquent, la grande majorité des articles d'économie traitant des réseaux sociaux et économiques ont tendance à se concentrer sur un seul type de relations et à le modéliser à l'aide d'un simple graphique.

Dans notre travail, nous fournissons un modèle simple où chaque individu a un type culturel et est intégré dans deux réseaux. Notre modèle est encadré en termes de sélection dans la migration. La nouvelle preuve que nous fournissons est basée sur la numérisation de données historiques au niveau des ménages du Tyrol du Sud, la province la plus septentrionale d'Italie, suite à un accord entre Hitler et Mussolini en 1939 qui donnait aux Tyroliens du Sud l'option de quitter leur patrie et de garder leur culture (à cette époque, ils/elles étaient tous germanophones) ou de rester en Italie et d'abandonner leur culture. Bien qu'elle semble spécifique à la littérature sur les migrations, l'approche de modélisation est pertinente pour l'étude de l'identité culturelle dans les réseaux sociaux multiplex, les gens interagissent ensemble en utilisant plusieurs outils médiatiques impliquant une corrélation des comportements sur plusieurs plateformes.

Pour étudier les composantes culturelles, sociales et économiques de l'auto-sélection dans la migration, nous formalisons un jeu avec des ménages stratégiques qui déclarent leurs intentions d'émigrer. Chaque ménage a un type culturel qui est son degré d'attachement à la culture germanique. Les ménages interagissent avec d'autres ménages au sein de leur cercle social, y compris les amis, les collègues de travail et les voisins. Le cercle social de chaque ménage représente son capital social. En outre, pour tenir compte des facteurs économiques, nous considérons la propriété qui est l'élément économique le plus pertinent dans une société agricole. Les propriétés des voisins qui décident d'émigrer sont des opportunités économiques potentielles qu'un ménage donné peut saisir. Nous montrons qu'en ne considérant que les interactions dans le réseau social, si la plupart des contacts sociaux d'un ménage donné ont l'intention d'émigrer, alors le ménage est également plus susceptible d'émigrer, sinon il perd son capital social. En outre, indépendamment des interactions dans le réseau social, les ménages sont moins incités à émigrer lorsque tous leurs voisins du réseau géographique ont l'intention d'émigrer. En effet, ils peuvent acheter la maison du voisin et améliorer leur bien-être économique. Deux individus qui sont liés à la fois dans le réseau social et dans le réseau géographique (par exemple, un voisin géographique qui est aussi un collègue de travail sympathique) ont une influence positive sur les décisions de migration de l'autre, mais plus la valeur de la propriété du voisin est élevée, plus l'effet est faible. Nous étudions un exemple exhaustif pour montrer comment les positions des réseaux se répercutent sur les gains migratoires nets. Un individu donné qui est central dans le réseau social mais périphérique dans le réseau géographique, a des gains migratoires négatifs lorsque ses contacts sociaux ont des types culturels faibles impliquant des intentions d'émigrer relativement faibles, même si l'individu lui-même a une culture germanique relativement forte.

(ii) *Revue générale de la littérature*

La présence croissante des réseaux depuis les années 1990 dans les modèles économiques a presque suivi l'évolution de la technologie numérique et l'interconnexion croissante du monde. L'inclusion de modèles d'interaction est devenue primordiale pour l'étude de plusieurs phénomènes contemporains ayant des conséquences directes sur l'économie tels que : la contagion financière, la résilience des systèmes économiques, le commerce international, les chaînes d'approvisionnement, la mobilité professionnelle, la mobilité humaine, les opinions

publiques, les mouvements sociaux, les accords sur le changement climatique, la coopération transnationale, les conflits, la diffusion de maladies, l'adoption de technologies, etc. La littérature peut être largement divisée en deux grands volets. Le premier volet concerne les jeux sur les réseaux (voir Bramoullé et Kranton (2016) [32] et Zenou (2016) [101]). Il étudie le comportement des agents en interaction dans un réseau fixe. Il met généralement en relation la position des agents décrits par une mesure de centralité du réseau avec l'action choisie et l'ampleur des externalités générées. Le second volet traite de la formation des réseaux (voir Mauleon et Vannetelbosch (2016) [75]). Cette littérature traite généralement des notions de stabilité des liens formés et de l'efficacité des réseaux formés dans une perspective sociétale. Enfin, il existe une littérature récente qui mélange les deux volets où l'on étudie simultanément les actions des agents d'un réseau et le choix des personnes avec lesquelles interagir (voir Vega-Redondo (2016) [96]).

La présente thèse s'inscrit dans le premier volet car dans les trois chapitres, le réseau est fixé. Le chapitre 1 s'appuie sur une ramification de cette littérature, qui étudie les problèmes ciblés dans les réseaux fixes - tels que la discrimination par les prix avec des réductions de prix à certains consommateurs en présence d'externalités de consommation (voir Bloch et Quérou (2013) [23] et Bimpikis et al. (2012) [33]) ou de récompenses inégales pour les efforts en équipe avec des synergies (voir Galeotti et al. (2019) [50]) ou d'injection de fonds dans le réseau financier à des agents spécifiques (Demange (2016) [43]). Cette littérature suppose que le planificateur et les agents disposent d'informations complètes sur la structure d'interaction. Si cette hypothèse peut être pertinente dans les petits réseaux (par exemple les banques), elle peut naturellement être remise en cause pour les réseaux ayant un nombre d'acteurs beaucoup plus élevé. L'hypothèse de l'information complète peut être assouplie de plusieurs façons. Dans le chapitre 1, nous supposons que la topologie du réseau est connue mais pas l'identité des agents. D'autres approches consistent à fournir au planificateur des statistiques agrégées sur le réseau, telles que le degré de distribution du réseau (Fainmesser et Galeotti (2015) [45]) ou le processus de génération du réseau (Wilder et al. (2018) [60]).

Le chapitre 2 est intégré dans une littérature qui étudie le partage de l'information et la formation de l'opinion. Les modèles économiques d'apprentissage en réseau contiennent deux grandes familles. La première est l'apprentissage bayésien, où des agents rationnels peuvent traiter les informations qu'ils reçoivent de leurs voisins. Cette approche requiert

une capacité mentale extraordinaire en raison de la complexité inhérente aux réseaux (nécessité de prendre en compte les croyances des voisins sur les croyances de leurs voisins, etc.) La deuxième famille est celle de l'apprentissage naïf et elle est souvent présentée comme une alternative crédible à l'apprentissage bayésien (voir Banerjee et al. (2018) [11]). Les agents mettent à jour leur opinion en utilisant des heuristiques simples. Les principaux modèles d'apprentissage naïf sont des modèles dans lesquels les agents mettent à jour leur opinion à chaque période, en prenant la moyenne des opinions de leurs amis à la période précédente (voir DeGroot (1974) [42], French (1956) [67] et Harary (1959) [57]). Ces deux familles de modèles d'apprentissage fournissent des résultats qui répondent aux questions suivantes : dans quelles conditions les agents parviennent-ils à un consensus d'opinion à long terme ? S'ils parviennent à un consensus, combien de temps faut-il pour converger vers ces croyances ? Les agents peuvent-ils agréger les informations disponibles dans le réseau ? La croyance consensuelle correspond-elle à la croyance correcte, lorsque l'objectif des agents est de connaître la valeur réelle d'un paramètre ? Le principal théorème de cette littérature (*The Wisdom of the Crowds*) est que les agents convergent vers la croyance correcte en l'absence de famille royale ou d'influenceurs (voir Bala et Goyal (1998) [9], Acemoglu et al. (2008) [2], Golub et Jackson (2012) [53]). Un nombre croissant de recherches enrichit cette littérature pour englober de nombreux aspects qui ont reçu peu d'attention, tels que le comportement stratégique (voir Grabish et al. (2018) [55]). Le chapitre 2 affine ces contributions en reliant la règle de mise à jour des croyances à la topologie du réseau. Je présente notamment deux types de personnes ayant des règles de mise à jour des avis différentes. Le type d'individus est déterminé par leur position dans le réseau (popularité). La première règle de mise à jour du premier type d'agents est une moyenne de la DeGroot, tandis que la seconde règle de mise à jour est une loi de mouvement qui intègre un effet attractif et un effet répulsif. Ce faisant, en fonction de la structure du réseau et des opinions initiales, les opinions à long terme peuvent former différentes catégories d'opinion, à savoir le consensus ou le désaccord.

Enfin, cette littérature économique aborde des questions spécifiques mais s'appuie également sur les recherches menées par les spécialistes des sciences sociales au cours des deux derniers siècles dans le but de comprendre la formation de l'opinion publique (et sa manipulation). Parmi les premières contributions fondamentales, on peut citer *La Psychologie des foules* de Gustave Le Bon (1895) [26], *Propaganda* d'Edward L. Bernays (1928) [17] et *Personal Influence* d'Elihu Katz & Paul Lazarsfeld (1955) [69].

Le chapitre 3, outre l'introduction d'un réseau multiplex comme mentionné précédemment, fait partie de la littérature émergente qui intègre les réseaux d'origine et de destination dans l'étude de la migration. Borjas (1987) [31] a introduit le modèle *Roy Model* pour l'étude de l'auto-sélection dans la migration. Ce modèle est un mastodonte dans la littérature sur la migration et il analyse les moteurs économiques de la migration : dans quelles conditions un individu décide-t-il de se déplacer ? La principale conclusion est que les individus décident de migrer lorsque les revenus correspondant à leur niveau de compétence sont plus élevés à l'étranger que dans leur pays d'origine. Munshi (2020) [80] fait valoir que ce modèle ne peut expliquer un certain nombre de faits stylisés qui incluent une plus grande mobilité des individus vers des destinations où d'autres personnes de leur origine se sont déjà déplacées auparavant. Blumenstock, Chi, Tan (2019) [24] explorent davantage les déterminants de la décision de migration en se concentrant sur la topologie spécifique du réseau. Leurs travaux confirment l'hypothèse selon laquelle les personnes sont plus susceptibles de migrer vers des destinations où elles ont plus de contacts sociaux et sont moins susceptibles de quitter leur lieu de résidence d'origine lorsqu'elles appartiennent à des réseaux sociaux plus vastes. En étudiant au chapitre 3 les liens du ménage simultanément dans un réseau géographique et un réseau social, nous étendons la littérature croissante sur les réseaux de migration. Cette littérature considère généralement un seul type d'interaction modélisé par un réseau. Pourtant, l'influence des décisions migratoires des ménages peut dépendre du type particulier de lien qu'ils ont entre eux.

Enfin, les trois chapitres visent à donner au lecteur de nouvelles perspectives. Notamment, l'inclusion de caractéristiques des réseaux sociaux modernes peut élargir les horizons des modèles classiques en économie des réseaux et apporter des réponses innovantes aux défis sociaux et économiques.

1. Targeting in Social Networks with Anonymized Information

(joint with Francis Bloch¹)

Abstract. This paper studies whether a planner who only has information about the network topology can discriminate among agents according to their network position. The planner proposes a simple menu of contracts, one for each location, in order to maximize total welfare, and agents choose among the menu. This mechanism is immune to deviations by single agents, and to deviations by groups of agents of sizes 2, 3 and 4 if side-payments are ruled out. However, if compensations are allowed, groups of agents may have an incentive to jointly deviate from the optimal contract in order to exploit other agents. We identify network topologies for which the optimal contract is group incentive compatible with side-payments: undirected networks and regular oriented trees, and network topologies for which the planner must assign uniform quantities: single root and nested neighborhoods directed networks.

Keywords: *social networks, targeting, privacy protection, mechanism design.*

JEL Classification: D85, D82, D42, L14.

¹Université Paris 1 Panthéon-Sorbonne and Paris School of Economics.

1.1. Introduction

Anonymizing personal and social data before they can be shared or used seems like a promising avenue, because it presumably preserves privacy without depriving the society from the benefits of data analysis. As is known from the statistical and computer science literature, privacy-preserving data disclosure can be achieved in multiple ways, from a complete elimination of any identification of the individual (anonymization) to algorithmic and statistical methods to avoid identification (differential privacy). The recent advances in privacy protection, together with the fast development of data sharing and data disclosure raises the following question: How valuable are anonymized data (rather than full individual data) for a third party whose objective is not aligned with the welfare of individuals? When will the third party be able to achieve the same value with anonymized data and with the full, personal and social data?

Anonymized social data describe the *network topology*, the architecture of the network without specifying the identity of agents at different nodes. For example, anonymized social network data can be geo-data describing a network of roads and houses with no specification of the inhabitants, or an organizational chart for a company or a criminal organization with no identification of the individuals, or a snapshot of a fraction of a large digital social network (Twitter, LinkedIn or Facebook) with lacunary evidence on the identity of the nodes.

We suppose that the social network describes local positive externalities in the choices of agents, and consider a third party whose objective is to maximize total surplus. The third party could be a monopolist, selecting discriminatory prices to exploit local consumption externalities (by lowering the price to stimulate consumption of individuals generating important externalities), or a firm who wishes to maximize output, when the social networks describes synergies in efforts by pairs of workers. In particular, agents have linear quadratic payoffs, yielding linear best-responses. In the presence of local network externalities, the objective of the planner and of individuals are not aligned: the planner internalizes positive externalities that an agent produces on other agents, raising consumption (in the case of the monopolist) or effort (in the case of the firm) beyond the optimum of individuals.

A planner can observe anonymized social data but does not know the identity of agents. To be more precise, anonymized social data are modeled by an unlabeled network structure, where each vertex corresponds to the location of a given agent, whose identity is not known

to the planner. Hence there exists a one-to-one mapping between the set of locations in the network and the set of agents. We assume that each agent only knows her location in the network. Hence the set of possible *types* is the set of all permutation of identities (labels) over the set of locations (vertices). The planner offers a menu of contracts (quantities or efforts), corresponding to the decisions at each location. Locations are correlated, in the sense that two agents can not occupy the same location and each can be assigned a single location in the network.

When full data are available, the planner can target agents and offer them the exact contract which maximizes her objective. But when the planner only knows the network topology, she is only able to offer a contract for each *location* in the network, with no certainty that agents will indeed choose the contract corresponding to that location. This is the incentive problem we study in this paper, where we analyze for which network architectures this incentive constraint is binding or not.

We first observe, unsurprisingly, that correlation in the locations of agents allows the planner to achieve his first-best outcome when agents can only make unilateral deviations. Indeed, because the set of locations in the network is fixed and finite, if the planner proposes a menu of contracts, one for each location, and all other agents pick the contract corresponding to their location, the deviating agent must be selecting the same contract as another agent. The planner can use this evidence to ascertain that one of the agents has lied, and punish all agents (for example by offering a level lower than any agent's optimum), thereby ensuring that all agents have an incentive to choose the contract corresponding to their true location.¹

We thus turn our attention to joint deviations, where subsets of agents decide to "lie" about their locations, and choose contracts which do not correspond to their true position in the network. Given that the objective of the planner is to maximize social surplus, why do groups of agents have an incentive to "lie" about their locations and exchange their contracts? Why would an agent who already receives a consumption or effort level above her optimal choice accept to exchange her contract with another agent with a higher level? The answer to these questions stems from the fact that, by exchanging their contracts, agents are able to increase the externalities they receive from other agents, while reducing the ex-

¹As usual, this reasoning only shows that truth-telling is one equilibrium of the revelation game played by the agents, but other equilibria exist as well – in fact, in this model, *any* choice of contracts by the agents is an equilibrium of the game.

ternalities they create for agents outside the deviating coalition. For example, consider a directed network architecture where the hub of the star (the “influencer”) produces positive consumption externalities on all other agents but receives no externalities in return. The planner, internalizing this externality, assigns a very high consumption to the hub, and a lower consumption to a peripheral agents. But the hub may have an incentive to exchange her contract with a peripheral agent, increasing her utility by reducing her consumption level, at the expense of all other peripheral agents, who now receive a much lower level of externalities.

The first main Proposition of the paper asserts that, when the size of the deviating group is two, three or four, the optimal contract is in fact immune to deviations by groups of agents. The proof of the Proposition in the case of pairwise deviations is clear. For an agent to accept to exchange her contract for a higher one (e.g. for a peripheral agent to accept the high level of the hub), whereas her current contract is already too high given the contracts of other agents, she must receive higher externalities in return from the members of the deviating coalition. But when a pair of agents deviates, the externalities inside the deviating coalition do not change, as agents only exchange their contracts and the externality they exert on each other remains the same. Hence, in any pairwise deviation, absent any compensation, the agent with the lower contract must be unwilling to exchange her contract with the other agent.²

Based on this observation, we consider a more restrictive group incentive compatibility condition, by allowing agents in the deviating coalition to make side-payments. Under this condition, a contract is group incentive compatible if no coalition, by choosing a permutation of the announcements of their locations, can increase the sum of utilities of its members. In the particular context of a linear-quadratic game, we show that this condition is equivalent to assuming that, whenever an agent i has a higher contract than an agent j , the sum of externalities received by agent i must be at least as large as the sum of externalities received by agent j .

This characterization drives the second main result of the paper. When externalities are reciprocal (the graph representing the social network is undirected), then the optimal contract of the planner is immune to group deviations with side-payments. The intuition un-

²Unfortunately, this intuition does not easily carry over to deviations by larger groups, where agents can benefit from a change in the level of externalities among agents in the deviating coalition. For groups of three and four agents, we are able to prove that one of the lower contract agents must be unwilling to exchange, but so far have been unable to generalize the result beyond groups of four agents.

derlying this result is that an agent will receive a higher contract in the first-best if and only if she is more central (in the sense of Katz-Bonacich) than the other agent. But this centrality measure is defined recursively: an agent has a higher Katz-Bonacich centrality if and only if the sum of the Katz-Bonacich centralities of her neighbors is higher. In other words, when influence is reciprocal, an agent with a higher contract in the first-best is also an agent who receives (and produces) higher externalities, and the sum of utilities of the agents cannot increase through an exchange of contracts.

The situation changes dramatically when externalities are asymmetric. We next focus on directed networks with a hierarchical structure: the set of agents can be partitioned into an ordered collection of subsets (or tiers of the hierarchy) such that agents at higher tiers influence agents at lower tiers, but agents at lower tiers never influence agents at higher tiers. In this family of hierarchical social networks, we first show that the optimal contract is group incentive compatible when side-payments are ruled out. But when agents in a deviating coalition can make side-payments, we uncover two situations where the optimal contract of the planner cannot be sustained.

First, when there exists one agent who influences all other agents *the single root* and the number of tiers is larger than two, then in an optimal contract which is immune to group deviations with side-payments, *all agents in tiers lower than two must receive the same contract as the root*. The intuition is as follows: if the root receives a higher contract than agents at lower tiers, there is an incentive to exchange the higher quantity of the root with the lower quantity of the lower tier agent, as the latter receives externalities but not the former. Hence, any contract which is immune to deviations by groups with side-payments must assign a lower contract to the root than to agents at tiers lower than two.³ However, as the root produces externalities on all other agents, whereas other agents may not influence or be influenced by all other agents, the planner also has an incentive to increase the contract of the root with respect to the contract of any other agent. Hence, at the optimum, the contract of the root must be exactly equal to the contract offered to all agents at tiers lower than two. Notice that this implies that targeting is impossible, and may result in large efficiency losses with respect to the first-best.

Second, suppose that *neighborhoods are nested*, so that an agent at a lower tier is influenced by all agents who also influence agents who influence him. Multiple roots can exist, and

³For agents at tier two, the only externalities they receive come from the root, and hence there is no restriction on the relative ranking of their contract with the contract of the root.

agents can be connected to multiple roots, but the structure of nested neighborhoods implies that the set of influencers of an agent is included in the set of influencers of the agent she influences. Assume furthermore that the size of the tiers is increasing: there are strictly more agents at lower tiers. In this hierarchical structure, as in the case of single roots, we show that the optimal contract is not group incentive compatible with side-payments, and that any root has an incentive to exchange her contract with an agent at lower tier. As in the case of a single root, the planner would like to increase the quantity of the root, so that *the optimal contract which is group incentive compatible with side-payments results in a uniform contract for all agents.*⁴ Again, the group incentive compatibility constraint is binding, and induces a total inability to discriminate among agents.

Finally, we consider a hierarchical structure where the optimal contract is immune to deviations by coalitions making side payments: *regular oriented tree*. In these structures, every agent is influenced by a single other agent; namely, the social network is an oriented tree. Furthermore, the number of agents influenced by an agent at any given tier is identical across agents of the same tier, and strictly decreasing with the tier. This is what we call a “regular” oriented tree. In that specific hierarchy, agents at higher tier must have a higher contract than the agents they influence, because every agent they influence receives his externalities from his immediate predecessor, but the predecessor receives externalities from other agents. When the structure is regular, the planner will also want agents at higher levels of the hierarchy to receive higher contracts, so that the first-best contract is indeed group incentive compatible with side-payments.

Our results thus show that the optimal contract of the planner is immune to deviations by groups whenever influence is reciprocal, or agents at higher levels of the hierarchy receive more influence than agents at lower levels of the hierarchy. When on the other hand agents at higher levels of the hierarchy produce more influence but receive less influence than agents at lower tiers, the first-best contract is susceptible to group deviations, and the planner may in extreme cases be forced to assign a uniform contract to all agents. These results thus show how the architecture of the network matters to determine whether anonymized data are sufficient or not for the planner to implement the first-best.

We also explore the robustness of our results with respect to the assumptions made in

⁴This time, under the assumption that agents at tier two are influenced by more than one root, they receive more externalities than a root, and hence the reasoning which implies that an agent must receive a higher contract than the root in a group incentive compatible contract with side-payments also holds for tier two agents.

the model. First, we note that we put a very strong restriction on the structure of deviating coalitions by assuming that they are adjacent (i.e. any pair of agents in the deviating coalition must be connected by a link). Absent this restriction, more coalitions can deviate and in fact, even in the case of undirected networks and regular oriented trees, the first-best contract may become impossible to implement.

Second, we analyze the effect of increasing the complexity of the contract, by allowing the planner to extract information about agents' local neighborhood. We show that if the planner can collect information about an agent's neighbors' identity, the first-best contract becomes immune to group deviations with side-payments. Finally, we explore what happens when the planner knows the identity of some agents in the network. For example, suppose that the network knows the identity of the main influencers (the roots of the directed social network). We show through an example that this partial information may greatly help the planner and allow him to discriminate much more effectively among the agents.

Related literature. Our paper is related to the study of targeting in social networks, as surveyed in Bloch [22]. The linear-quadratic model of interaction in networks we consider was introduced in Ballester, Calvo-Armengol and Zenou [10]. General problems of targeting where the planner seeks to maximize social welfare in the presence of complementarities have been recently studied by Demange [43] and Galeotti, Golub and Goyal [50].

The problem of a monopolist pricing in the presence of local consumption externalities has first been studied by Candogan, Bimpikis and Ozdaglar [33] and Bloch and Qu erou [23]. Candogan, Bimpikis and Ozdaglar [33] and Bloch and Qu erou [23] both look at the relation between the centrality of consumers in a given network and the prices and quantities they are offered by a monopolist; in the context of perfect knowledge of the network structure. They both show that when the network is undirected - that is two directly connected consumers influence each other equally - consumers are offered quantities proportional to their Bonacich centrality for the same price. Fainmesser and Galeotti [45] consider a monopolist which can price discriminate based on her knowledge of either the in-degrees of consumers (their influence), the out-degree of consumers (their susceptibility to influence) or both. They show that the knowledge of in-degree (respectively out-degree) is more valuable when the dispersion in in-degrees (respectively out-degree) in the network is higher.

A handful of recent papers extend the analysis of monopoly pricing and targeting to situations of incomplete information. Jadbabaie and Kakhbod [63] consider the pricing problem

when the strength of influence among agents is private information. Ata, Belloni and Candogan [8], take a different approach and assume that there is a seller who faces two groups of agents: observable and latent. There are local consumption externalities among members of the same group and across groups. The seller can observe past purchasing decisions only of the group of *observable* agents and can try to deduce from these observations information on influence patterns. Finally, in a paper which shares our motivation about the planner's incentives to elicit information about the network, Shi and Xing [90] consider a model where buyers draw their in and out-degrees independently from a common distribution. They characterize the optimal contracts in the framework of random graphs. The novelty in our model is to introduce incomplete information with regard to the network itself by assuming that the structure is observable but the locations or identities of agents is private information.

1.2. The model

1.2.1. Agents, utilities and network effects

We consider a set N of agents (labeled "he"), indexed by $i = 1, 2, \dots, n$. Agents are connected in a social network, which we represent by a graph g , with nodes $i = 1, 2, \dots, n$ and edges $g_{ij} \in \{0, 1\}$. We denote by \mathbf{G} the adjacency matrix of the graph g .

Each agent i chooses an action $x_i \in \mathbb{R}_+$, which will be interpreted as his consumption or effort. The utility of agent i depends on his action and the actions of his neighbors in the social network. Following a well-established model, initiated by Ballester, Calvo-Armengol and Zenou [10], we assume that utilities are quadratic, so that

$$U_i = a_i x_i - \frac{b_i}{2} x_i^2 + \alpha \sum_{j \in N} g_{ij} x_i x_j,$$

where $\alpha > 0$ is a strictly positive externality parameter.⁵ For the main part of the analysis, we assume that all agents are identical, except for their location in the social network. For homogeneous agents, $a_i = a$ and we normalize $b_i = b = 1$ so that

⁵With $\alpha > 0$, the payoff function yields a game of pure complements. However, had the the externality parameter been strictly negative, the payoff function would yield a game of pure substitutes (e.g. Cournot game with n firms) and our results would change substantially. For a discussion, see Bramoulle and Kranton (2016) [32].

$$U_i = ax_i - \frac{1}{2}x_i^2 + \alpha \sum_{j \in N} g_{ij}x_i x_j.$$

1.2.2. The planner's first-best

We consider a planner (labeled "she") with perfect knowledge of the network g who chooses the vector of actions $\mathbf{x} = (x_1, \dots, x_n)$ in order to maximize the sum of utilities of the agents,

$$V = \sum_{i \in N} U_i = \sum_{i \in N} ax_i - \frac{1}{2}x_i^2 + \alpha \sum_{j \in N} g_{ij}x_i x_j, \quad (1.1)$$

where $\alpha > 0$. The first-order conditions result in a system of linear equations:

$$a - x_i + \alpha \sum_{j \in N} (g_{ij} + g_{ji})x_j = 0$$

Letting $\mathbf{1}$ denote an $n \times 1$ vector of ones, $\mathbf{0}$ an $n \times 1$ vector of zeros and \mathbf{I} the $n \times n$ identity matrix, we can rewrite the above conditions in matrix⁶ form:

$$a\mathbf{1} - \mathbf{I}\mathbf{x} + \alpha(\mathbf{G} + \mathbf{G}^T)\mathbf{x} = \mathbf{0}.$$

Let λ denote the largest eigenvalue of the matrix real symmetric matrix $(\mathbf{G} + \mathbf{G}^T)$. Following Ballester, Calvo-Armengol and Zenou [10] we compute the optimal solution of the planner's problem as the solution to the system of linear equations when external effects are not too large,

Proposition 1.1 *If $1 > \alpha\lambda$ the optimal solution of the planner's problem is given by*

$$\mathbf{x}^* = (\mathbf{I} - \alpha(\mathbf{G} + \mathbf{G}^T))^{-1}a\mathbf{1}.$$

Next recall that the *Katz-Bonacich* (Katz [70] and Bonacich [27]) centrality measure $\beta(\mathbf{R}, \delta)$ of an agent in a matrix of relationships R with discount factor δ , is the discounted sum of

⁶Rearranging we get $(\mathbf{I} - \alpha(\mathbf{G} + \mathbf{G}^T))\mathbf{x} = a\mathbf{1}$. If $(\mathbf{I} - \alpha(\mathbf{G} + \mathbf{G}^T))$ is a nonsingular matrix the unique solution is $\mathbf{x} = (\mathbf{I} - \alpha(\mathbf{G} + \mathbf{G}^T))^{-1}a\mathbf{1}$. Recall that *M-matrices* are real nonsingular matrices such that all their entries are non-positive and all the entries of their inverse are non-negative. From Meyer (2000) [78], A is an *M-matrix* if and only if there exists a matrix $B \geq 0$ and a real number $r > \rho(B)$ s.t. $A = r\mathbf{I} - B$, with $\rho(B)$ the spectral radius of B .

walks originating from that agent. Formally, for $\delta\lambda < 1$ where λ is the largest eigenvalue of R ,

$$\begin{aligned}\beta(\mathbf{R}, \delta) &= \sum_{k=0}^{\infty} \delta^k \mathbf{R}^k \mathbf{1}, \\ &= (\mathbf{I} - \delta \mathbf{R})^{-1} \mathbf{1}.\end{aligned}$$

It is easy to check that, the planner's solution is to assign to each agent i an action which is proportional to its Katz-Bonacich centrality measure in the network $(\mathbf{G} + \mathbf{G}^T)$ under discount factor $\alpha > 0$,

$$\mathbf{x}^* \propto \beta((\mathbf{G} + \mathbf{G}^T), \alpha).$$

Hence the optimal choice of the planner is to assign discriminatory actions to the agents, in proportion to their Bonacich centrality measure in the network⁷ $(\mathbf{G} + \mathbf{G}^T)$. We now observe that the planner's general problem encompasses different situations of pricing and targeting in networks.

Pricing with network externalities

Consider, as in Candogan, Bimpikis and Ozdaglar [33] and Bloch and Qu  rou [23] a monopolist setting prices on a market where consumers experience positive consumption externalities. For simplicity, assume that the monopolist faces a constant marginal unit cost c . Consumer i 's utility depends both on his own consumption and on the consumption of his neighbors in the social network. If the monopolist sets a unit price p_i to consumer i , and offers a quantity x_i , the utility of the consumer is given by

$$U_i = ax_i - \frac{1}{2}x_i^2 + \alpha \sum_{j \in N} g_{ij}x_i x_j - p_i x_i.$$

Assuming that the monopolist can fully discriminate among consumers, she will select a price p_i to capture the entire surplus of the consumer and make a profit

⁷Strictly speaking, $(\mathbf{G} + \mathbf{G}^T)$ is a weighted network, with weights equal to 0, 1 or 2 on each link. Whenever G is directed, $(\mathbf{G} + \mathbf{G}^T)$ could be interpreted as the underlying graph of G . Recall that Ballester, Calvo-Armengol and Zenou [10] assume that the network G is undirected hence in their paper $\beta(G + G^T, \alpha) = \beta(2G, \alpha)$. While here we are precisely interested in considering networks G where some agents receive more influence than they exert over neighbors and vice versa.

$$\pi = \sum_{i \in N} (p_i - c)x_i = \sum_{i \in N} \left((a - c)x_i - \frac{1}{2}x_i^2 + \alpha \sum_{j \in N} g_{ij}x_ix_j \right),$$

an objective function which is equivalent to the objective function V in equation (1.1) after a renormalization.

Targeting and efforts

Next we consider as in Galeotti, Golub and Goyal [50] the decision x_i chosen by an agent with utility function

$$U_i = ax_i - \frac{1}{2}x_i^2 + \alpha \sum_{j \in N} g_{ij}x_ix_j.$$

The planner intervenes by subsidizing or taxing the marginal stand-alone utility of the effort, resulting in $\hat{a}_i = a + t_i$, where t_i is the tax or subsidy of agent i . Any intervention leading to a change from a to \hat{a}_i will result in a change in the effort level x_i . Hence an intervention can be interpreted as a choice of decisions x_i to maximize the sum of utilities of the agents

$$V = \sum_{i \in N} \left(ax_i - \frac{1}{2}x_i^2 + \alpha \sum_{j \in N} g_{ij}x_ix_j \right).^8$$

1.2.3. Network architectures and information structure

Proposition 1.1 characterizes the first-best solution when the planner has complete information on the social network. We now introduce the constraints faced by the planner when she only has access to *anonymized* information on the network architecture. To be more precise, vertices in a network correspond to locations $\ell = \{l_1, l_2, \dots, l_n\}$. There is a one-to-one mapping ρ which assigns to each agent in the set $N = \{1, 2, \dots, n\}$ a unique location in the network. This mapping is unknown to the planner and this is precisely how we model anonymized information. Hence, the set of types is the set of all permutations of agents in the set N over locations ℓ ; which is exactly the set of all vertex-labelled graphs. Formally,

⁸Note that Galeotti, Golub and Goyal [50] assume the planner faces a fixed budget and incurs a cost which is quadratic in the difference between the initial value a and the target value \hat{a}_i . This results in an additional constraint in the planner's problem. When the budget is sufficiently large, this additional constraint is not binding and the problem is equivalent to the problem we consider here.

the planner observes a network structure, which we denote by $\hat{g}(\ell)$, where an *identifier* in ℓ is assigned to each vertex. Agents (*i*) know their own location, that is a given agent *i* knows at which location in ℓ he is at, and (*ii*) observe the locations in ℓ that correspond to his neighborhood and the neighbors of neighbors, without knowing the identity of any of them.

1.2.4. A menu of contracts

In the benchmark case, we restrict the planner to offer a menu of contracts. The planner offers a menu of contracts $\mathbf{x}(\hat{g}(\ell)) = \{x_{\ell_1}, \dots, x_{\ell_n}\}$, corresponding to the decisions at each location in the observed network structure \hat{g} . Every agent then selects a contract among the menu $\mathbf{x}(\hat{g}(\ell)) = \{x_{\ell_1}, \dots, x_{\ell_n}\}$. If all agents select different contracts, each agent obtains the contract he selected. Otherwise, if several agents select the a contract for the same location, the planner punishes all agents and chooses an outcome $\mathbf{x}(\hat{g}(\ell)) \equiv \{\mathbf{0}\}$.

The menu of contracts $\mathbf{x}(\hat{g}(\ell)) = \{x_{\ell_1}, \dots, x_{\ell_n}\}$ is to ex-post incentive compatible for any network structure \hat{g} , whenever agents have an incentive to select the contract intended for their location rather the contract designed for any other location. Formally

Definition 1.1 Denote by l_i be the true location of agent $i \in N$. A menu of contracts $\mathbf{x}(\hat{g}(\ell)) = \{x_{\ell_1}, \dots, x_{\ell_n}\}$ is ex-post incentive compatible at a network architecture \hat{g} if for any i , for any $l_i \neq l_j \in \ell$,

$$U_i(\mathbf{x}(\hat{g}(l_i, l_{-i}))) \geq U_i(\mathbf{x}(\hat{g}(l_j, l_{-j}))).$$

As we will see, correlation of agents' types makes it easy for the planner to construct ex-post incentive compatible contracts. We thus extend the definition of ex-post incentive compatibility to allow for joint deviations by coalitions of players. A set of agents can deviate by exchanging among themselves the contracts intended for each others' locations. In the benchmark model, we assume that coalition S must contain adjacent players. Formally

Definition 1.2 A coalition S contains adjacent players if and only if, for all $i, j \in S$, $g_{ij} + g_{ji} \geq 1$.

The idea underlying this requirement is that, in order to engineer a joint deviation, players in a coalition must communicate with each other, and hence be directly connected in the

social network. In Section 3.6, we will also discuss what happens when coalitions of nonadjacent players can misreport, and show that this may result in additional constraints on the contract.

We assume that if a coalition S deviates, the identities of all members of S are revealed to each other, so that agents in the deviating coalition can exactly evaluate their payoff following the deviation. (This assumption is reminiscent of the definition of the “fine core” in Wilson (1978) [98], where agents pool their information to engineer a deviation in an exchange economy with incomplete information). Absent this assumption, agents would be unable to identify the quantities consumed by their neighbors and hence the level of externalities they enjoy. Notice that agents in S will thus refine their information after the deviation is proposed. However, as the planner is unaware that the coalition S forms and jointly misreports, she will be unable to exploit this fact in the design of the contract.

Definition 1.3 *A contract $\mathbf{x}(\hat{g}(\ell)) = \{x_{\ell_1}, \dots, x_{\ell_n}\}$ is ex-post group incentive compatible at a network architecture \hat{g} if there does not exist a coalition $S \subset N$ containing adjacent players and a mapping ρ_S from S to S such that, for every agent $i \in S$,*

$$U_i((x_{\ell_{\rho(j)}})_{j \in S}, (x_{\ell_j})_{j \notin S}) > U_i((x_{\ell_j})_{j \in S}, (x_{\ell_j})_{j \notin S}),$$

where we abuse notation and write $(x_{\ell_{\pi(j)}})_{j \in S}$ to denote the actions of players in the coalition S when they exchange among each other their contracts according to the permutation π , and $(x_{\ell_j})_{j \notin S}$ the actions of players outside the coalition S . We will argue that this condition makes deviations very hard, because it requires all players to be strictly better off from the joint misreporting. A weaker notion, allowing for side-payments across players in the deviating coalition S , will result in easier deviations (thereby making positive results on the optimal contract even stronger). Intuitively, some players in the coalition individually may not make a higher payoff, but if the sum of payoffs when they exchange their contracts is strictly positive, then the collective gain made by the coalition can be distributed among its members.

Definition 1.4 *A contract $\mathbf{x}(\hat{g}(\ell)) = \{x_{\ell_1}, \dots, x_{\ell_n}\}$ is ex-post group incentive compatible with side-payments at a network architecture \hat{g} if there does not exist a coalition S containing adjacent*

players and a mapping ρ from S to S such that,

$$\sum_{i \in S} U_i((x_{l_{\rho(j)}})_{j \in S}, (x_{l_j})_{j \notin S}) > \sum_{i \in S} U_i((x_{l_j})_{j \in S}, (x_{l_j})_{j \notin S}).$$

The menu of contracts only uses information about agents' locations, and not any additional information that the agents can have about the network (such as the identity of their neighbors). In Section 3.6, we explore more complex contracts when agents have additional information on the network, and show that the planner will often be able to implement her first-best contract, exploiting the correlation between agents' types. However, we defend the menu of contracts as a contract which is a realistic approximation of the contracts offered in reality. In particular, this contract only relies on information on the network topology (the locations of the different nodes) and does not require any additional information, such as the identity of the agents. In addition, as we show below, if the agents only know their location in the network, and do not have any other information, the menu of contracts is equivalent to any other mechanism chosen by the planner.

Proposition 1.2 *Suppose that the agents know their location in the network. If the menu of contracts $\mathbf{x}(\hat{g}(\ell)) = \{x_{\ell_1}, \dots, x_{\ell_n}\}$ is incentive compatible then every agent i chooses the contract corresponding to his location, i.e.*

$$U_i(\mathbf{x}(\hat{g}(\ell))) \geq U_i(\mathbf{x}(\hat{g}(\ell_j, \ell_{-i}))), \forall \ell_j \neq \ell_i.$$

If the mechanism is ex-post group incentive compatible then there does not exist a coalition S of adjacent players and a mapping ρ_S from S to S such that, for every agent i in S ,

$$U_i((x_{l_{\rho(j)}})_{j \in S}, (x_{l_j})_{j \notin S}) \geq U_i(\mathbf{x}(\hat{g}(\ell))).$$

If the mechanism is ex-post group incentive compatible with side-payments then there does not exist a coalition S of adjacent players and a mapping ρ_S from S to S

$$\sum_{i \in S} U_i((x_{l_{\rho(j)}})_{j \in S}, (x_{l_j})_{j \notin S}) > \sum_{i \in S} U_i(\mathbf{x}(\hat{g}(\ell))).$$

1.2.5. Efficient incentive compatible contracts

When agents do not jointly deviate, the planner can exploit correlation across types to implement an efficient contract. Whenever more than one agent selects the same contract intended for a given location, showing that one agent has not selected the contract not intended for his location, the planner chooses to punish all agents. More specifically we prove the following result.

Proposition 1.3 *Any contract $\mathbf{x}(\hat{g}(\ell))$ where agents obtain positive payoffs is ex-post incentive compatible at any network architecture \hat{g} .*

Proof of Proposition 1.3: Suppose two agents $i \neq j \in N$ select the same contract x_{ℓ_i} , the planner chooses to assign 0 to both. Then, for any agent $i \in N$ at location l_i ,

$$U_i(\mathbf{x}(\hat{g}(\ell_i, \ell_{-i}))) = ax_i - \frac{1}{2}x_i^2 + \alpha \sum_{j \in N} g_{ij}x_i x_j \geq 0 = U_i(\mathbf{x}(\hat{g}(\ell_j, \ell_{-i}))),$$

showing that the ex-post incentive compatibility constraint is always satisfied. \square

In particular, the optimal contract $\mathbf{x}^*(\hat{g}(\ell))$ is incentive compatible. In the remainder of the paper, for simplicity when there is no confusion about the type ℓ , we write \mathbf{x}^* instead of $\mathbf{x}^*(\hat{g}(\ell))$.

1.2.6. Group incentive compatible contracts

As individuals cannot gain by individually deviating and selecting a contract that is not intended for their true location, we now turn our attention to deviations by groups of agents. First notice, as in the proof of Proposition 1.3, that if an agent $i \in S \subset N$ selects a contract intended for a location of an agent j who does not belong to S , the planner will be able to detect that one of the agents has lied, and punish all agents with a decision 0. Hence, if a deviation by a group of agents in S is profitable, it must involve a selection of contracts which is a permutation of the contracts of agents inside the coalition S . We obtain the following important result:

Proposition 1.4 *The optimal contract \mathbf{x}^* is group incentive compatible for any network architecture \hat{g} if the size of the group $|S|$ is 2, 3 or 4.*

Proof of proposition 1.4: the proof is moved to Appendix 1.7.1.

Proposition 1.4 shows that no coalition of size 2, 3 or 4 can construct a group deviation, which makes all agents strictly better off. The logic of the Proposition is transparent in the case of deviations by pairs, when $|S| = 2$. The optimal contract \mathbf{x}^* always assigns quantities which are higher than the optimal quantities of the agents, as the planner takes into account the positive externalities resulting from higher quantities. Hence, for a fixed set of quantities of the other players k , a player i has no incentive to accept the quantity offered to player j when $x_i^* < x_j^*$. When a pair (i, j) of players exchanges quantities, the externalities they experience from each other remain identical, equal to $x_i^* x_j^*$. Hence there is no possibility for both players to benefit from the exchange as one of them must be accepting a higher quantity than the quantity he obtains by reporting truthfully his type. The same type of argument applies in a more complex way when $|S| = 3$ or $|S| = 4$ by considering different joint ways of misreporting, and focusing attention on those agents who trade their quantity x_i^* for a higher quantity. When $|S|$ is strictly greater than 4, whether the optimal contract is group incentive for arbitrary graph architectures remains an open question. Proposition 1.4 suggests that asking for all agents to increase their payoff by a joint deviation might be too demanding. Hence it will often be important to study the weaker deviation concept, when side-payments are allowed.

A contract $\mathbf{x} = (x_1, \dots, x_n)$ is group incentive compatible with side-payments if there does not exist a coalition S of adjacent players and a permutation ρ_S of agents in S such that

$$\sum_{i \in S} \left(ax_{\rho_S(i)} - \frac{1}{2} x_{\rho_S(i)}^2 + \alpha \sum_{j \notin S} g_{ij} x_{\rho_S(i)} x_j \right) > \sum_{i \in S} \left(ax_i - \frac{1}{2} x_i^2 + \alpha \sum_{j \notin S} g_{ij} x_i x_j \right)$$

To interpret the last condition, notice that by definition

$$\sum_{i \in S} \sum_{j \in S} g_{ij} x_{\rho_S(i)} x_{\rho_S(j)} = \sum_{i \in S} \sum_{j \in S} g_{ij} x_i x_j.$$

In addition, as agents are homogeneous, $\sum_{i \in S} (ax_{\rho_S(i)} - \frac{1}{2} x_{\rho_S(i)}^2) = \sum_{i \in S} (ax_i - \frac{1}{2} x_i^2)$, and the condition for group incentive compatible contracts with side-payments reduces to

$$\sum_{i \in S} \sum_{j \notin S} g_{ij} x_{\rho_S(i)} x_j \leq \sum_{i \in S} \sum_{j \notin S} g_{ij} x_i x_j \quad \forall S, \quad \forall \rho_S \quad (1.2)$$

1.3. Undirected Networks

In this Section, we consider undirected network architectures, where agents' influence is reciprocal, and the adjacency matrix G is symmetric. This corresponds for example to team production with synergy across workers, or consumption externalities for agents using the same software, same standard or same communication device. Our main Proposition shows that, in that case, the optimal contract is group incentive compatible with side-payments.

Proposition 1.5 *If the network g is undirected, then the optimal contract x^* is group incentive compatible with side-payments.*

Proof. See Appendix 1.7.2. ■

Proposition 1.5 shows that when influence is reciprocal, coalitions of adjacent agents cannot benefit from jointly misreporting their locations when the planner proposes the optimal contract. This result is due to the fact that in an undirected network agents with higher levels in the first-best contract are agents who receive and produce the larger externalities. Hence, constructing a group deviation against the first-best contract becomes impossible, because agents with higher levels are also agents with higher externalities. This is a very strong result because the definition of coalitional deviations is very permissive – we allow for any transfer across coalition members. The only restriction we place on the deviating coalition is that it should contain adjacent agents. As the proof demonstrates, if we only consider pairs of agents swapping their announcements, this condition is not needed: the optimal contract is robust to any deviation by pairs of players, whether they are adjacent or not. However, for larger coalitions, the restriction that only adjacent players can form a deviating coalition is meaningful, as we will illustrate in section 3.6

1.4. Hierarchical Networks

We next consider very asymmetric structures. Agents are organized in a hierarchy. Agents at higher levels of the hierarchy influence agents below them, without being influenced by them. For any pair of agents, externalities only flow in one direction, from agents at higher tiers to agents at lower tiers. Formally, we partition the set of agents into M tiers of a hierarchy, A_1, \dots, A_M with $M \geq 2$ such that $g_{ij} = 1$ if and only if $i \in A_m, j \in A_q$ and $q < m$.⁹

⁹We later consider an alternative formulation where $q \leq m$, i.e. agents in the same level of the hierarchy influence each other.

Proposition 1.6 *In a hierarchical network, the optimal contract \mathbf{x}^* is group incentive compatible.*

Proof: Suppose that there exists a coalition S of adjacent agents and a permutation ρ such that every agent i in S has a higher payoff by exchanging locations according to ρ at the optimal contract \mathbf{x}^* . There must be a set of agents S^- for whom $x_{\rho(i)}^* > x_i^*$. Among those agents, pick an agent i belonging to the highest level of the hierarchy so that $g_{ij} = 0$ for any $j \in S^-$. For that agent, $(x_{\rho(j)}^* - x_j) < 0$ for any $j \in S$, such that $g_{ij} = 1$. Hence, as $(x_{\rho(i)}^* - x_i^*) > 0$,

$$\begin{aligned} \Delta U_i &= \alpha \left[- \sum_{j \notin S} (x_{\rho(i)}^* - x_i^*) g_{ji} x_j^* - \frac{(x_{\rho(i)}^* - x_i^*)^2}{2} + \sum_{j \in S} g_{ij} x_{\rho(i)}^* (x_{\rho(j)}^* - x_j) - \sum_{j \in S} g_{ji} x_j^* (x_{\rho(i)}^* - x_i^*) \right]. \\ &< 0 \end{aligned}$$

contradicting the fact that agent i has an incentive to deviate. \square

Proposition 1.6 shows that groups of players, even if they are larger than three, cannot organize exchanges of locations which make all the agents strictly better off. Among agents who are assigned higher quantities than the first-best quantity, the agent in the highest tier of the hierarchy only enjoys externalities from agents who exchange their quantities for lower quantities. Hence the agent can surely not increase his utility by receiving a higher quantity.

1.4.1. Single root

While Proposition 1.6 shows that the optimal contract is immune to group deviations in hierarchical networks, we will now see that, contrary to the case of undirected networks, the sum of utilities of agents in a deviating coalition can increase. In order to analyze group incentive compatible contracts with side-payments, we first consider hierarchical networks with a single root.

Proposition 1.7 *Suppose that the hierarchy has a single root, $|A_1| = 1$ and that all agents are connected to the root and that the hierarchy contains more than two tiers. Then agents at tiers $m = 3, \dots, M$ must receive the same quantity as the root agent in the optimal contract satisfying group incentive compatibility with side-payments.*

Proof. See Appendix 1.7.3. ■

Proposition 1.7 shows that agents in tiers $m = 3, \dots, M$ must receive the same quantity as the root agent. Suppose by contradiction that the root agent receives a higher quantity. We claim that she has an incentive to exchange her location with any other agents at a tier $m \geq 3$, as those agents receive positive externalities from agents at tier 2, in addition to the positive externalities from the root agent, whereas the root agent does not receive any positive externality. Hence a contract which satisfies group incentive compatibility with side-payments must assign to the root agent a quantity which is lower or equal than the quantity of agents at tiers $m \geq 3$. As the sum of utilities increases when the quantity assigned to the root agent increases, this implies that the best contract must assign uniform quantities to agents in tiers $m = 1, 3, \dots, M$.

As long as not all agents are connected to each other (in which case they all receive the same quantity at the optimal contract \mathbf{x}^*), some agents must receive different quantities at the optimal contract \mathbf{x}^* . This shows that *the optimal contract is not group incentive compatible with contracts*. The incentive constraint binds and forces the planner to equalize quantities offered agents at different tiers of the hierarchy.

Interestingly, Proposition 1.7 does not restrict the quantities offered at tier $m = 2$. Agents at tier $m = 2$ have no incentive to exchange their location with the root agent, as the sum of utilities remains exactly the same after the exchange. Whether agents at tier 2 have an incentive to exchange locations with agents at lower tiers depends on the precise structure of the network, and cannot be ascertained in general. Hence, in general, the choice of quantities to agents in the second tier of the network remains unrestricted.

Finally, we note that, because quantities at the second level of the hierarchy are unrestricted, Proposition 1.7 does not hold when $M = 2$. In that case, the group incentive constraint with side-payments is not binding, and the optimal contract \mathbf{x}^* is immune to group deviations with side-payments.

We next consider a situation where either the first level of the hierarchy contains multiple roots, or agents are not all connected to the single root.

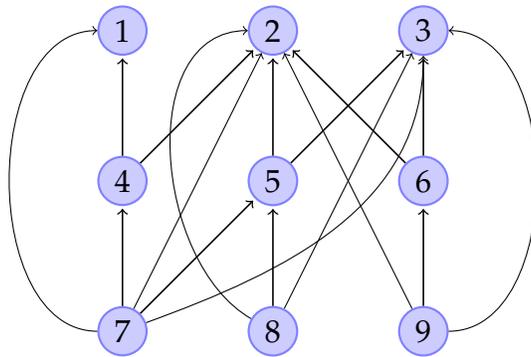
1.4.2. Nested neighborhoods

We first analyze a hierarchy with multiple roots, where neighborhoods are nested, so that agents at lower tiers of the hierarchy are influenced by the same agents as agents at higher tiers of the hierarchy.

Formally, for any three tiers m, p, q with $m < p < q$, we have

$$\forall i, \in A_q, \forall j \in A_p, \text{ such that } g_{ij} = 1, \{k | g_{ik} = 1\} \cup A_m = \{k | g_{jk} = 1\} \cup A_m.$$

This definition covers the case where agents at tier q are influenced by *all agents at higher tiers of the hierarchy*. But it also covers cases where agents are only influenced by a subset of agents at higher tiers of the hierarchy, as long as the influence relation is equal to its transitive closure: namely, for $i \in A_q, j \in A_p, k \in A_m, g_{ik} = 1$ if and only if $g_{ij} = 1$ and $g_{jk} = 1$. Figure 1.4.2 illustrates a hierarchy with nested neighborhoods and three tiers.



Proposition 1.8 Consider a hierarchy with nested neighborhoods such that all agents are connected to more than one root, and $|A_m| \leq |A_{m+1}|$ for all $m = 1, \dots, M - 1$. Then all agents receive the same quantity in the optimal contract satisfying group incentive compatibility with side-payments.

Proof: Let i be any agent at a tier $A_m, m \geq 2$ and j be a root agent. For the contract to be immune to a deviation by the pair (i, j) , we must have

$$\sum_{k \neq j} g_{ik} x_j x_k \leq \sum_{k \neq j} g_{ik} x_i x_k,$$

Because every agent is connected to at least two root agents, $\sum_{k \neq j} g_{ik} \neq 0$, so that we must have

$$x_j \leq x_i.$$

Now, suppose by contradiction that $x_j < x_i$ and compute the effect of a small decrease in x_i couple with a small increase in x_j on the sum of utilities:

$$\sum \Delta U = \epsilon(x_i - x_j) - \epsilon^2 - \alpha\epsilon \sum_k (g_{ik} + g_{ki})x_k + \alpha\epsilon \sum_k (g_{jk} + g_{kj})x_k.$$

To complete the proof of the Proposition, we need to show $\sum_k g_{kj}x_k \geq \sum_k (g_{ik} + g_{ki})x_k$. Consider first agents $k \in A_q$ with $q > m$ so that $g_{ik} = 0$. By the definition of nested neighborhoods, if $g_{ki} = 1$ then $g_{kj} = 1$, so

$$\sum_{k \in A_q} g_{kj}x_k \geq \sum_{k \in A_q} g_{ki}x_i.$$

Consider next agents $k \in A_q$, $1 < q < m$. Suppose that $g_{ik} = 1$ and $g_{kj} = 0$. Then, by the definition of nested neighborhoods we must have $g_{ij} = 0$, contradicting the fact that i is connected to j . So we must have $g_{kj} = 1$ and hence again

$$\sum_{k \in A_q} g_{kj}x_k \geq \sum_{k \in A_q} g_{ki}x_i.$$

We finally consider $k \in A_1$, $k \neq j$ and $k \in A_m$, $k \neq i$. Recall that if $k \in A_1$ and $k' \in A_m$, $x_k \leq x_{k'}$. Furthermore, by assumption $|A_m| \geq |A_1|$, so that

$$\sum_{k' \in A_m} g_{k'j}x_{k'} \leq \sum_{k \in A_1} g_{ik}x_k,$$

completing the proof of the Proposition. \square

Proposition 1.8 displays a family of hierarchies with multiple roots for which the group incentive compatibility constraint forces the planner to select *uniform* quantities in the network. The planner is unable to target agents at different locations, as agents always have an incentive to exchange their locations with one of the root agents. This may result in vary large efficiency losses for the planner, as the optimal contract clearly discriminates among agents, offering higher quantities to agents at higher tiers in the hierarchy. Observe that Proposition 1.8 relies on the fact that the number of agents is nondecreasing at lower levels of the hierarchy. As the following example shows, if this condition fails, the optimal contract may very well be group incentive compatible with side-payments.

Example 1 Suppose that $n = 4$, agents 1, 2, 3 belong to the first level of the hierarchy and a single agent 4 to the second level A_2 , i.e. $g_{41} = g_{42} = g_{43} = 1$ and $g_{ij} = 0$ for all other ij . Then in the first-best contract we have

$$\begin{aligned}x_1^* &= a + \alpha x_4^*, \\x_4^* &= a + 3\alpha x_1^*,\end{aligned}$$

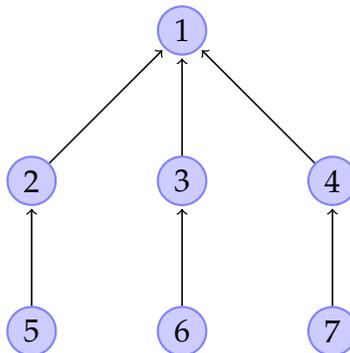
so that

$$x_1^* = x_2^* = x_3^* = \frac{a(1 + \alpha)}{1 - 3\alpha^2} < x_4^* = \frac{a(1 + 3\alpha)}{1 - 3\alpha^2}.$$

The optimal contract assigns a higher quantity to the agent at the lower level A_2 , so that the group incentive compatibility with side-payments constraint holds.

1.4.3. Regular Oriented Trees

We next consider hierarchies with a single root, but where agents are not all connected to the root. Instead, the hierarchy is a tree oriented towards the root, so that any node i has a *single* neighbor in the directed network, located at the tier immediately above him. We let σ_i denote the set of immediate predecessors of agent i in the oriented tree. We suppose that at any tier A_m of the hierarchy, $|\sigma_i| = |\sigma_j|$ for all $i, j \in A_m$. Hence the oriented tree is *regular*: there are the same number of predecessor nodes for any node at the same tier of the tree. Figure 1.4.3 illustrates a regular tree oriented towards the root with $n = 7$ agents. In this example $\sigma_1 = \{2, 3, 4\}$, $\sigma_2 = \{5\}$, $\sigma_3 = \{6\}$, $\sigma_4 = \{7\}$, $\sigma_5 = \sigma_6 = \sigma_7 = \emptyset$.



The following Proposition shows that, when the number of predecessors is higher for higher tiers in the oriented tree, the optimal contract is immune to group deviations with side-payments.

Proposition 1.9 *Consider a regular oriented tree such that, $|\sigma_i| < |\sigma_j|$ if $i \in A_m, j \in A_q$ and $m < q$. Then the optimal contract \mathbf{x}^* is group incentive compatible with side-payments.*

Proof. see Appendix 1.7.4. ■

Proposition 1.9 shows that in a regular oriented tree, the optimal contract can be sustained by the planner. When every agent is influenced by a single predecessor, the contract is immune to exchange by two adjacent agents if and only if the predecessor has a greater quantity than the successor. The optimal contract \mathbf{x}^* assigns to every agent a quantity which is proportional to her Katz-Bonacich centrality measure in the undirected tree. When the tree is regular, the Katz-Bonacich centrality measure will be higher at higher tiers whenever the number of agents connected to another agent is lower at lower tiers of the hierarchy.

Notice that the assumption that the number of agents connected to another agent is lower at lower levels of the hierarchy is crucial for the proof of Proposition 1.9. Consider for example a line of five agents, $n = 5$, $g_{12} = g_{23} = g_{34} = g_{45} = 1$, $g_{ij} = 0$ for all other pairs (i, j) . It is easy to check that the Katz-Bonacich centrality measure is highest at the center of the line, and decreases towards the edges, so that $x_3^* > x_2^* = x_4^* > x_1^* = x_5^*$. But then the condition $x_4^* \geq x_3^*$ is violated, and the optimal contract is not immune to a deviation by the two players, 3 and 4.

1.4.4. Connected agents at the same tier of the hierarchy

Next consider an alternative formulation, where agents are influenced by other agents at the same level of the hierarchy, i. e. $g_{ij} = 1$ if $i, j \in A_m$. In that case, we do not know whether Proposition 1.6 holds, as the argument showing that the optimal contract is group incentive compatible relies on the fact that the agent at the highest level of the hierarchy in S^- does not have any connection to other agents in S^- . Proposition 1.7 may also be violated, as the first-best contract can assign a higher quantity to agents at lower levels of the hierarchy. For example, suppose that $n = 5$, $A_1 = \{1\}$, $A_2 = \{2\}$, $A_3 = \{3, 4, 5\}$. We have $g_{21} = g_{31} = g_{32} = g_{34} = g_{35} = g_{41} = g_{42} = g_{43} = g_{45} = g_{51} = g_{52} = g_{53} = g_{54} = 1$ and $g_{ij} = 0$ otherwise. The first-best contract solves

$$\begin{aligned}x_1^* &= a + \alpha x_2^* + 3\alpha x_3^*, \\x_2^* &= a + \alpha x_1^* + 3\alpha x_3^*, \\x_3^* &= a + \alpha x_1^* + \alpha x_2^* + 4\alpha x_3^*\end{aligned}$$

so that

$$x_1^* = x_2^* = \frac{a(1 - \alpha)}{1 - 5\alpha - 2\alpha^2} < x_3^* = x_4^* = x_5^* = \frac{a(1 + \alpha)}{1 - 5\alpha - 2\alpha^2},$$

and the first-best contract satisfies the group incentive compatibility with side-payments condition.

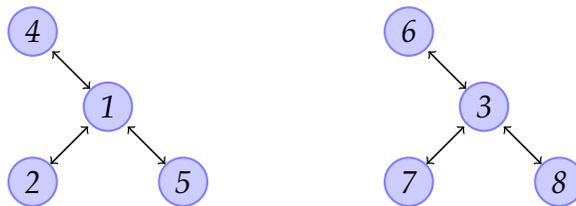
1.5. Robustness and Extensions

1.5.1. Nonadjacent agents

When the deviating coalitions comprise any set of players (not necessarily adjacent), the first-best contract may fail to be immune to group deviations with side-payments. We illustrate this fact by looking at the two families of network architectures for which the first-best contract has been shown to be immune to deviations by coalitions of adjacent agents: undirected networks and regular oriented trees.

Undirected networks

Example 2 Consider 8 agents organized in two disjoint stars as in the following picture



In a star with three peripheral nodes, the Katz-Bonacich centrality measure of the hub is given by $x^* = \frac{1+3\alpha}{1-3\alpha^2}$ and the Katz-Bonacich centrality of a peripheral node by $y^* = \frac{1+\alpha}{1-3\alpha^2}$ for $\alpha < \frac{1}{\sqrt{3}}$. Notice that $3y^* > x^*$. It is easy to check that no pair of agents has an incentive to

deviate: if the star and the hub exchange their quantities, the difference in utilities is given by

$$(y^* - x^*)3y^* + (x^* - y^*)x^* = (x^* - y^*)(x^* - 3y^*) < 0,$$

Suppose now that three agents deviate, the hubs of the two stars, agents 1 and 3 and a peripheral agent, agent 2. Furthermore, suppose that agents 2 and 3 exchange their quantities. The difference in utilities for the coalition of three agents is given by

$$2(x^* - y^*)x^* - 3y^*(x^* - y^*) = (x^* - y^*)(2x^* - 3y^*).$$

Now, whenever $\alpha > \frac{1}{3}$, this difference in utilities is positive. The loss in utilities of agent 3 (due to the decrease in its quantity from x^* to y^* is more than compensated by the gain of agents 1 and 2, who can choose a transfer to convince agent 3 to participate in the deviating coalition. Hence there exist intermediate values of α , $\frac{1}{3} < \alpha < \frac{1}{\sqrt{3}}$ for which the coalition of three non-adjacent agents has an incentive to deviate even though no two-player coalition has an incentive to deviate.

Regular oriented tree

Example 3 Consider the 7 agent network of Figure 1.4.3.

In the optimal contract of the planner

$$\begin{aligned} x_1^* &= a + 3\alpha x_2^*, \\ x_2^* &= a + \alpha x_1^* + \alpha x_5^*, \\ x_5^* &= a + \alpha x_2^* \end{aligned}$$

yielding

$$x_1^* = \frac{1 + 2\alpha^2 + 3\alpha}{1 - 4\alpha^2} > x_2^* = \frac{1}{1 - 4\alpha^2} > x_3^* = \frac{1 - \alpha}{1 - 4\alpha^2}.$$

Now consider a deviation by the non-adjacent pair (1, 5). Then the difference in utilities is given by

$$(x_1^* - x_5^*)x_2^* > 0,$$

so that the pair has an incentive to deviate at the first-best contract.

1.5.2. Richer information structures and mechanisms

Agents know the identity of their neighbors

Suppose that agent i knows the identity of the agents he is influenced by, i.e. agents for which $g_{ij} = 1$. Under this assumption, the planner can increase the set of possible messages, and ask agents to report the identity of their neighbors, consistent with the network architecture \hat{g} . We show that the planner can exploit the correlation between the announcements of agents to implement the first-best contract.

Proposition 1.10 *Suppose that all agents know the identity of their neighbors. Then there exist a mechanism such that the optimal contract \mathbf{x}^* is group incentive compatible with side-payments.*

Proof: Consider a contract where the planner assigns $\{x_1^*, \dots, x_n^*\}$ to agents in location $l = 1, \dots, n$ when announcements are consistent and $\{0, \dots, 0\}$ if announcements are inconsistent. Suppose that there exists a profitable deviation by some coalition S resulting in decisions $\mathbf{x} \neq \mathbf{x}^*$. Then clearly announcements must be consistent. By definition, the Because \mathbf{x}^* maximizes the sum of utilities of the agents, if $\sum_{i \in S} U_i(\mathbf{x}) > \sum_{i \in S} U_i(\mathbf{x}^*)$, there must exist an agent $j \notin S$ such that $U_j(\mathbf{x}) < U_j(\mathbf{x}^*)$. As announcements are consistent and j truthfully announces his location, he must receive x_j^* . Hence if $U_j(\mathbf{x}) < U_j(\mathbf{x}^*)$, $\sum_k g_{jk}x_k \neq \sum_k g_{jk}x_k^*$. This means that the set of agents who declare to be at locations k such that $g_{jk} = 1$ cannot be the true set of neighbors of j . But then, there must be an inconsistency in the announcements, contradicting the fact that $\mathbf{x} \neq \{0, \dots, 0\}$. \square

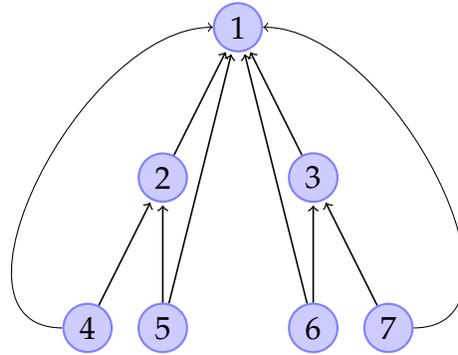
Proposition 1.10 shows that, if agents announce both their location and their list of neighbors, the planner can implement the optimal contract at no cost. This mechanism extracts complex information from the agents, checking the consistency of the list of neighbors and locations announced by each agent. If instead one were to restrict attention to *simple* contracts, which are only based on announcements of locations, then the analysis of the benchmark case would prevail. Hence we can reinterpret the benchmark case with finer information structure of the agents, as a restriction on the complexity of the contract. It assumes

that the contract is a simple sorting mechanism, where agents only announce locations, and obtain the payoff corresponding to their announced location if a simple consistency check (no two agents announce the same location) is satisfied.

1.5.3. Partial information

The planner knows the identity of the root

Consider a 7 agent network, with a three-tier hierarchy with a single root, 1, two agents in the second tier 2, 3 and four agents in the third tier, 4, 5, 6, 7.



In the first-best contract, we have

$$x_1^* = a + 2\alpha x_2^*,$$

$$x_2^* = a + \alpha x_1^* + 2\alpha x_4^*,$$

$$x_4^* = a + \alpha x_1^* + \alpha x_2^*$$

resulting in

$$x_1^* = \frac{1 + 2\alpha + 6\alpha^2}{1 - 4\alpha^2 + 4\alpha^3}, x_2^* = \frac{1 + 3\alpha + 2\alpha^2}{1 - 4\alpha^2 + 4\alpha^3} > x_3^* = \frac{1 + 2\alpha + \alpha^2}{1 - 4\alpha^2 + 4\alpha^3}.$$

By Proposition 1.7, if the planner ignores the identity of the root, she must set $x_1 = x_4 = x_5 = x_6$. Now suppose that the planner knows the identity of the root, so that the only possible deviating coalitions are coalitions of agents at tiers 2 and 3. The first-best contract assigns $x_4^* < x_2^*$, and the group incentive compatibility constraint writes:

$$(x_2^* - x_4^*)(x_1^* - x_1^*) \geq 0.$$

which is always satisfied. Hence the optimal contract satisfies the group incentive compatibility constraint with side-payments as soon as the planner knows the identity of the root.

1.5.4. Heterogenous marginal returns to own efforts

In this subsection, we relax the assumption about agents being homogenous regarding their marginal return to own effort or quantity consumed. That is there exists $i \neq j \in N$, such that $a_i \neq a_j$. Let 1 and 2 be two adjacent players at locations l_1 and l_2 . Let x_1 and x_2 be the contracts offered by the planner for respectively locations 1 and 2. Suppose that both agents have different marginal return to own action, that is $a_1 \neq a_2 > 0$. When agent 1 selects the contract intended for location l_2 , the difference in her payoff writes:

$$\begin{aligned} \Delta\pi_1 &= [a_1x_2 + \frac{1}{2}x_2^2 + \alpha \sum_{j \neq 2 \in N} g_{1j}x_jx_2 + \alpha x_2x_1] - [a_1x_1 + \frac{1}{2}x_1^2 + \alpha \sum_{j \neq 2 \in N} g_{1j}x_jx_1 + \alpha x_1x_2] \\ &= (x_2 - x_1)(a_1 + \frac{(x_1 + x_2)}{2}) + \alpha \sum_{j \neq 2 \in N} g_{1j}x_j \end{aligned}$$

Similarly, when agent 2 selects the contract intended for location l_1 while agent 1 selects the contract intended for location l_2 , her difference of payoffs is given by:

$$\Delta\pi_2 = (x_1 - x_2)(a_2 + \frac{(x_1 + x_2)}{2}) + \alpha \sum_{j \neq 1 \in N} g_{2j}x_j.$$

Hence,

$$\Delta\pi_1 + \Delta\pi_2 = (x_2 - x_1)([a_1 + \sum_{j \neq 2 \in N} g_{1j}x_j] - [a_2 + \alpha \sum_{j \neq 1 \in N} g_{2j}x_j]) \quad (1.3)$$

The optimal solution of the planner's problem given by Proposition 1.1 write with heterogenous marginal return to own action:

$$\mathbf{x}^* = (\mathbf{I} - \alpha(\mathbf{G} + \mathbf{G}^T))^{-1} a_d \mathbf{1},$$

where $a_d = \text{diag}(a_1, \dots, a_n)$ is a diagonal matrix with $a_i \neq a_j$ for at least one pair of connected agents $i \neq j \in N$. Proposition 1.5 still holds. It states that if the network g is undirected, then the optimal contract \mathbf{x}^* is group incentive compatible with side-payments. To see this, recall that for agents 1 and 2 to have a profitable deviation, it must be the case that :

$$(x_2 - x_1) \left([a_1 + \sum_{j \neq 2 \in N} g_{1j} x_j] - [a_2 + \alpha \sum_{j \neq 1 \in N} g_{2j} x_j] \right) > 0$$

We use the same argument as in the proof of Proposition 1.5. Recall that $\mathbf{x}^* = (\mathbf{I} - 2\alpha\mathbf{G})^{-1} a_d \mathbf{1}$ is simply the Katz-Bonacich measure and by construction of this measure it can be expressed recursively as a function of the Katz-Bonacich centrality of an agent's neighbors. Hence if, $x_2 > x_1$ then the above inequality cannot be strictly positive.

1.6. Conclusion

This paper studies whether a planner who only has information about the network topology can discriminate among agents according to their network position. The planner proposes a simple menu of contracts, one for each location, in order to maximize total welfare, and agents choose among the menu. This mechanism is immune to deviations by single agents, and to deviations by groups of agents of sizes 2, 3 and 4 if side-payments are ruled out. However, if compensations are allowed, groups of agents may have an incentive to jointly deviate from the optimal contract in order to exploit other agents. We identify network topologies for which the optimal contract is group incentive compatible with side-payments: undirected networks and regular oriented trees, and network topologies for which the planner must assign uniform quantities: single root and nested neighborhoods directed networks.

The analysis of this paper is a first step in the general study of the possible exploitation of anonymized social data by third parties. However, we realize that the setting we consider is very specific, and further research is needed to understand better how a planner can use partial knowledge of the social network to target agents. In particular, we would like to study how a planner can use simple instruments such as referral fees to extract information

about agents' local neighborhoods, without any information about the network topology.

1.7. Appendix

1.7.1. Proof of proposition 1.4

Fix a coalition S of adjacent players and consider the permutation ρ mapping i to $\rho(i)$ for any $i \in S$. (We dispense with the index S as the coalition is fixed). We first compute the difference of utility of any player i in S when players in the coalition use the permutation and when they report their true locations for any contract \mathbf{g} ,

$$\begin{aligned} \Delta U_i &= a(x_{\rho(i)} - x_i) - \frac{(x_i^2 - x_{\rho(i)}^2)}{2} + \alpha \left[\sum_{j \in S} g_{ij}(x_{\rho(i)}x_{\rho(j)} - x_ix_j) + \sum_{j \notin S} g_{ij}(x_{\rho(i)} - x_i)x_j \right], \\ &= (x_{\rho(i)} - x_i) \left[a - \frac{(x_i + x_{\rho(i)})}{2} + \alpha \sum_{j \notin S} g_{ij}x_j \right] + \alpha \sum_{j \in S} g_{ij}(x_{\rho(i)}x_{\rho(j)} - x_ix_j). \end{aligned}$$

At the optimal contract,

$$x_i^* = a + \alpha \sum_j [g_{ij} + g_{ji}]x_j^*.$$

So

$$\begin{aligned} x_i^* + x_{\rho(i)}^* &= 2a + \alpha \sum_j ([g_{ij} + g_{ji}]x_j^* + [g_{\rho(i)j} + g_{j\rho(i)}]x_j^*), \\ &= 2a + 2\alpha \sum_j [g_{ij} + g_{ji}]x_j^* + \alpha \left(\sum_j [g_{\rho(i)j} + g_{j\rho(i)}]x_j^* - \sum_j ([g_{ij} + g_{ji}]x_j^*) \right), \\ &= 2a + 2\alpha \sum_j [g_{ij} + g_{ji}]x_j^* + (x_{\rho(i)}^* - x_i) \end{aligned}$$

Replacing we obtain

$$\begin{aligned}
\Delta U_i &= \alpha \left[(x_{\rho(i)}^* - x_i^*) \left(- \sum_{j \notin S} g_{ji} x_j^* + \frac{(x_{\rho(i)}^* - x_i^*)}{2} \right) \right. \\
&\quad \left. + \sum_{j \in S} (g_{ij} x_{\rho(i)}^* x_{\rho(j)}^* - x_i^* x_j^* - x_j^* (x_{\rho(i)}^* - x_i^*)) - g_{ji} x_j^* (x_{\rho(i)}^* - x_i^*) \right] \\
&= \alpha \left[- \sum_{j \notin S} (x_{\rho(i)}^* - x_i^*) g_{ji} x_j^* - \frac{(x_{\rho(i)}^* - x_i^*)^2}{2} + \sum_{j \in S} g_{ij} x_{\rho(i)}^* (x_{\rho(j)}^* - x_j^*) - \sum_{j \in S} g_{ji} x_j^* (x_{\rho(i)}^* - x_i^*) \right].
\end{aligned}$$

Hence $\Delta U_i > 0$ if and only if

$$\sum_{j \in S} g_{ij} x_{\rho(i)}^* (x_{\rho(j)}^* - x_j^*) - \sum_{j \in S} g_{ji} x_j^* (x_{\rho(i)}^* - x_i^*) > \sum_{j \notin S} (x_{\rho(i)}^* - x_i^*) g_{ji} x_j^* + \frac{(x_{\rho(i)}^* - x_i^*)^2}{2}. \quad (1.4)$$

We now show that when $|S| \leq 3$, there must be some agent for whom inequality (1.4) fails. Suppose first that $|S| = 2$. Suppose without loss of generality that $x_2^* > x_1^*$ and consider ΔU_1 . Inequality (1.4) implies that

$$\Delta U_1 = (g_{12} - g_{21}) x_2^* (x_2^* - x_1^*) > 0,$$

a contradiction as $x_2^* > x_1^*$ and $g_{12} = g_{21}$.

Suppose next that $|S| = 3$ and suppose without loss of generality that $x_1^* < x_2^* < x_3^*$. Agent 1 must trade his quantity x_1^* for a higher quantity. Suppose first that $\rho(1) = 3, \rho(3) = 2, \rho(2) = 1$ so that agent 1 is the only agent who trades his quantity x_1^* for a higher quantity. Then inequality (1.4) yields

$$\Delta U_1 = g_{12} x_3^* (x_1^* - x_2^*) + g_{13} x_3^* (x_2^* - x_3^*) - x_2^* (x_2^* - x_1^*) - x_3^* (x_2^* - x_1^*) > 0,$$

a contradiction as $x_3^* > x_2^* > x_1^*$.

Suppose now that $\rho(1) = 2, \rho(2) = 3$ and $\rho(3) = 1$. (This is the only other permutation to consider, as any other permutation on S will leave one of the players at the same position.) Then inequalities (1.4) for agents 1 and 2 result in

$$\Delta U_1 = g_{12}x_2^*(x_3^* - x_2^*) + g_{13}x_2^*(x_1^* - x_3^*) - g_{21}x_2^*(x_2^* - x_1^*) - g_{31}x_3^*(x_2^* - x_1^*) > 0,$$

$$\Delta U_2 = g_{21}x_3^*(x_2^* - x_1^*) + g_{23}x_3^*(x_1^* - x_3^*) - g_{12}x_1^*(x_3^* - x_2^*) - g_{32}x_3^*(x_3^* - x_2^*) > 0.$$

First notice that in both inequalities, the only positive term is the first term, which implies that both g_{12} and g_{21} must be equal to 1 for the two inequalities to hold. Second, recall, that because the coalition is connected $g_{13} + g_{31} \geq 1$ and $g_{23} + g_{32} \geq 1$. Now, as $x_2^*(x_3^* - x_1^*) > x_2^*(x_2^* - x_1^*)$, the first inequality is more likely to be satisfied when $g_{13} = 0$ and $g_{31} = 1$. Similarly, because $x_3^*(x_3^* - x_1^*) > x_3^*(x_3^* - x_2^*)$, the second inequality is more likely to be satisfied when $g_{23} = 0$ and $g_{32}=1$. Hence if both inequalities are satisfied, we must have

$$x_2^*(x_3^* - x_2^*) > x_2^*(x_2^* - x_1^*) + x_3^*(x_2^* - x_1^*),$$

$$x_3^*(x_2^* - x_1^*) > x_1^*(x_3^* - x_2^*) + x_3^*(x_3^* - x_2^*)$$

But this implies

$$x_2^*(x_3^* - x_2^*) > x_2^*(x_2^* - x_1^*) + x_3^*(x_2^* - x_1^*) > x_2^*(x_2^* - x_1^*) + x_1^*(x_3^* - x_2^*) + x_3^*(x_3^* - x_2^*) > x_3^*(x_3^* - x_2^*),$$

a contradiction that completes the proof for $|S| = 3$. Now suppose that the size of the group is $|S| = 4$ and suppose without loss of generality that $x_4^* > x_3^* > x_2^* > x_1^*$. There are six possible permutations where each member of the group gets a different contract from the one intended for their location and no pair exchanges pairwise their contracts. There are 3^6 possible *ways* in which the members of the group can be linked because each link $(g_{ij}, g_{ji}) \in \{(0, 1), (1, 0), (1, 1)\}$.

Recall that :

$$\begin{aligned}\Delta U_i &= -\frac{\alpha}{2}(x_{\rho(i)} - x_i)^2 - \alpha(x_{\rho(i)} - x_i) \sum_{j \in S} g_{ji} x_j \\ &\quad - \alpha(x_{\rho(i)} - x_i) \sum_{j \notin S} g_{ji} x_j \\ &\quad + \alpha x_{\rho(i)} \sum_{j \in S} g_{ij} (x_{\rho(j)} - x_j)\end{aligned}$$

1. Consider: $\rho(1) = 4, \rho(2) = 1, \rho(3) = 2$ and $\rho(4) = 3$

Note that all agents get a lower quantity except agent 1. The difference in utility of agent 1 writes:

$$\begin{aligned}\Delta U_1 &= -\frac{\alpha}{2}(x_4^* - x_1^*)^2 - \alpha(x_4^* - x_1^*)[g_{21}x_2 + g_{31}x_3 + g_{41}x_4] - \alpha(x_4^* - x_1^*) \sum_{j \notin S} g_{j1}x_j \\ &\quad + \alpha x_{\rho(1)}[g_{12}(x_1^* - x_2^*) + g_{13}(x_1^* - x_3^*) + g_{14}(x_1^* - x_4^*)] < 0\end{aligned}$$

Recall that $x_1^* < x_2^* < x_3^* < x_4^*$, hence there exists at least one agent that is strictly worse off with ρ for all network structures i.e. for all $i \neq j \in \{1, 2, 3, 4\}$ and $g_{ij} \in \{0, 1\}$ such that $g_{ij} + g_{ji} \geq 1$.

2. Consider: $\rho(1) = 4, \rho(2) = 3, \rho(3) = 1$ and $\rho(4) = 2$.

For all the network structures such that $g_{12} = 0$ and the remaining links take the value 0 or 1 such that for all $i, j \in S$ $g_{ij} + g_{ji} \geq 1$, $\Delta U_1 < 0$. Similarly, for $g_{21} = 0$, $\Delta U_2 < 0$.

Moreover, when $g_{12} = g_{21} = 1$ and $g_{13} = 1$ and $g_{31} \in \{0, 1\}$: $\Delta U_1 < 0$. For $g_{12} = g_{21} = 1$ and $g_{14} = 1$ and $g_{41} \in \{0, 1\}$: $\Delta U_1 < 0$. The remaining networks that we need inspect are networks such that $g_{12} = g_{21} = 1$ and $g_{13} = 0, g_{31} = 1, g_{14} = 0, g_{41} = 1$. But again, when $g_{31} = 1$ and the remaining links take values 0 or 1 such that for $i, j \in S$, $g_{ij} + g_{ji} \geq 1$, we get $\Delta U_1 < 0$.

3. Consider: $\rho(1) = 3, \rho(2) = 4, \rho(3) = 2$ and $\rho(4) = 1$.

For all the network structures such that $g_{12} = 0$ and the remaining links take the value 0 or 1 such that for all $i, j \in S$ $g_{ij} + g_{ji} \geq 1$, $\Delta U_1 < 0$. Similarly, when $g_{21} = 0$, $\Delta U_2 < 0$. Hence we check structures where $g_{12} = g_{21} = 1$.

For $g_{12} = g_{21} = 1$ and $g_{24} = 1$ and $g_{42} \in \{0, 1\}$: $\Delta U_2 < 0$. Similarly, when $g_{12} = g_{21} = 1$ and $g_{14} = 1$ and $g_{41} \in \{0, 1\}$: $\Delta U_1 < 0$.

The remaining networks that we need to inspect are such that $g_{12} = g_{21} = 1$ and $g_{14} = 0$, $g_{41} = 1$, $g_{24} = 0$, $g_{42} = 1$ and the remaining links take the value 0 or 1, and for all $i, j \in S$ $g_{ij} + g_{ji} \geq 1$. Consider ΔU_2 :

$$\begin{aligned} \Delta U_2 &= -\frac{\alpha}{2}(x_4^* - x_2^*)^2 - \alpha(x_4^* - x_2^*) \sum_{j \notin S} g_{j2} x_j - \alpha(x_4^* - x_2^*)(x_1^* + g_{32} x_3^* + x_4^*) \\ &\quad + \alpha x_4^*((x_3^* - x_1^*) + g_{23}(x_2^* - x_3^*)) \end{aligned}$$

If $(x_3^* - x_1^*) < (x_4^* - x_2^*)$ then $\Delta U_2 < 0$. If not, consider ΔU_1 :

$$\begin{aligned} \Delta U_1 &= -\frac{\alpha}{2}(x_3^* - x_1^*)^2 - \alpha(x_3^* - x_1^*) \sum_{j \notin S} g_{j1} x_j - \alpha(x_3^* - x_1^*)(x_2^* + g_{31} x_3^* + x_4^*) \\ &\quad + \alpha x_3^*((x_4^* - x_2^*) + g_{13}(x_2^* - x_3^*)) \end{aligned}$$

If $(x_3^* - x_1^*) \geq (x_4^* - x_2^*)$ then $\Delta U_1 < 0$. A contradiction which completes the proof for this permutation ρ .

4. Consider: $\rho(\mathbf{1}) = \mathbf{3}$, $\rho(\mathbf{2}) = \mathbf{1}$, $\rho(\mathbf{3}) = \mathbf{4}$ and $\rho(\mathbf{4}) = \mathbf{2}$.

For all the network structures such that $g_{31} = 0$ and the remaining links take the value 0 or 1 such that for all $i, j \in S$ $g_{ij} + g_{ji} \geq 1$, $\Delta U_3 < 0$. Similarly, when $g_{13} = 0$, $\Delta U_1 < 0$. Moreover for $g_{13} = g_{31} = 1$ and $g_{14} = 1$ and $g_{41} \in \{0, 1\}$, it follows that $\Delta U_1 < 0$.

The remaining networks to inspect are ones where $g_{13} = g_{31} = 1$ and $g_{14} = 0$ and $g_{41} = 1$. Let $d_1 = x_3^* - x_1^*$ and $d_2 = x_4^* - x_3^*$. The variation in payoffs of agents 1 and 3

writes:

$$\begin{aligned}\Delta U_1 &= -\alpha d_1 \left[\frac{d_1}{2} + (g_{21}x_2^* + x_3^* + x_4^*) + \sum_{j \notin S} g_{j1}x_j \right] - \alpha x_3^* g_{12}(x_2^* - x_1^*) + \alpha d_2 x_3^* \\ \Delta U_3 &= -\alpha d_2 \left[\frac{d_2}{2} + (g_{23}x_2^* + x_1^* + g_{43}x_4^*) + \sum_{j \notin S} g_{j3}x_j \right] - \alpha x_4^* [g_{32}(x_2^* - x_1^*) + g_{34}(x_4^* - x_2^*)] \\ &\quad + \alpha d_1 x_4^*\end{aligned}$$

If $d_1 \geq d_2$ then $\Delta U_1 < 0$ because the positive term $\alpha d_2 x_3^*$ is strictly smaller than $-\alpha d_1 x_4^*$. We study below the remaining case: $d_2 > d_1$. Suppose by contradiction that $\Delta U_3 > 0$ and recall that $g_{34} + g_{43} \geq 1$ because players 3 and 4 are part of the same group S . For $g_{43} = 1$ and $g_{34} \in \{0, 1\}$:

$$\begin{aligned}\Delta U_3 > 0 &\Leftrightarrow \alpha d_1 x_4^* > \alpha d_2 \left[\frac{d_2}{2} + (g_{23}x_2^* + x_1^* + x_4^*) + \sum_{j \notin S} g_{j3}x_j \right] + \alpha x_4^* [g_{32}(x_2^* - x_1^*) \\ &\quad + g_{34}(x_4^* - x_2^*)] (> 0) \\ &\Leftrightarrow \alpha x_4^* (d_1 - d_2) > \alpha d_2 \left[\frac{d_2}{2} + (g_{23}x_2^* + x_1^*) + \sum_{j \notin S} g_{j3}x_j \right] + \alpha x_4^* [g_{32}(x_2^* - x_1^*) \\ &\quad + g_{34}(x_4^* - x_2^*)] (> 0)\end{aligned}$$

A contradiction because $\alpha x_4^* (d_1 - d_2) < 0$. Hence $\Delta U_3 < 0$. Recall that $\alpha d_1 x_4^*$ is upper bounded by $\alpha d_2 x_4^*$ because we are considering the case $d_2 > d_1$. For $g_{34} = 1$ and $g_{43} \in \{0, 1\}$:

$$\begin{aligned}\alpha d_2 x_4^* &> (\alpha d_1 x_4^* >) \alpha d_2 \left[\frac{d_2}{2} + (g_{23}x_2^* + x_1^* + g_{43}x_4^*) + \sum_{j \notin S} g_{j3}x_j \right] \\ &\quad + \alpha x_4^* [g_{32}(x_2^* - x_1^*) + (x_4^* - x_2^*)] (> 0) \\ &\Leftrightarrow \alpha d_2 x_4^* - \alpha x_4^* (x_4^* - x_2^*) > \alpha d_2 \left[\frac{d_2}{2} + (g_{23}x_2^* + x_1^*) + \sum_{j \notin S} g_{j3}x_j \right] + \alpha x_4^* g_{32}(x_2^* - x_1^*) (> 0) \\ &\Leftrightarrow \alpha x_4^* (x_4^* - x_3^* - x_4^* + x_2^*) > \alpha d_2 \left[\frac{d_2}{2} + (g_{23}x_2^* + x_1^*) + \sum_{j \notin S} g_{j3}x_j \right] + \alpha x_4^* g_{32}(x_2^* - x_1^*) (> 0)\end{aligned}$$

A contradiction because $x_1^* < x_2^* < x_3^* < x_4^*$ implies $\alpha x_4^* (x_4^* - x_3^* - x_4^* + x_2^*) < 0$ and the *RHS* is strictly positive. Hence $\Delta U_3 < 0$, which completes the proof for this permutation.

5. Consider: $\rho(1) = 2, \rho(2) = 4, \rho(3) = 1$ and $\rho(4) = 3$.

For all the network structures such that $g_{21} = 0$ (link with the other agent who gets a higher quantity) and the remaining links take the value 0 or 1 such that for all $i, j \in S$ $g_{ij} + g_{ji} \geq 1, \Delta U_2 < 0$. Similarly, for $g_{12} = 0, \Delta U_1 < 0$. Moreover, for $g_{12} = g_{21} = 1$ and $g_{23} = 1$ and $g_{32} \in \{0, 1\}$: $\Delta U_2 < 0$. We inspect the remaining group structures for which we did not show a contradiction: $g_{12} = g_{21} = 1$ and $g_{23} = 0$ and $g_{32} = 1$. Following the same reasoning in the previous case, let $d_3 = x_2^* - x_1^*$ and $d_4 = x_4^* - x_2^*$. If $d_3 \geq d_4 = 0$ then $\Delta U_1 < 0$. If $d_4 > d_3$ then $\Delta U_3 < 0$ (consider the links $g_{42} + g_{24} \geq 1$).

6. Consider: $\rho(1) = 2, \rho(2) = 3, \rho(3) = 4$ and $\rho(4) = 1$

Let $d_1 = x_2^* - x_1^*, d_2 = x_3^* - x_2^*, d_3 = x_4^* - x_3^*$ and $d_4 = x_4^* - x_1^*$ (clearly $d_4 = d_1 + d_2 + d_3$). The variation in payoffs of agents 1, 2, 3 write:

$$\Delta U_1 = -\alpha d_1 \left[\frac{d_1}{2} + g_{21}x_2^* + g_{31}x_3^* + g_{41}x_4^* + \sum_{j \in S} g_{j1}x_j \right] + \alpha x_2^* [g_{12}d_2 + g_{13}d_3 - g_{14}d_4]$$

If $g_{14} = 1, g_{12} \in \{0, 1\}$ and $g_{13} \in \{0, 1\}$ then $\Delta U_1 < 0$ because $d_4 > d_2 + d_3$. If $g_{12} = g_{13} = 0$ and $g_{14} \in \{0, 1\}$ then $\Delta U_1 < 0$. Hence we consider networks where $g_{14} = 0$ and $g_{12} + g_{13} \geq 1$.

$$\Delta U_2 = -\alpha d_2 \left[\frac{d_2}{2} + g_{12}x_1^* + g_{32}x_3^* + g_{42}x_4^* + \sum_{j \in S} g_{j2}x_j \right] + \alpha x_3^* [g_{21}d_1 + g_{23}d_3 - g_{24}d_4]$$

If $g_{24} = 1, g_{21} \in \{0, 1\}$ and $g_{23} \in \{0, 1\}$ then $\Delta U_2 < 0$. If $g_{21} = g_{23} = 0$ and $g_{24} \in \{0, 1\}$ then $\Delta U_2 < 0$. Hence we consider networks where $g_{24} = 0$ and $g_{21} + g_{23} \geq 1$.

$$\Delta U_3 = -\alpha d_3 \left[\frac{d_3}{2} + g_{13}x_1^* + g_{23}x_2^* + g_{43}x_4^* + \sum_{j \in S} g_{j3}x_j \right] + \alpha x_4^* [g_{31}d_1 + g_{32}d_2 - g_{34}d_4]$$

If $g_{34} = 1, g_{31} \in \{0, 1\}$ and $g_{32} \in \{0, 1\}$ then $\Delta U_3 < 0$. If $g_{31} = g_{32} = 0$ and $g_{34} \in \{0, 1\}$ then $\Delta U_3 < 0$. Hence we consider networks where $g_{34} = 0$ and $g_{31} + g_{32} \geq 1$.

Now we consider the remaining group structures such that: $g_{14} = g_{24} = g_{34} = 0$ (and $g_{41} = g_{42} = g_{43} = 1$), $g_{12} + g_{13} \geq 1, g_{21} + g_{23} \geq 1$ and $g_{31} + g_{32} \geq 1$. In addition to the

three previous inequalities, $g_{ij} + g_{ji} \geq 1$ for $i \neq j \in S$. This leaves us with exactly 10 group structures to inspect. The variation in payoffs rewrite:

$$\Delta U_1 = -\alpha d_1 \left[\frac{d_1}{2} + g_{21}x_2^* + g_{31}x_3^* + \sum_{j \notin S} g_{j1}x_j \right] - \alpha d_1 x_4^* + \alpha x_2^* [g_{12}d_2 + g_{13}d_3]$$

$$\Delta U_2 = -\alpha d_2 \left[\frac{d_2}{2} + g_{12}x_1^* + g_{32}x_3^* + \sum_{j \notin S} g_{j2}x_j \right] - \alpha d_2 x_4^* + \alpha x_3^* [g_{21}d_1 + g_{23}d_3]$$

$$\Delta U_3 = -\alpha d_3 \left[\frac{d_3}{2} + g_{13}x_1^* + g_{23}x_2^* + \sum_{j \notin S} g_{j3}x_j \right] - \alpha d_3 x_4^* + \alpha x_4^* [g_{31}d_1 + g_{32}d_2]$$

- (a) $g_{12} = g_{31} = g_{23} = 1$ and $g_{13} = g_{32} = g_{21} = 0$: if $d_3 > d_1$ then $\Delta U_3 < 0$. If $d_3 < d_1$ and $d_2 > d_3$ then $\Delta U_2 < 0$. If $d_3 < d_1$ and $d_3 > d_2$ then it implies $d_1 > d_2$ and $\Delta U_1 < 0$.
- (b) $g_{12} = g_{31} = g_{23} = 0$ and $g_{13} = g_{32} = g_{21} = 1$: if $d_3 > d_2$ then $\Delta U_3 < 0$. If $d_2 < d_3$ and $d_2 > d_1$, $\Delta U_2 < 0$. If $d_2 > d_3$ and $d_1 > d_2$ then it implies $d_1 > d_3$, $\Delta U_1 < 0$.
- (c) $g_{12} = g_{31} = g_{32} = g_{23} = g_{21} = 1$ and $g_{13} = 0$: if $d_1 > d_2$, $\Delta U_1 < 0$. If $d_2 > d_1$, $\Delta U_2 < 0$.
- (d) $g_{13} = g_{31} = g_{23} = g_{21} = 1$ and $g_{12} = g_{32} = 0$: if $d_1 > d_3$, $\Delta U_1 < 0$. If $d_3 > d_1$, $\Delta U_3 < 0$.
- (e) $g_{12} = g_{13} = g_{32} = g_{23} = 1$ and $g_{31} = g_{21} = 0$: if $d_2 > d_3$, $\Delta U_2 < 0$. If $d_3 > d_2$, $\Delta U_3 < 0$.
- (f) $g_{12} = g_{13} = g_{32} = g_{23} = g_{21} = 1$ and $g_{31} = 0$: if $d_1 > d_2$ and $d_1 > d_3$ then $\Delta U_1 < 0$. If $d_1 > d_2$ and $d_3 > d_1$ then it implies $d_3 > d_2$ and $\Delta U_3 < 0$. If $d_2 > d_1$ and $d_1 > d_3$ then it implies $d_2 > d_3$ and $\Delta U_2 < 0$. If $d_2 > d_1$ and $d_3 > d_1$ then $\Delta U_3 < 0$.
- (g) $g_{12} = g_{13} = g_{31} = g_{32} = g_{21} = 1$ and $g_{23} = 0$. If $d_2 > d_1$ then $\Delta U_2 < 0$. If $d_1 > d_2$ and $d_1 > d_3$ then $\Delta U_3 < 0$. If $d_1 > d_2$ and $d_3 > d_1$, it implies $d_3 > d_1 > d_2$. If $d_3 > d_1 > d_2$ and $d_3 > d_1 + d_2$ then $\Delta U_3 < 0$. Last case: $d_3 > d_1 > d_2$ and $d_3 < d_1 + d_2$. We

rewrite ΔU_1 and find a (negative) upper bound using $d_3 < d_1 + d_2 \Leftrightarrow d_3 - d_1 < d_2$:

$$\begin{aligned}\Delta U_1 &= -\alpha d_1 \left(\frac{d_1}{2} + \sum_{j \notin S} g_{j1} x_j \right) - \alpha d_1 x_3^* + \alpha d_2 x_2^* - \alpha d_1 x_4^* + \alpha x_2^* (d_3 - d_1) \\ &< -\alpha d_1 \left(\frac{d_1}{2} + \sum_{j \notin S} g_{j1} x_j \right) - \alpha d_1 x_3^* + \alpha d_2 x_2^* - \alpha d_1 x_4^* + \alpha x_2^* d_2 \\ &< 0 \text{ because } d_1 > d_2\end{aligned}$$

(h) $g_{12} = g_{13} = g_{23} = g_{32} = g_{21} = 1$ and $g_{31} = 0$. If $d_1 > d_2$ and $d_1 > d_3$ then $\Delta U_1 < 0$. If $d_1 > d_2$ and $d_3 > d_1$ then $d_3 > d_2$ and $\Delta U_3 < 0$. If $d_2 > d_1$ and $d_1 > d_3$ then $d_2 > d_3$ and $\Delta U_2 < 0$. Last two cases: if $d_2 > d_1$ and $d_3 > d_1$ and $d_2 > d_3$. We rewrite ΔU_2 :

$$\Delta U_2 = -\alpha d_2 \left(\frac{d_2}{2} + \sum_{j \notin S} g_{j2} x_j \right) - d_2 x_1^* - x_3^* (d_2 - d_3) - d_2 x_4^* + d_1 x_3^* < 0$$

If $d_2 > d_1$ and $d_3 > d_1$ and $d_3 > d_2$, then $\Delta U_3 < 0$ (because here $g_{32} = 0$).

(i) $g_{31} = g_{12} = g_{21} = g_{23} = 1$ and $g_{32} = g_{13} = 0$. If $d_3 > d_1$ then $\Delta U_3 < 0$. If $d_1 > d_3$ and $d_1 > d_2$ then $\Delta U_1 < 0$. Last two cases: if $d_2 > d_1 > d_3$ and $d_2 > d_1 + d_3$ then $\Delta U_2 < 0$. If $d_2 > d_1 > d_3$ and $d_2 < d_1 + d_3$, we rewrite ΔU_1 :

$$\begin{aligned}\Delta U_1 &= -\alpha d_1 \left[\frac{d_1}{2} + \sum_{j \notin S} g_{j1} x_j \right] - \alpha d_1 (x_2^* + x_3^* + x_4^*) + \alpha d_2 x_2^* \\ &< -\alpha d_1 \left[\frac{d_1}{2} + \sum_{j \notin S} g_{j1} x_j \right] - \alpha d_1 (x_2^* + x_3^* + x_2^*) + \alpha d_2 x_2^* \text{ (because } -x_4^* < -x_2^*) \\ &= -\alpha d_1 \left[\frac{d_1}{2} + \sum_{j \notin S} g_{j1} x_j \right] - \alpha d_1 x_3^* - 2\alpha d_1 x_2^* + \alpha x_2^* d_2 \\ &< 0 \text{ because } (d_2 < d_1 + d_3 < 2d_1)\end{aligned}$$

(j) $g_{31} = g_{13} = g_{32} = g_{23} = g_{21} = 1$ and $g_{12} = 0$. If $d_1 > d_3$ then $\Delta U_1 < 0$. If $d_3 > d_1$ and $d_2 > d_3$ then $\Delta U_2 < 0$. Last two cases: if $d_3 > d_2 > d_1$ and $d_3 > d_1 + d_2$ then

$\Delta U_3 < 0$. If $d_3 > d_1 > d_2$ and $d_3 < d_1 + d_2$, we rewrite ΔU_1 :

$$\begin{aligned} \Delta U_1 &= -\alpha d_1 \left(\frac{d_1}{2} + \sum_{j \notin S} g_{j1} x_j \right) - \alpha d_1 (x_2^* + x_3^* + x_4^*) + \alpha d_3 x_2^* \\ &< -\alpha d_1 \left(\frac{d_1}{2} + \sum_{j \notin S} g_{j1} x_j \right) - \alpha d_1 (x_2^* + x_3^* + x_2^*) + \alpha d_3 x_2^* \text{ because } -x_4^* < -x_2^* \\ &= -\alpha d_1 \left(\frac{d_1}{2} + \sum_{j \notin S} g_{j1} x_j \right) - \alpha d_1 x_3^* - 2\alpha d_1 x_2^* + \alpha d_3 x_2^* \\ &< 0 \text{ because } d_3 < d_1 + d_2 < 2d_1 \end{aligned}$$

a contradiction that completes the proof of the Proposition.

1.7.2. Proof of proposition 1.5

We first show that there does not exist a pair of connected agents (i, j) who can benefit from a deviation. Suppose that this were the case, then we would have

$$x_j^* \sum_{k \neq j \in N} g_{ik} x_k^* + x_i^* \sum_{k \neq i \in N} g_{jk} x_k^* > x_i^* \sum_{k \neq j \in N} g_{ik} x_k^* + x_j^* \sum_{k \neq i \in N} g_{jk} x_k^*,$$

As $g_{ij} = g_{ji} = 1$, we can add $2x_i^* x_j^*$ on both sides of the inequality, to obtain

$$x_j^* \sum_{k \in N} g_{ik} x_k^* + x_i^* \sum_{k \in N} g_{jk} x_k^* > x_i^* \sum_{k \in N} g_{ik} x_k^* + x_j^* \sum_{k \in N} g_{jk} x_k^*,$$

and collecting terms

$$(x_j^* - x_i^*) \left(\sum_{k \in N} g_{ik} x_k^* - \sum_{k \in N} g_{jk} x_k^* \right) > 0. \quad (1.5)$$

When the network is undirected we have $G + G^T = 2G$. Now recall that $\mathbf{x}^* = (\mathbf{I} - 2\alpha \mathbf{G})^{-1} a \mathbf{1}$, the Katz-Bonacich centrality measure vector. By construction, the Katz-Bonacich centrality of an agent i can be recursively expressed as a function of the Katz-Bonacich centrality of agent i 's neighbors:

$$b_i(\mathbf{G}, 2\alpha) = 1 + 2\alpha \sum_{k \in N} g_{ik} b_k(\mathbf{G}, 2\alpha).$$

But then, if $b_j(\mathbf{G}, 2\alpha) > b_i(\mathbf{G}, 2\alpha)$, we must have $\sum_{k \in N} g_{jk} b_k(\mathbf{G}, 2\alpha) > \sum_{k \in N} g_{ik} b_k(\mathbf{G}, 2\alpha)$,

contradicting equation (1.5).

In the second step of the proof, we show that there cannot exist a coalition of adjacent players of size $k > 2$ which has a profitable deviation. Suppose that it were the case and let ρ denote the permutation of players in S which results in a profitable deviation. We must then have

$$\sum_{i \in S} \sum_{j \notin S} g_{ij} x_{\rho_S(i)} x_j > \sum_{i \in S} \sum_{j \notin S} g_{ij} x_i x_j.$$

Because $g_{ij} = 1$ for all $i, j \in S$, we can add to the left and right sides of the inequality $2 \sum_{i, j \in S} x_i x_j = 2 \sum_{i, j \in S} x_{\rho_S(i)} x_{\rho_S(j)}$ to obtain

$$\sum_{i \in S} \sum_{j \neq i} g_{ij} x_{\rho_S(i)} x_j > \sum_{i \in S} \sum_{j \neq i} g_{ij} x_i x_j. \quad (1.6)$$

Next order the vector of \mathbf{x} so that $x_i \leq x_{i+1}$ for all $i \in S$ and let all indices outside S be larger than the indices in S . Then we can rewrite equation (1.6) as

$$(x_{\rho_S(1)} - x_1) \sum_j g_{1j} x_j + (x_{\rho_S(2)} - x_2) \sum_j g_{2j} x_j + \dots + (x_{\rho_S(s)} - x_s) \sum_j g_{sj} x_j > 0. \quad (1.7)$$

Next consider the set of agents in S such that $\rho(i) \neq i$, discard all other agents in S and reorder if needed the indices so that all remaining agents in S have lower indices. Let t be the size of the remaining set of indices. For any $j \in \{1, \dots, t\}$ define the following two sets of indices:

$$\begin{aligned} A_j &= \{i | \rho(i) \geq j + 1 > j \geq i\}, \\ B_j &= \{i | i \geq j + 1 > j \geq \rho(i)\}. \end{aligned}$$

Clearly $A_j \cap B_j = \emptyset$. We now show by induction that $|A_j| = |B_j|$ and that $\max\{i | i \in A_j\} < \min\{i | i \in B_j\}$ for all j .

Consider the initial step at $j = 1$. Then clearly $\rho(1) > 1$ and we have $A_1 = \{1\}$. Let k be the unique antecedent of 1, i.e. $\rho(k) = 1$. Then $B_1 = \{k\}$ and hence $|A_1| = |B_1| = 1$ and $k > 1$.

Next assume that $|A_{j-1}| = |B_{j-1}|$ and $\max\{i | i \in A_{j-1}\} < \min\{i | i \in B_{j-1}\}$. Consider the

sets A_j and B_j . By construction we have

$$\begin{aligned} A_j &= A_{j-1} \setminus \{i | \rho(i) = j > j-1 \geq i\} \cup \{i | \rho(i) \geq j+1 > j = i\} \\ B_j &= B_{j-1} \setminus \{i | i = j > j-1 \geq \rho(i)\} \cup \{i | i \geq j+1 > j = \rho(i)\} \end{aligned}$$

Now let k be the antecedent of j in the permutation ρ , i.e. $\rho(k) = j$. We consider the four following cases:

Case 1. $\rho(j) > j, j > k$ In that case $A_j = A_{j-1} \setminus \{k\} \cup \{j\}$, $B_j = B_{j-1}$ and hence $|A_j| = |B_j|$. Furthermore, $j > i$ for all $i \in A_{j-1}$ so $j = \max\{i | i \in A_j\}$. Finally, $i \geq j$ for all $j \in B_{j-1}$ and as $j \notin B_j$, $\min\{i | i \in B_j\} > j = \max\{i | i \in A_j\}$.

Case 2. $\rho(j) > j, k > j$ In that case $A_j = A_{j-1} \cup \{j\}$, $B_j = B_{j-1} \cup \{k\}$ and hence $|A_j| = |B_j|$. As $k > j$, if $\max\{i | i \in A_{j-1}\} < \min\{i | i \in B_{j-1}\}$, we must also have $\min\{i | i \in B_j\} > j = \max\{i | i \in A_j\}$.

Case 3. $\rho(j) < j, j > k$ In that case $A_j = A_{j-1} \setminus \{k\}$, $B_j = B_{j-1} \setminus \{j\}$ and hence $|A_j| = |B_j|$. Furthermore, as $A_j \subset A_{j-1}$, $\max\{i | i \in A_j\} \leq \max\{i | i \in A_{j-1}\}$ and as $B_j \subset B_{j-1}$, then $\min\{i | i \in B_j\} \geq \min\{i | i \in B_{j-1}\}$. Hence $\min\{i | i \in B_j\} \geq \max\{i | i \in A_j\}$.

Case 4. $\rho(j) < j, k > j$ In that case $A_j = A_{j-1}$ and $B_j = B_{j-1} \setminus \{j\} \cup \{k\}$ and hence $|A_j| = |B_j|$. In addition as $k > j$, $\min\{i | i \in B_j\} \geq \min\{i | i \in B_{j-1}\} > \max\{i | i \in A_{j-1}\} = \max\{i | i \in A_j\}$.

Now rewrite equation (1.7) as

$$\sum_{j=1}^{t-1} (x_{j+1} - x_j) \left(\sum_{i \in A_j} \sum_k g_{ik} x_k - \sum_{i \in B_j} \sum_k g_{ik} x_k \right) > 0. \quad (1.8)$$

Now because $|A_j| = |B_j|$, we can pick for any $i \in A_j$ a corresponding index l in B_j and as $l > i$, we also have $x_l > x_i$. But by equation (1.5), if (i, l) does not have a profitable deviation and $x_l > x_i$ we must have

$$\sum_k g_{lk} x_k \geq \sum_k g_{ik} x_k.$$

But this implies that for all j

$$\left(\sum_{i \in A_j} \sum_k g_{ik} x_k - \sum_{i \in B_j} \sum_k g_{ik} x_k \right) \leq 0,$$

contradicting equation (1.8). This contradiction completes the proof of the Proposition.

1.7.3. Proof of proposition 1.7

Consider a two-player deviation with side-payments between any agent i in A_m , $m > 2$ and the root agent j in A_1 . We must have

$$\sum_{k \neq j} g_{ik} x_j x_k + \sum_{k \neq i} x_i x_k \leq \sum_{k \neq j} g_{ik} x_i x_k + \sum_{k \neq i} g_{jk} x_j x_k,$$

Next, as $g_{jk} = 0$ for all k , this condition amounts to

$$\sum_{k \neq j} g_{ik} x_j x_k \leq \sum_{k \neq j} g_{ik} x_i x_k,$$

or $x_j \leq x_i$. Notice that this condition must hold for any pair (i, j) where j is the root agent and $i \in A_k$, $k > 2$.

Suppose that the planner selects a contract \mathbf{x} such that $x_j < x_i$ for some $i \in A_m$ for $m > 2$. Consider the marginal effect of increasing x_j by ϵ and reducing x_i by ϵ on the sum of payoffs:

$$\begin{aligned} \sum \Delta U &= -\frac{(x_i - \epsilon)^2 - x_i^2}{2} - \frac{(x_j + \epsilon)^2 - x_j^2}{2} - \alpha \epsilon \sum_k (g_{ik} + g_{ki}) x_k + \alpha \epsilon \sum_k (g_{jk} + g_{kj}) x_k \\ &= \epsilon(x_i - x_j) - \epsilon^2 - \alpha \epsilon \sum_k (g_{ik} + g_{ki}) x_k + \alpha \epsilon \sum_k (g_{jk} + g_{kj}) x_k. \end{aligned}$$

Now recall that in a hierarchical network, $g_{ik} + g_{jk} \leq 1$. Because player j is connected to all agents, $g_{jk} + g_{kj} = 1$ for all $k \neq j$. Player i is at best connected to all agents so $g_{ik} + g_{ki} \leq 1$ for all $k \neq i$. Now

$$\begin{aligned} \sum_k (g_{jk} + g_{kj}) x_k - \sum_k (g_{ik} + g_{ki}) x_k &= (x_i - x_j) + \sum_{k \neq i, j} (1 - (g_{ik} + g_{ki})) x_k \\ &\geq (x_i - x_j) \\ &> 0 \end{aligned}$$

so that

$$\sum \Delta U > 0$$

for small enough ϵ . This shows that the optimal contract satisfying group incentive compatible with side-payments must assign the same quantity to x_i and x_j .

1.7.4. Proof of proposition 1.9

Because every agent has a single neighbor, the only adjacent coalitions are pairs of agents (i, j) such that $g_{ij} = 1$. We show that the optimal contract is immune to deviations by pairs of players with side-payments. First note that if $j \in A_1, i \in A_2$, then, as there is no $k \neq j$ such that $g_{ik} = 1$ and by definition $g_{jk} = 0$ for all k . Hence, any pair of quantities is group incentive compatible with side-payments.

Next pick two agents j, i in A_m, A_{m+1} with $m \geq 2$. Notice that there is no agent different from j such that $g_{ik} = 1$ whereas $g_{jk} = 1$ for some agent $k \in A_{m-1}$. Hence the condition

$$\sum_{k \neq j} g_{ik} x_j x_k + \sum_{k \neq i} x_i x_k \leq \sum_{k \neq j} g_{ik} x_i x_k + \sum_{k \neq i} g_{jk} x_j x_k,$$

is equivalent to

$$x_i \leq x_j \text{ for all } i, j \text{ such that } g_{ij} = 1, j \neq A_1. \quad (1.9)$$

Hence any contract for which $x_i \leq x_j$ whenever $g_{ij} = 1$ is group incentive compatible with side-payments.

Next, we show that the optimal contract \mathbf{x}^* satisfies condition (1.9). The proof is by induction on the tier of the hierarchy. Notice that because the tree is regular, $x_i^* = x_j^*$ for any i, j at the same tier of the hierarchy. Hence, we let x^{*m} denote the common quantity at tier m of the hierarchy.

Consider (i, j) such that $g_{ij} = 1$ and $i \in A_M, j \in A_{M-1}$. Then $|\sigma_i| = \emptyset$. Let l be the successor of j in the oriented tree. The optimal contract satisfies

$$\begin{aligned} x_i^* &= a + \alpha x_j^*, \\ x_j^* &= a + \alpha x_i^* + \alpha \sum_{k|g_{kj}=1} x_k^* + \alpha x_l^* \end{aligned}$$

so that

$$(x_i^* - x_j^*)(1 - \alpha) = -\alpha \sum_{k|g_{kj}=1} x_k^* + \alpha x_l^* < 0.$$

This shows that $x^{*M} < x^{*M-1}$.

Next suppose that $x^{*q} < x^{*q-1}$ for any $q > m$ and consider a pair (i, j) with $g_{ij} = 1$, $i \in A_m, j \in A_{m-1}$. Let l be the successor of j . The optimal contract satisfies

$$\begin{aligned} x_i^* &= a + \alpha x_j^* + \alpha \sum_{k|g_{ki}=1} x_k^*, \\ x_j^* &= a + \alpha x_i^* + \alpha \sum_{k \neq i, g_{kj}=1} x_k^* + \alpha x_l^*. \end{aligned}$$

Hence

$$(x_i^* - x_j^*)(1 - \alpha) = -\alpha \left(\sum_{k \neq i, g_{kj}=1} x_k^* - \sum_{k|g_{ki}=1} x_k^* \right) - \alpha x_l^*.$$

Now, by assumption $|\sigma_j| \geq |\sigma_i| + 1$, so $|\{k \neq i, g_{kj} = 1\}| \geq |\{k|g_{ki} = 1\}|$. In addition, by the induction hypothesis, $x_k^* > x_l^*$ for any k, l such that $g_{ki} = g_{lj} = 1$ as $k \in A_{m+1}, l \in A_m$ and by the induction hypothesis $x^{*m} > x^{*m+1}$. This shows that $x_i^* < x_j^*$, completing the proof of the Proposition.

2. Hidden Opinions

Abstract. This paper widens the scope of analysis of opinion dynamic models by introducing a novel heuristic: individuals choose to express their opinion or hide it, as a function of their local popularity. Intuitively, individuals who hide their opinion could be interpreted as individuals who have a low popularity such that even if they speak-up (or *tweet*) they will not be heard. Local popularity captures the idea that immediacy causes higher influence. Locally popular individuals express their opinion and can interact with like-minded or ideologically-opposed peers, namely expression entails debates and discussions. In the presence of hidden opinions, I show that the interactions between locally popular individuals and the magnitude of their influence explains whether consensus or polarization prevails. The primary mechanism at play is that the influence structure allows for consensus of opinion locally but communication between ideologically opposed expressers lead to global disagreement. The main contribution of this paper is to provide a unifying theoretical framework to assess different long-run opinion patterns with a focus on the topology of the network. I provide a measure of polarization and I run simulations to show the extent to which the topology of the network affects long-run opinion patterns.

Keywords: *Naive learning, repulsive influence, opinion polarization.*

JEL Classification: D83, D91, Z1.

2.1. Introduction

Are all people equal on social media or do popular voices dominate the conversation? My paper investigates this question by introducing a new heuristic to the study of opinion formation in social contexts. Namely I introduce an *expression heuristic*, which assumes that individuals choose whether to express their opinion or hide it based on how popular they are within their social network. Individuals who hide their opinion can be interpreted as individuals who have a low popularity such that even if they speak-up they will not be heard or considered. One can think of an individual who has very few followers on Twitter. Alternatively, a second interpretation can be that hiding one's opinion is less costly than expressing it. The cost of expression can be the time spent arguing with more eloquent and persuasive peers or the cost of social isolation when one's opinion drifts from the average group viewpoint.

The *expression heuristic* departs from early models of opinion formation, such as French (1956) [67], Harary (1959) [57], DeGroot (1971) [42]. In these seminal models, individuals pool their opinions by taking the average of opinions expressed by all of their direct contacts at every period of interaction, irrespective of differences in levels of expertise or influence. When referring to this classic opinion updating rule, I say that individuals update their opinions à la DeGroot. In that framework, regardless of the specific topology of the network, as long as it is (strongly) connected and aperiodic, individuals reach consensus of opinion in the long-run. The present paper refines these contributions by relating the opinion updating rule to the topology of the network. Namely, I introduce two types of individuals with different opinion updating rules. In doing so, I provide a unifying framework to assess different long-run opinion patterns, such as consensus, polarization of opinion or total disagreement.

Formally, I develop a model where a set of individuals are connected through an undirected network and exchange opinions about a given issue over a large number of periods. Each individual is exogenously allocated an initial opinion in a bounded interval, which represents their initial stance or attitude concerning the issue to be discussed. The influence of each individual is summarized by a centrality measure, which I call *local popularity*. This measure is defined and discussed further. As mentioned, there are two types of individuals: individuals who choose to *hide* and individuals who choose to *express*. If the local popularity of an individual is below a given threshold (hereafter *expression-threshold*), then they choose to *hide* and update their opinions à la DeGroot. Otherwise, individuals choose to *express* (hereafter *expresser*) and update their opinions according to a law of motion, which could incorporate either assimilation of opinions or distancing. At a given period, two di-

rectly connected expressers can either be like-minded or ideologically-opposed depending on the difference of their respective viewpoints. In particular, expressers only interact with peers who also choose to express. To be more precise, individuals who choose to express undergo an attractive effect (positive influence or assimilation) when interacting at a given period with like-minded peers who also choose to express. But when they interact with expressing peers that are ideologically-opposed, they undergo a repulsive effect (negative influence or distancing) and push each other to the upper and lower bound of the opinion interval. Intuitively, expression allows for a debate or a discussion to take place. This discussion could lead to agreement or can escalate to conflict.

To account for centrality, I define and use a measure that is: (i) increasing in the degree of a given individual (immediacy), (ii) decreasing in the average degree of direct friends of a given individual (strength). In other words, *local popularity* is defined as the degree of an individual divided by the average degree of friends. A locally popular individual has many friends, who themselves have very few influence sources. Naturally local and global centrality measures capture different aspects of influence. For example, when considering eigenvector centrality, the influence of an individual is high, when they are linked to other influential individuals. Global centrality measures require the knowledge of the whole network structure and are pertinent for many applications, e.g. the importance of website pages. Nevertheless, a local centrality measure may be better suited for a model of exchange of opinions or attitudes. In particular, well established strands of literature in social psychology, reviewed in section 2.2, argue that immediacy and strength of interaction are associated with higher social influence. Hence, an individual who may not be *globally* influential on the scale of the entire social network, can still have a high impact over their direct friends and local clustering of opinions can occur. It is the idea of having a devoted fan base.

I provide a novel unifying framework that explains how different opinion patterns prevail in the long-run. In the presence of hidden opinions, the study of the interactions between locally *popular* individuals who interact with like-minded or ideologically-opposed peers can explain whether consensus or polarization prevails. Since influence is stronger locally, clusters can form. But some members within a given cluster who are popular enough and interact with ideologically-opposed peers can fall into disagreement, which causes opinions to polarize across clusters. Expressers only pay attention to neighbors who also express, while consensual individuals who update their opinions à la DeGroot pay attention to all of their direct social contacts. This paper makes several contributions. First, I start by characterizing the opinions of expressers. Second I characterize the overall process of interpersonal influence with the two types of individuals and the two opinion updating rules. I show

that opinions converge in the long-run. Third, I provide simulations to relate the specific topology of the network to the level of opinion polarization; defined as the variance in the long-run opinions.

When an expresser has no neighbors who also choose to express, this individual remains stubborn and their long-run opinion corresponds to their initial opinion. Furthermore, expressers can have neighbors who are locally popular and choose to express. For instance, consider three expressers 1, 2 and 3. Individual 1 has a direct link to 2 and 2 has a direct link to 3, but 1 and 3 are not directly connected. Assume that all the remaining friends of those three individuals are not popular enough to express. The opinion update of individual 1 depends on their opinion difference with individual 2 and indirectly depends on the interaction between individual 2 and 3. Individuals 1, 2 and 3 form a set of connected expressers. Namely, this set is defined as follows: all the elements of the set are nodes (individuals) who choose to express and any two nodes are connected by a path of expressers within this same set.

With this definition in hand, proposition 2.1 considers a group of individuals who belong to the same set of connected expressers and shows that, if each member of the set has only like-minded neighbors, then the long-run opinions of each member is exactly the average of the initial opinions of the members of this set. Each expresser of the group undergoes only the attractive effect. In particular, the group can have members who are ideologically-opposed that are not neighbors. But since they are not directly linked, they do not repulse each other to the upper and lower of the opinion interval and become extreme. Hence the whole group can reach consensus by gathering different viewpoints. This result sheds light on the design of media tools or the formation of discussion groups for initiatives related to participatory democracy.¹

Proposition 2.2 characterizes the long-run opinions within a connected set of expressers, when a pair of initially ideologically-opposed neighbors belong to the group. I show that, generically, the long-run opinions of the group of connected expressers reach the upper or lower bound of the opinion interval. That is, all the members of the connected group of expressers become extreme. One special case where moderate opinions of expressers can survive in the long-run occurs when an expresser has like-minded neighbors. But they are indirectly connected to at least two ideologically-opposed expressing individuals. This case depicts a political left-right spectrum where parties on the far right and far left are ideologically-opposed and interact often together but between both parties, many moderate parties survive.

¹For example, think of the Citizens' convention for climate (*convention citoyenne pour le climat*) in France, where 150 citizens were randomly selected to work in smaller groups on different propositions.

Unlike linear models with assimilative influence, my model is non-linear; in the sense that the influence structure or the *hearing* matrix² can vary across periods in the presence of expressers who repulse or attract each other. Two expressers can initially be like-minded and influence each other positively. But in subsequent periods they can become ideologically-opposed and influence each other negatively, if one of them is repulsed by another expressing neighbor. Intuitively, this situation occurs when two individuals are somehow like-minded initially. But one of them starts adopting an extreme point of view in an unreasonable fashion under the influence of a third expressing peer, so her friend starts bringing to the table arguments which support the opposing view. In other words, the weights in the hearing matrix can depend on the opinion itself, as in Hegselmann and Krause (2002) [58].

Lemma 2.1 shows that the time or number of periods that two directly linked expressers take to repulse each other to the upper and lower bound of the opinion interval when they are ideologically-opposed, is smaller than the time they take to reach an agreement when they are like-minded. This motivates the study of the process of interpersonal influence with both types of individuals, starting from the time period by which repulsion opportunities in the course of a discussion are exhausted. I build a hearing matrix which accounts for both updating rules and show that opinions converge in the long-run. In particular, I show that the opinions of consensual individuals (who choose to *hide*) vanish in the long-run and remain hidden forever. Their long-run opinions simply become convex combinations of opinions of expressers to whom they are connected. Long-run consensus of opinions corresponds to an opinion vector such that the difference between any two opinions is small. Long-run bi-polarization of opinions corresponds to an opinion vector such that opinions belong to two groups of equal size: within each group opinions are close enough (below a threshold) and across both groups the opinion difference is large.

Proposition 2.3 provides necessary and sufficient conditions for consensus to prevail in the long-run. First, long-run consensus of opinions is obtained when there is a unique expresser. Second, consensus prevails when there exists multiple sets of connected expressers such that: (i) within each set, all members have like-minded neighbors, (ii) the average of initial opinions across two sets of connected expressers is close enough. Long-run opinions of consensual individuals depend on the long-run opinions of expressers to whom they are connected. Since those expressers all have similar viewpoints, consensual individuals adopt similar viewpoints when they take the average of opinions expressed within their social circle. The *type* of consensus obtained will depend first on how many expressers there are in the society. Their number is directly related to the structure of the network. Second, it

²This matrix provides information on who listens to whom or who pays attention to whom and the magnitude of this attention.

depends on the expressed opinions within the neighborhood of each expresser.

Mapping back a long-run bi-polarized opinion vector to an exact set of network structures and initial opinion distributions is tedious³. Consensual individuals are influenced by expressers to whom they are linked directly and indirectly. Put simply, they receive influence of different magnitudes from many expressers depending on how far or close they are from those expressers in the network. That is, one needs to account for all the possible paths of different length connecting consensual individuals to expressers who may be ideologically-opposed.

I provide compelling necessary conditions for long-run bi-polarization of opinions in lemma 2.2. I show that if a long-run opinion vector is bi-polarized then necessarily, (i) individuals who remain moderate in the long-run do not exist, (ii) both extreme influence groups influence an equal share of the society. The first condition means that there does not exist individuals who receive an equal amount of influence from two ideologically-opposed extreme opinion groups. Those individuals occupy a very particular location in the network, because they are not locally popular enough to express and they are equally influenced by two ideologically-opposed groups of expressers. Those individuals could be interpreted as neutral TV hosts, or non-biased journalists or intermediaries in general.

Simulations. I explore the model through simulations.⁴ The objective is to relate the topology of the network to the long-run opinion patterns. Polarization of long-run opinions is measured by taking the variance of final opinions. I generate initial opinions uniformly at random in a bounded interval and study the evolution of opinions of a large set of individuals in scale-free networks. The degree distribution within scale-free networks follows a power law. Due to this inequality in the degree distribution, expressers and consensual individuals co-exist. In particular, this exercise shows that average polarization level can be relatively low even if expressers are densely connected among each other. This happens precisely for network topologies where consensual individuals are connected to many influence sources.

Related literature. The study of opinion dynamics is a multi-disciplinary topic. Different fields such as economics (learning in networks, for surveys see Golub and Sadler (2017) [54] and Acemoglu and Ozdaglar (2011) [3]), sociology (the community cleavage problem, see Flache et al. (2017) [46] or Friedkin (2015) [49]), statistical physics and computer science (community detection, Malliaros and Vazirgiannis (2013) [74] survey the literature) have

³Hegselmann and Krause (2002) [58] have a non-linear model and they discuss this point extensively: *though elementary, the model is nonlinear in that the structure of the model changes with the states of the model given by the opinions of the agents (see Section 2). Not only that helpful mathematical tools like Markov chains are no longer applicable, it turns out, moreover, that rigorous analytical results are difficult to obtain..*

⁴The Matlab code can be found in the following repository <https://github.com/shadenshabayek/Hidden-Opinions>.

tackled this problem from different angles.

This paper is closely related to a strand of the literature on Naive Learning which introduces *stubborn* agents or agents that remain attached to their initial opinion to a certain extent, in order to model disagreement (See Friedkin and Johnsen (1990) [48] and Friedkin(2015) [49], Acemoglu et al (2013) [1]). Two papers that are closest to mine are Yildiz et al. (2013) [100] and Sadler (2019) [87]. Both papers introduce *stubborn* agents in a voter model set-up where opinions are discrete and can take only two discrete real values either a or b . Players can be either stubborn, that is they never update their opinion, or they can update à la DeGroot. Nevertheless, the stubbornness of a player is independent of their network position. Hence a stubborn player who is nor locally nor globally central can have a great impact on the long-run opinions of all individuals in the network. With that respect, I extend this approach by relating the impact a stubborn player can have on others' long-run opinions, to their popularity.

My paper also fits in the family of bounded confidence models (See Hegselmann and Krauss (2002) [58], Jager and Amblard (2005) [64]). The key ingredient of those models is to consider the difference between the opinions of individuals when opinion updating is taking place. In particular, Hegselmann and Krauss (2002) [58] consider a model where agents update their opinions by taking an average over the opinions of neighbors whose opinion difference falls within a confidence interval. When neighbors opinions fall outside the confidence interval they are ignored. I extend this literature by introducing an opinion updating rule which treats the opinions of neighbors differently, depending on whether the opinions of neighbors fall within or outside the confidence interval.

Furthermore my paper is related to an active line of research about disagreement in social contexts and oppositional identity. Both literatures use different methodologies (dyadic interaction between agents versus interaction with a proportion of agents) but address very similar questions. Melguizo (2018) [77] studies persistent disagreement. She allows interactions and attitudes to co-evolve, hence departing from the time independent weights used in averaging neighbors' opinions in models à la DeGroot. The key ingredient used is to assume that each individual has several attributes. Relations with other individuals sharing similar attributes become more intense, while relationships with dissimilar others deteriorate. Furthermore, disagreement can modeled by introducing repulsive or negative influence when an individual interacts with dissimilar others or by modeling individuals who are similarity biased (See Flache et al. (2017) [46] for a survey). As for oppositional identity, Bisin et al. (2016) [21] provide a model which incorporates cultural conformity and cultural distinction, in the context of marriage choice within the same ethnic group or outside. Individuals select their optimal choice by considering the psychological cost of interacting with the proportion of dissimilar others.

Finally, explaining polarization has been tackled by a handful of recent papers, yet there is no consensus in the literature about its main drivers. Bolletta and Pin (2019) [25] introduce a network formation model and argue that under certain conditions when agents optimally choose their links, the network can become disconnected and consensus of opinions cannot be reached. Banisch and Olbrich (2019) [12] explain the emergence of polarization by introducing reinforcement learning, where agents optimally adopt one viewpoint when they get positive feedback from peers. But their focus is not on the network structure itself and how it could be one of the drivers, in particular they fix a random geometric network to account for the structure of interactions.

Outline. The remainder of the paper is organized as follows. Section 2.2 reviews relevant literatures in social psychology that lay the ground for the main behavioral assumptions of my model. In section 3.3, I present the model and I characterize the overall process of interpersonal influence in section 2.4. In section 2.5, through simulations I relate the topology of networks to long-run opinion patterns. Section 3.5 concludes.

2.2. Related work in social psychology

The two main novel ingredients of the model are the expression heuristic and the local popularity measure. The former ingredient is grounded in a well known literature in social psychology called Hidden Profiles⁵ and the latter is borrowed from the Dynamic Impact Social Theory. In what follows I briefly review both literatures.

2.2.1. Hidden Profiles

Stasser and Titus (1985) [94] document how individuals in social contexts, do not always share their own opinion or the information they hold. The starting point of their research is to challenge the common belief that a group of individuals should be able to take a better decision than each individual on their own by pooling the members' knowledge and expertise. Namely, group discussion or communication is believed to have a corrective function because members can each have incomplete information but together they can gather the different pieces of the puzzle. The authors ran an experiment in which they simulate a political set-up where a group has to elect one of three candidates: *Best*, *Okay* and *Ohum*. In a first protocol, they distributed a different subset of desirable traits of *Best* and a different subset of *Okay*'s undesirable traits over the members of the group, such that from each one's individual perspective *Okay* appeared more positive than *Best*. Before discussion

⁵See the survey *Hidden Profiles: a brief history* by Stasser and Titus (2003) [95]

Best received 25% of endorsement. Since the whole group had complete (but dispersed) information about *Best* they could exchange it and come to the conclusion that *Best* was actually the best candidate. Yet after group discussion, surprisingly the percentage of endorsement for *Best* remained at 24%. This finding suggests that unique information held by some members of the group about candidates were not being shared. In a later study, Stasser, Taylor and Hanna (1989) [93] showed that unique pieces of information are less likely to be mentioned during group discussion. One explanation is that social status, expertise or popularity can be a driver for expression of opinion. In fact, Larson et al. (1996) [66] suggest that repeating a unique piece of information, leading to the formation of group opinion during a discussion, is more likely by higher status members (experts, leaders, etc.) rather than lower status members. They ran an experiment with residents, interns and 3rd-year medical students and they show that residents were more likely to repeat (unique) information when compared to interns and students.

Using the findings of the above literature, I introduce an expression heuristic to a dynamic opinion formation model. An individual chooses to express her opinion or hide it based on a popularity measure that is meant to capture different hierarchical and expertise levels.

2.2.2. Dynamic social Impact theory

Latané's Dynamic Social Impact Theory (1981) [72] suggests that social influence has three determinants: strength, immediacy and the number of influence sources. Strength refers to social status, level of expertise or persuasiveness, while immediacy refers to closeness in space, time or the possibility of direct contact. The theory bridges the influence processes at an individual level using these three determinants, to outcomes at the level of a social system. The main statement of the theory is that *total impact of a group of people on an individual is a multiplicative function of their strength, immediacy, and number*. Latané, Nowak and Liu (1994) [73] use this theory to study through simulations the dynamics of attitude change in groups and societies. Rather than studying attitude distribution as the usual percentage frequencies of different attitude choices, they study the distribution of attitudes in *space*. Using *immediacy*, they are able to explain phenomena such as attitude clustering because individuals are more influenced by nearby individuals. In particular, *distance* between persons is used to compute *immediacy* and it is used along with *strength* to compute for each person the total persuasive impact and supportive impact. *A person will change his/her position if and only if the total persuasive impact (the pressure to change to a different position) outweighs the pressure to maintain one's own position (the strength of the initial position plus any supportive impact)*. Using computer simulations, they find that individuals cluster in the social space in terms

of position similarity.

To incorporate *immediacy* in my model, I use a local popularity measure as opposed to global measures (e.g. eigenvector centrality). Individuals choose to express when their local popularity is above a given threshold. To incorporate *strength*, the local popularity of a given individual is inversely related to the average level of neighborhood connectivity. Meaning that an individual will have a higher impact over their neighbors if those same neighbors are not exposed to many other influence sources.

2.3. The model

2.3.1. Set-up

A group of individuals $N = \{1, \dots, n\}$ is embedded in a connected and symmetric network G of interpersonal relationships, with typical entries $g_{ij} = g_{ji} \in \{0, 1\}$. Each node represents an individual. The set of friends, colleagues or acquaintances of individual $i \in N$ is denoted by $N_i = \{j \in N : g_{ij} = 1\}$ and $d_i = |N_i|$ is the cardinality of N_i . For all $i \in N$, I assume that $g_{ii} = 1$. A chain of friends of friends of length l between two individuals $i \neq j \in N$, hereafter called *path*, is defined as follows: there exists a sequence of distinct individuals $i = k_0, k_1, \dots, k_l = j \in N$ such that $g_{ik_1} \times g_{k_1k_2} \times \dots \times g_{k_{l-1}k_l} > 0$. Finally, the social status or relative expertise is represented by a local centrality measure which I call *local popularity*.

Definition 2.1 (Local popularity) Let $i \in N$ be an individual and N_i the set of her/his direct friends. Local popularity is:

$$\delta_i = \frac{d_i - 1}{\frac{1}{d_i - 1} \sum_{j \neq i \in N_i} (d_j - 1)} \quad (2.1)$$

Notice that the network G is connected and contains self-loops the above centrality measure is always defined. Alternatively, without making any assumptions, for any individual $i \in N$, if $d_i = 1$ then set $\delta_i \equiv 0$.

Assumption 1 Each individual $i \in N$ knows d_i and d_j for all $j \in N_i$.

2.3.2. Expression heuristic

Each individual is endowed with an exogenous initial opinion $\alpha_{i,0} \in [-1, 1]$ which represents their attitude or stance⁶ about an issue or a policy $\theta \in [-1, 1]$. Individuals exchange opinions about the issue over periods $t \geq 0$. First, each individual $i \in N$ observes opinions

⁶Here an opinion is represented as a real number.

in their neighborhood N at period t , denoted by $\alpha_{j,t} \in [-1, 1]$ for all $j \in N_i$. Second, each individual chooses in period $t + 1$, whether to express or hide their opinion based on their local popularity parameter (2.1). They use the following heuristic:

$$\begin{cases} \text{play hide} & \text{if } \delta_i < \delta^* \\ \text{play express} & \text{if } \delta_i \geq \delta^*, \end{cases}$$

where δ^* denotes an exogenous expression threshold. In the remainder of the paper, the set of individuals who choose to express will be labelled $E = \{i \in N, \text{ s.t. } \delta_i \geq \delta^*\}$. Similarly, the set of individuals who choose to hide their opinions will be labelled $C = \{i \in N, \text{ s.t. } \delta_i < \delta^*\}$.

Express & Hide. When an individual chooses to express, her/his opinion update will depend on her/his neighbors, who also choose to express. To account for the difference in opinions of two neighbors who interact, I compare their opinion difference to a threshold $\tau \in (0, 1)$.

Definition 2.2 *Two individuals $i \neq j \in N$ who choose to express are : (i) like-minded at period $t \geq 1$ if $|\alpha_{i,t} - \alpha_{j,t}| < \tau$, (ii) ideologically-opposed at period $t \geq 1$ if $|\alpha_{i,t} - \alpha_{j,t}| \geq \tau$. Moreover, $\underline{N}_{i,t} = \{j \in N_i \cap E, |\alpha_{i,t} - \alpha_{j,t}| < \tau\}$ and $\bar{N}_{i,t} = \{j \in N_i \cap E, |\alpha_{i,t} - \alpha_{j,t}| \geq \tau\}$.*

Expression allows for a debate or discussion and opinion update of expressers follows the law of motion described below. It incorporates an attractive and a repulsive effect among direct neighbors who express their opinions. For $\forall i \in E$ and $\mu \in (0, 1/2)$:

$$\alpha_{i,t} = \alpha_{i,t-1} + \Delta_{\alpha_{i,t-1}} \text{ s.t. } \alpha_{i,t} \in [-1, 1], \quad (2.2)$$

where

$$\Delta_{\alpha_{i,t-1}} = \mu \sum_{j \in \underline{N}_{i,t-1}} (\alpha_{j,t-1} - \alpha_{i,t-1}) - \mu \sum_{j \in \bar{N}_{i,t-1}} (\alpha_{j,t-1} - \alpha_{i,t-1}).$$

Notice that a positive weight is assigned to the opinion of like-minded expressing neighbors at period t . While a negative weight is assigned to the opinion of ideologically-opposed expressing neighbors at period t . Moreover, the weights assigned to the opinion of expressing neighbors can change across periods. This happens because expressing neighbors receive influence from their own expressing neighbors (if any). Hence, an initially like-minded expressing neighbor of individual i can become in subsequent periods ideologically-opposed, if their opinion difference with i becomes larger than τ .

When an individual chooses to hide, they update their opinion by taking the average of opinions expressed within their social circle at the previous period:

$$\alpha_{i,t} = \bar{\alpha}_{i,t-1} = \frac{1}{d_i} \sum_{j \in N} g_{ji} \alpha_{j,t-1}. \quad (2.3)$$

Example 4 Consider a network G with only two connected individuals 1 and 2 who choose to express. Suppose that $\tau = 0.5$ and initial opinions are $\alpha_{1,0} = -0.7$ and $\alpha_{2,0} = 0.7$. In period $t = 1$, $\alpha_{1,1} = \alpha_{1,0} + \mu(\alpha_{1,0} - \alpha_{2,0}) = -0.7(1 + \mu) - \mu 0.7 < \alpha_{1,0} = -0.7$ and $\alpha_{2,1} = \alpha_{2,0} + \mu(\alpha_{2,0} - \alpha_{1,0}) = 0.7(1 + \mu) + \mu 0.7 > \alpha_{2,0} = 0.7$. The updated opinion of individual 1 becomes more negative or pushed-down towards -1 , while the updated opinion of individual 2 more positive or pushed-up towards 1. Individuals 1 and 2 repulse each other.

2.3.3. Micro-foundation

The *expression heuristic* can be micro-founded by a simple game where individuals face a social cost of expression. Intuitively this cost is a psychological cost of disagreement with peers or, the cost of a long lasting debate because of strong opposing views. At each time period, individuals observe the actions (opinions) of neighbors at the previous period and myopically best-respond at the current period. The payoff of individual $i \in N$ given action $a_{i,t} \in \{\text{express, hide}\} = \{\alpha_{i,t}, \bar{\alpha}_{i,t-1}\}$ is the following:

$$\pi_i(a_{i,t}, a_{-i,t-1}) = -f(\delta_i)(\alpha_{i,t} - a_{i,t})^2 - (1 - f(\delta_i))(\bar{\alpha}_{i,t-1} - a_{i,t})^2. \quad (2.4)$$

The function f has support $(0, 1)$ and $f' > 0$. The first term of the payoff function (2.4) is the cost borne by each player when they does not express their opinion, in other words it is the cost of being consensual. The second term is the cost of expression, which is a cost borne when a player expresses their own opinion rather than their neighborhood average opinion. Both terms are weighted by an increasing function of the local popularity parameter δ_i . The higher δ_i is, the higher the influence of player i within their own neighborhood and the lower the cost of expression or social disagreement. Given the payoff function (2.4), any $i \in N$ such that $\delta_i \geq \delta^* = f^{-1}(\frac{1}{2})$ best-responds with *express*, while any $i \in N$ such that $\delta_i < \delta^* = f^{-1}(\frac{1}{2})$ best-responds with *hide*. Hence the local centrality parameter of each individual is a sufficient statistic for the choice of the optimal binary action at each stage-game and individuals can be classified into two groups: $E = \{i \in N, \delta_i \geq \delta^*\}$ and $C = \{i \in N, \delta_i < \delta^*\}$.

Assumption 2 For the remainder of the paper, the expression threshold is normalized to $\delta^* = 1$.

Given the local popularity measure δ_i of individual i , the threshold $\delta^* = 1$ allows all

individuals to play *express* when they are all equally popular. Namely, in any d -regular network where each individual has degree d , any player $i \in N$ expresses because $\delta_i = \frac{d}{\frac{1}{d}d \times d} = \delta^*$. Intuitively, d -regular networks have perfect assortativity meaning that there is no difference in the level of popularity, expertise or leadership.

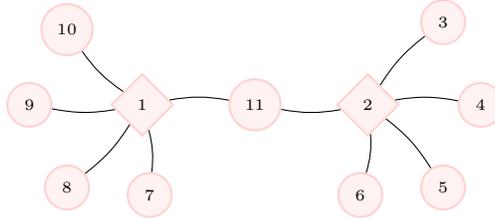


Figure 2.1: Network G_1 , diamond shaped nodes correspond to individuals who choose to *express*.

Example 5 Consider the network G_1 in Figure 2.1. Individuals 1 and 2 both have local popularity $\delta_1 = \delta_2 = 5 / ((1/5)(1 + 1 + 1 + 1 + 2)) = 25/6 > \delta^*$, therefore they play *express* and $E = \{1, 2\}$. While individuals $j \in \{3, \dots, 9, 10\}$ have a local popularity $\delta_j = 1/5 < \delta^*$ and individual 11 has $\delta_{11} = 4/10 < \delta^*$, hence they play *hide* and $C = \{3, \dots, 10, 11\}$.

2.4. Opinion Dynamics

Given the above model, I am interested in studying the long-run opinions, for a network structure G and a vector of initial opinions α_0 . Influence can flow from expressers to consensual individuals but not vice versa. I start by considering the evolution of opinions of expressers. Then I characterize in section 2.4.2 the overall dynamics of opinions and show that long-run opinions always converge.

2.4.1. Long-run opinions of expressers

Individuals who choose to express update their opinions at each time period according to the law of motion (2.2). Consequently, their opinion update will directly depend on the opinions of neighbors who choose to *express* (if any). The opinions of the latter will depend on the opinions of their own neighbors (if any) who choose to *express* and so on. Recall that E is the set of individuals who choose to express because their local popularity is higher than the expression threshold δ^* . In order to account for the indirect effect of the opinions of expressers on other individuals who also express, I give a formal definition of a connected set of expressers.

Definition 2.3 (Connected set of expressers) Let G be a given network structure and let \mathcal{E} be a set of individuals such that (i) $\forall i \in \mathcal{E}, \delta_i \geq \delta^*$, (ii) $\mathcal{E} \subseteq E$ and (iii) any pair of individuals $i \neq j \in \mathcal{E} \subseteq E$ are connected in network G by a path of length l_{ij} of expressers belonging to the set \mathcal{E} , that is $\exists g_{ik_1} \times g_{k_1k_2} \dots \times g_{k_lj} > 0$ for $k_1, \dots, k_l \in \mathcal{E} \subseteq E$. When expressers $k_1, \dots, k_l \in \mathcal{E} \subseteq E$ all have like-minded neighbors, I say that individual i is linked to individual j through a path of like-minded expressers and denote it by l_{ij}^+ .

In a given network there could be multiple connected sets of expressers $\mathcal{E}_1, \dots, \mathcal{E}_k$ such that $E = \bigcup_{i=1}^k \mathcal{E}_i$. Those sets could be singletons or they could contain more than one expresser. In the network G_1 in figure 2.1, each of both individuals 1 and 2 form a set of connected expresser(s) on their own: $\mathcal{E}_1 = \{1\} \subset E$ and $\mathcal{E}_2 = \{2\} \subset E$. Moreover, $E = \mathcal{E}_1 \cup \mathcal{E}_2 = \{1, 2\}$.

For a given set of connected expressers, if each pair of neighbors are like-minded, then their long-run opinions will be the average of their initial opinions. I formalize this idea in the following proposition.

Proposition 2.1 Let G be a network of interpersonal relationships, α_0 an initial opinion vector and consider $\mathcal{E} \subseteq E$ a given set of connected expressers.

(i) (Stubborn) If $|\mathcal{E}| = 1$ and $\mathcal{E} = \{i\}$ then

$$\forall t \geq 1, \alpha_{i,t} = \alpha_{i,0}$$

(ii) (Like-minded) If $|\mathcal{E}| = \kappa > 1$ and $\forall i \neq j \in \mathcal{E}, \forall j \in N_i \cap \mathcal{E}, |\alpha_{i,0} - \alpha_{j,0}| < \tau$ then for $\mu \in (0, 1/\kappa)$ and $j_1, \dots, j_\kappa \in \mathcal{E}$,

$$\exists t^* \geq 1, \forall t \geq t^*, \alpha_{i,t} = \frac{\alpha_{i,0} + \alpha_{j_1,0} + \dots + \alpha_{j_\kappa,0}}{\kappa}$$

Proof. See Appendix 2.7.1. ■

When a set of connected expressers contains at least one pair of ideologically-opposed neighbors, *generically* long-run opinions become extreme, that is they take the value 1 or -1 . For very specific initial opinion distributions some expressers within the same set can hold moderate opinions (strictly between -1 and 1) in the long-run. I give two examples to make clear: (i) why opinion updating within a set of connected expressers is non-monotonic, (ii) why moderate opinions can survive in the long-run. Finally, I summarize the discussion in proposition 2.2. In all the subsequent examples, I assume that the opinion difference of two directly linked individuals at a given time period is $\tau = 0.5$.

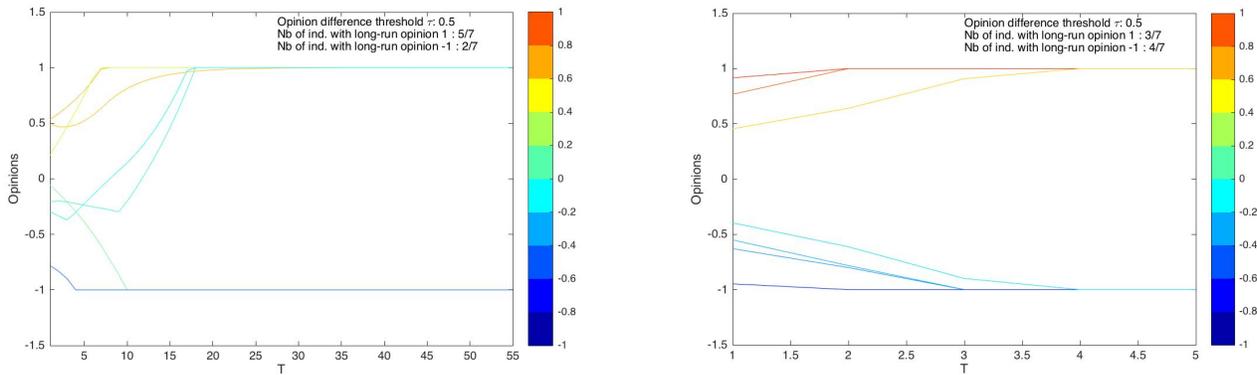


Figure 2.2: Non-monotonic and monotonic opinion updating in a circle

Non-monotonic opinion updating

The coexistence of pairs of directly linked individuals who are like-minded and pairs of directly linked individuals who are ideologically-opposed within the same connected set of expressers can lead to either monotonic or non-monotonic opinion updating. Non-monotonic opinion updating can occur when an individual is indirectly connected to an ideologically-opposed individual through a chain of like-minded expressers.

To see this, consider a wheel G of seven individuals $\{1, \dots, 7\}$ such that $g_{12} = g_{23} = \dots = g_{67} = g_{71} = 1$ and the remaining entries of G are zero. All individuals express because their local popularity is higher or equal to δ^* and together all seven form a unique connected set of expressers. Consider the following initial opinion vector:

$$\alpha_0 = (-0.21, -0.06, 0.53, 0.49, 0.21, -0.78, -0.29).$$

Individuals 5 and 6 are initially ideologically-opposed with initial opinions $\alpha_{5,0} = 0.21$ and $\alpha_{6,0} = -0.78$. The remaining individuals $\{1, 2, 3, 4, 7\}$ have initially like-minded neighbors. The evolution of opinions is plotted in the left panel in figure 2.2, with time periods on the x -axis and opinions on the y -axis. The colormap on the east side of the figure is associated with the opinion interval $[-1, 1]$ and the colors of the curves correspond to initial opinions.

Individuals 5 and 6 are pushed to the upper and lower bound of the opinion interval $[-1, 1]$ after few periods of interaction, as they are ideologically opposed. The evolution of their opinion is monotonic. The opinion of individual 5 becomes more and more positive, while the opinion of individual 6 becomes more and more negative. However, individuals 1 and 7 update their opinions non-monotonically. Individual 7 and 6 are initially like-minded and the attractive effect is at play in the first few periods of interaction. Hence the opinion of individual 7 starts becoming more negative, because it converges towards the opinion of individual 6. But after a few periods, individual 6 becomes extreme by reaching the lower

bound -1 . The opinion difference with their direct neighbor individual 7 becomes larger and larger, until a point where this difference becomes higher than the threshold τ and individual 7 starts getting repulsed by the extreme opinion of individual 6. In other words, the repulsive effect takes over. Intuitively, this situation occurs when a given individual i is having a discussion with an individual j who is initially more or less like-minded, but individual i is more neutral than j . As the discussion goes on, individual j becomes too extreme in an unreasonable fashion, such that individual i starts defending the opposite view.

Finally, in the right panel of figure 2.2, I provide an example where opinion updating is monotonic. Initial opinions are given by $\alpha_0 = (-0.55, 0.77, -0.63, -0.9478, 0.92, -0.4, 0.45)$. In other words, each individual has an ideologically-opposed neighbor and opinions converge *monotonically* to the upper or lower bound.

Moderate long-run opinions

There exist initial opinion distributions such that *moderate* opinions (i.e. with an opinion that is neither 1, nor -1) survive in the long-run within a set of connected expressers containing at least one ideologically-opposed pair. To see this, consider again a wheel G of 7 individuals such that $g_{12} = g_{23} = \dots = g_{67} = g_{71} = 1$ and all the remaining entries of G are zeros. Suppose that initial opinions are: $\alpha_0 = (-0.9, -0.7, -0.4, -0.1, 0.2, 0.5, 0.8)$. Individuals 1 and 7 are initially ideologically-opposed. Individuals $\{2, \dots, 6\}$ all have neighbors who are initially like-minded. The evolution of opinions is plotted in the left panel of figure 2.3. Individuals $i \in \{2, \dots, 5\}$ remain moderate in the long-run and never adopt an extreme opinion of 1 or -1 . To understand why moderate opinions can persist in the long-run, first notice that individual 7 is repulsed by her direct ideologically-opposed neighbor individual 1. Second, the opinion difference of individuals 6 and 7 remains smaller than τ even when 7 becomes extreme. Similarly, the opinion difference of individuals 1 and 2 remains smaller than τ even when 1 becomes extreme. Intuitively, extreme individuals 1 and 7 influence in opposite *directions* the chain of like-minded individuals that separate them so that in the long-run, each of these intermediate individuals have like-minded neighbors and remain moderate. This situation bears a resemblance to the left-right political spectrum in some countries, where the moderate parties survive in the long-run. To summarize, moderate opinions of expressers can persist in the long-run if such expressers only have like minded-neighbors and are linked to at least two ideologically-opposed expressers by a path of like-minded neighbors.

In the right panel of figure 2.3, all individuals hold extreme opinions in the long-run. Similarly to the setting in the left panel of figure 2.3, individuals $i \in \{2, \dots, 6\}$ have ini-

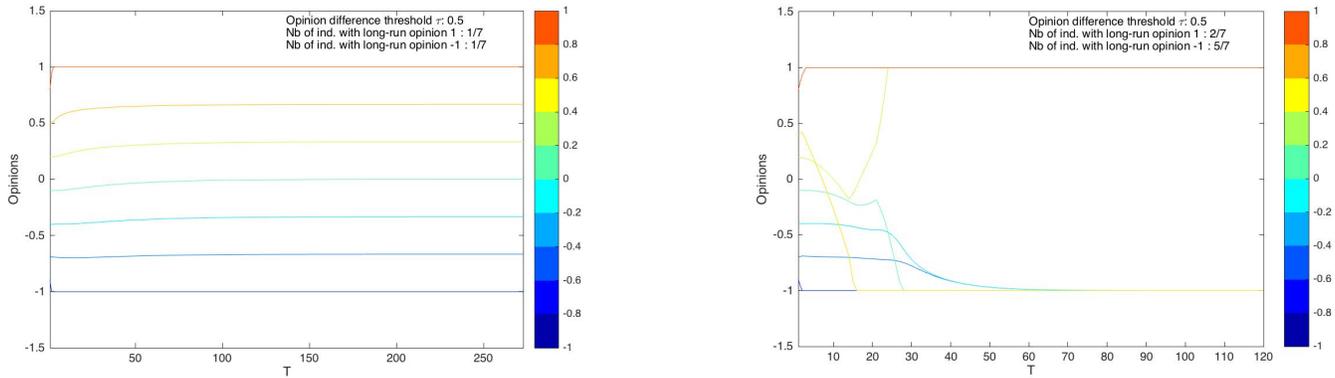


Figure 2.3: Moderate expressers

tially like-minded neighbors and are connected to individuals 1 and 7 directly or indirectly through a chain of like-minded neighbors. But the opinion difference of individuals 6 and 7 is initially larger than in the previous case (left panel of figure 2.3) and as individual 7 is pushed towards holding an extreme opinion, her opinion difference with individual 6 becomes larger and larger, up to a point where this difference becomes larger than τ .

Finally, consider a wheel G of three individuals such that $g_{12} = g_{23} = 0$ and all the remaining entries of the network G are zeros. All three individuals express and all three form together a unique connected set of expressers. Suppose that the initial opinion of individual 2 is $\alpha_{2,0} = 0$. Suppose that individuals 1 and 3 hold respectively the following initial opinions $\alpha_{1,0} = -1$ and $\alpha_{3,0} = 1$. In this case, the long-run opinion of individual i is $\alpha_{i,\infty} = 0$ because she is equally repulsed by both neighbors but in opposite directions.

To summarize, when a given connected set of expressers is formed of only initially like-minded expressers, influence is positive and opinions get *attracted* to the average opinion of the group. However, when the set contains at least one ideologically-opposed pair of neighbors, the repulsive and attractive effect can both be at play. I show in the following proposition, that when a set contains at least one ideologically-opposed pair of expressers, generically opinions get to the upper and lower bound of the opinion interval. Intuitively, some expressers are not updating their opinions by taking a convex combination of the opinions of their neighbors. Hence at each time step the length of the interval given by the opinion difference of an ideologically-opposed pair keeps growing until it reaches the bounds. Such expressers also drive their like-minded neighbors to reach the upper and lower bound of the opinion difference. This happens, because when those like-minded neighbors are updating their opinion, they assign a positive weight to the opinions of their neighbors. Except that the opinion of the latter keeps getting pushed to either bounds under the effect of their own

ideologically-opposed neighbor, rather than converging to the opinion of their like-minded neighbor. I given in points (i) and (ii) of the proposition, the two exceptions where some opinions remain moderate as explained in the previous examples.

Recall that $\underline{N}_{i,t}$ and $\overline{N}_{i,t}$ are the sets of neighbors of $i \in N$ who are respectively like-minded and ideologically-opposed at period $t \geq 0$. Denote the set of expressers that have at least one initially ideologically-opposed neighbor by $IO(\mathcal{E}) = \{i \in \mathcal{E} : \exists j \in \mathcal{E}, i \in \overline{N}_{j,0}, j \in \overline{N}_{i,0}\}$.

Proposition 2.2 *Let G be a network of interpersonal relationships, α_0 an initial opinion vector and consider $\mathcal{E} \subseteq E$ a set of connected expressers. Suppose that $IO(\mathcal{E}) \neq \emptyset$.*

- (i) *If there exists at each $t \geq 0$ at least two paths of expressers with only like-minded neighbors connecting $i \notin IO(\mathcal{E})$ to at least two elements of $\{i_1, i_2 \dots i_k\} \subset IO(\mathcal{E})$ then $\alpha_{i,\infty} \in \text{conv}(\alpha_{i_1,\infty}, \dots, \alpha_{i_k,\infty})$.*
- (ii) *Let $i_1, i_2 \dots i_k \in IO(\mathcal{E})$. If there exists $i \in IO(\mathcal{E})$ such that (a) $\alpha_{i,0} = 0$, (b) $\forall t \geq 0 \underline{N}_{i,t} = \emptyset$, (c) and $\sum_{j \in \overline{N}_{i,t}} \alpha_{j,t} = 0$ then $\alpha_{i,\infty} = 0$.*
- (iii) *Otherwise if (i) and (ii) don't hold and $IO(\mathcal{E}) \neq \emptyset$ then $\forall i \in \mathcal{E}, \alpha_{i,\infty} \in \{-1, 1\}$.*

Proof. See Appendix 2.7.2. ■

Furthermore, when a pair of expressers are initially like-minded they take a longer *time* to reach consensus but when the pair is ideologically-opposed they disagree at a much faster rate. I formalize this idea in the following lemma.

Lemma 2.1 *Let $i \neq j \in \mathcal{E} \subset E$ such that $|\mathcal{E}| = 2$ and $t_{\alpha_\infty} = \min\{t : |\alpha_t - \alpha_\infty| < \epsilon\}$. If α_∞^a is the long-run opinion vector when $|\alpha_{i,0} - \alpha_{j,0}| < \tau$ and α_∞^r is the long-run opinion vector when $|\alpha_{i,0} - \alpha_{j,0}| \geq \tau$ then $t_{\alpha_\infty^r} < t_{\alpha_\infty^a}$.*

Proof. See Appendix 2.7.3. ■

2.4.2. The process of interpersonal influence

I build a hearing matrix which takes into account who listens to whom. This hearing matrix takes into account the two opinion updating rules depending on the type of individual: expresser (updates according to the law of motion (2.2)) or consensual (updates à la DeGroot). In particular I study the long-run behavior starting at the time period where pairs

of ideologically-opposed expressing neighbors have repulsed each other towards the most extreme opinion. As I have shown in the previous section, given the parameter μ in the law of motion (2.2), a pair of ideologically-opposed individuals repulse each other at a faster rate than a pair of like-minded individuals who debate to reach a consensus. Formally let $t^* \geq t$ be the time period by which the least ideologically-opposed pair of directly connected expressers, in the group of individuals N , have repulsed each other to reach opinions at the upper and lower bound of the opinion interval. That is, for any period t beyond time period t^* , ideologically-opposed neighbors who express, are no longer updating their opinions and have long-run opinions that are either 1 or -1 . Given a network G representing interpersonal relationships, denote by \tilde{G} the hearing matrix with typical entries \tilde{g}_{ij} .

Consensual individuals. For each individual $i \in N$ such that $\delta_i < \delta^*$, the entries in the hearing matrix become $\tilde{g}_{ij} = g_{ij}/d_i, \forall j \in N$.

Expressers. For individuals who choose to express there are four cases to consider.

- (i) For all $i \in N$ in a connected set of expressers \mathcal{E} such that $|\mathcal{E}| = 1$ (stubborn), the entries of the hearing matrix \tilde{G} are: $\tilde{g}_{ii} = 1$ and $\tilde{g}_{ij} = 0$ for all $j \in N_i$.
- (ii) For all $i \in N$ in a connected set of expressers \mathcal{E} such that $|\mathcal{E}| = \kappa > 1$ with like-minded neighbors at period t^* , the entries of the hearing matrix \tilde{G} are: $\tilde{g}_{ii} = 1 - |N_i \cap \mathcal{E}| \mu$, $\tilde{g}_{ij} = \mu$ for $j \in N_i \cap \mathcal{E}$ and $\tilde{g}_{ij} = 0, \forall j \notin N_i \cap \mathcal{E}$.
- (iii) For all $i \in N$ in a connected set of expressers \mathcal{E} such that $|\mathcal{E}| = \kappa > 1$ with ideologically-opposed neighbors, i.e. $\forall i \neq j \in \mathcal{E}$ and $j \in N_i \cap \mathcal{E}$, $|\alpha_{i,t^*} - \alpha_{j,t^*}| \geq \tau$, the entries in the hearing matrix \tilde{G} are: $\tilde{g}_{ii} = \tilde{g}_{jj} = 1$, $\tilde{g}_{ik} = 0$ for all $k \in N_i$ and $\tilde{g}_{jk} = 0$ for all $k \in N_j$.

Remark 2.1 All the entries of the hearing matrix \tilde{G} are positive and each row sums to one.

Example 6 Consider network G in figure 2.1. The hearing matrix \tilde{G} has the following entries for expressers 1 and 2 who are both stubborn: $\tilde{g}_{11} = \tilde{g}_{22} = 1$ and $\tilde{g}_{1j} = \tilde{g}_{2j} = 0$ for all $j \in N_1 \cup N_2$. For the consensual individuals $i \in \{3, 4, 5\}$, $\tilde{g}_{ii} = \tilde{g}_{i2} = 1/2$. For $i \in \{6, 7, 8\}$, $\tilde{g}_{ii} = \tilde{g}_{i1} = 1/2$. Finally $\tilde{g}_{99} = \tilde{g}_{91} = \tilde{g}_{92} = 1/3$. All the remaining entries are zeros.

For a given network structure G , the process of interpersonal influence describing the evolution of opinions at period $t \geq t^*$ is given by the following equation:

$$\alpha_{t+1} = \tilde{G}\alpha_t \quad (2.5)$$

By induction, the opinions at period $t \geq t^*$ are given by $\tilde{G}^t \alpha_{t^*}$ and the limit yields the long-run opinions. A few comments are in order.

First, the entries of the hearing matrix \tilde{G} are all non-negative and all the diagonal entries are strictly positive. Moreover it has rows and columns that sum to one. Hence, the eigenvalues of \tilde{G} are all lower or equal to 1 and $\lim_{t \rightarrow \infty} \tilde{G}^t$ exists. The entry on the row i and column j of the matrix $\lim_{t \rightarrow \infty} \tilde{G}^t$ is the weight (between 0 and 1) that the opinion of individual i at period t^* has in the final opinion of individual j .

Second, the hearing matrix \tilde{G} is a reducible. To see this, recall that consensual individuals account for the opinions of all their neighbors, while expressers only account for the opinions of neighbors who also express (when such neighbors exist). Hence, there always exists at least one path starting at a node that represents a consensual individual and that ends at a node representing an expresser. However, there does not exist any paths that start at a node representing an expresser and that end at a node representing a consensual player. In particular, a set of individuals $\mathcal{C} \subset N$ is called an *essential class* (Seneta (1981) [89]) if there does not exist a path starting at an individual $i \in \mathcal{C}$ and ending at an individual $j \in N \setminus \mathcal{C}$.

Third, the multiplicity of the eigenvalue 1 is equal to the number of essential classes in the hearing matrix \tilde{G} . To see this simply, consider a circle as a network structure with exactly k individuals, where each individual has two neighbors and where initial opinions are such that each individual has at least one neighbor who is ideologically-opposed. For this network structure, given the expression threshold $\delta^* = 1$, all individuals choose to express. Since each individual has at least one ideologically-opposed neighbor, each individual reaches an extreme opinion of 1 or -1 after few periods of interaction. In this setting, individuals no longer take into account the opinions of other expressers in the long-run and each individual forms an essential class on their own. Hence, the hearing matrix \tilde{G} is simply the identity matrix of size k and the multiplicity of the eigenvalue 1 is exactly k . Beyond this example, the only case where an essential class is not a singleton is the case where there is a group of individuals that form a connected set of expressers (see definition 2.3) that are like-minded. In other words, there exists a path connecting each pair in this connected set of expressers at each time period of interaction, but no paths from any of those expressers to an individual outside this set. I summarize the above discussion in the following theorem and provide a proof which makes use of standard linear algebra results.

Theorem 2.1 *Given $\alpha_{t^*} \in [-1, 1]^n$ a vector of opinion at period t^* and a hearing matrix \tilde{G} associated with the network structure G , the long-run opinions are :*

$$\alpha_{\infty} = \left(\lim_{t \rightarrow \infty} \tilde{G}^t \right) \alpha_{t^*} = \mathcal{G} \mathbf{a}_{t^*} < \infty,$$

where \mathcal{G} is the spectral projector associated with the eigenvalue 1. Moreover, the algebraic multiplicity of the eigenvalue 1 is equal to the number of essential classes of the hearing matrix \tilde{G} .

Proof. See Appendix 2.7.4 ■

The columns corresponding to consensual individuals in the matrix \mathcal{G} are all zero, meaning that in the long-run the initial opinions of such individuals vanish. Their opinions remain hidden through out the periods of interaction. As for the columns corresponding to expressers, they have at least one strictly positive entry. In particular, the long-run opinions of consensual individuals are exactly convex combinations of initial opinions of expressers. In other words, the long-run opinion of consensual individuals is affected by the long-run opinions of all the expressers to whom they are connected to through a path of other consensual individuals. Hence, the total impact of the initial opinion of a given expresser $i \in N$ over long-run opinions can be assessed by considering the total weight an expresser has in the long-run opinions of other individuals. This motivates the introduction of the following statistic.

Definition 2.4 (Spectral influence) *Given a network structure G , a hearing matrix \tilde{G} and its limit \mathcal{G} , the spectral influence of individual $i \in N = \{1, \dots, n\}$ is:*

$$s_i = \frac{1}{n}(\mathcal{G}' \mathbf{1}_n)_i,$$

where $\mathbf{1}_n$ is a column vector of ones.

Example 7 *Consider network G in figure 2.1 and suppose that the initial opinions of expressers 1 and 2 are $\alpha_{0,1}$ and $\alpha_{0,2}$. Since both are not directly connected nor are they connected via a chain of expressers, each of them forms an essential class on their own. The spectral projector \mathcal{G} associated to the eigenvalue 1 of the hearing matrix \tilde{G} is a symmetric matrix of size 9 and is given by: $\mathcal{G}_{11} = \mathcal{G}_{i1} = 1$ for $i \in \{6, 7, 8\}$. That is the long-run opinion of individuals 6, 7 and 8 is fully determined by the initial opinion of individual 1. As for individual 2, $\mathcal{G}_{22} = \mathcal{G}_{j2} = 1$ for $j \in \{3, 4, 5\}$. The long-run opinion of individual 9 is equally determined by the initial opinions of expressers 1 and 2, that is $\mathcal{G}_{91} = \mathcal{G}_{92} = 1/2$. All the remaining entries of the matrix \mathcal{G} are zero. Hence, long-run opinions are $\alpha_{\infty,i} = \alpha_{0,1}$ for $i \in \{1, 6, 7, 8\}$, $\alpha_{\infty,j} = \alpha_{0,2}$ for $j \in \{2, 3, 4, 5\}$ and $\alpha_{\infty,9} = \frac{1}{2}\alpha_{0,1} + \frac{1}{2}\alpha_{0,2}$. Expressers 1 and 2 have an identical spectral influence equal to $s_1 = s_2 = 1/2$.*

2.4.3. Patterns of long-run opinions: consensus and bi-polarization

In this section, I characterize patterns of long-run opinions. I focus on two patterns. First, I study consensus as a benchmark, where all the individuals in the long-run become like-minded.

Definition 2.5 (Consensus) *Long-run opinions form consensus if $\forall i \neq j \in N, |\alpha_{\infty,i} - \alpha_{\infty,j}| < \tau$.*

The above definition of consensus reflects the idea that the limiting opinions need not to be identical but the pairwise difference needs to be at most τ . In other words, the matrix \mathcal{G}

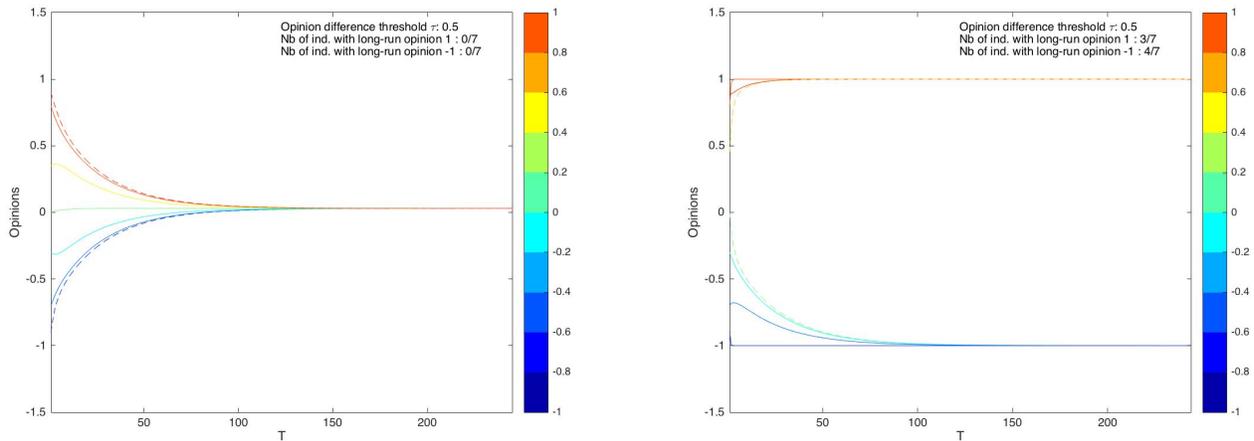


Figure 2.4: Consensus and bi-polarization

(see Theorem 2.1) in general will not have identical rows.⁷ Second I consider bi-polarization⁸ of long-run opinions, that is given a network structure and an initial opinion vector, the society of N individuals gets divided into two opinion groups of approximately equal size with strong agreement within each group and disagreement across both groups.

Definition 2.6 (Bi-polarization) Let $\epsilon > 0$ be a strictly positive real number. Long-run opinions are bi-polarized if (i) the society of N individuals is divided into two groups of size $|N_1|$ and $|N_2|$ such that $||N_1| - |N_2|| < \epsilon$, (ii) for $k \in \{1, 2\}, \forall i \neq j \in N_k, |\alpha_{\infty,i} - \alpha_{\infty,j}| < \tau$, (iii) $|\bar{\alpha}_{\infty,N_1} - \bar{\alpha}_{\infty,N_2}| \geq \tau$ where $\bar{\alpha}_{\infty,N_i}$ is the average long-run opinion of group N_i for $i \in \{1, 2\}$.

In general, the higher the number of expressers induced by a given network structure G and the more the long-run opinion pattern depends on the distribution of initial opinions within each expresser's neighborhood. To see this, consider a situation where most individuals can *express*, as it is the case in regular network structures. When each individual has like-minded neighbors, then the group converges to consensus; even if there exists a pair of expressers that are not directly connected and that hold ideologically-opposed views. For example, let G be a line network with 7 individuals such that $g_{12} = g_{23} = g_{34} = g_{45} = g_{56} = g_{67} = 1$ and the remaining entries are all zeros. Consider the following ini-

⁷The spectral projector has identical rows if and only if the algebraic multiplicity of the unit eigenvalue is 1; meaning that computing the perron vector is sufficient to obtain the long-run opinions. But having more than two sets of connected expressers is translated into an algebraic multiplicity of the unit eigenvalue strictly higher than 1.

⁸I focus on opinion bi-polarization rather than the more general case of opinion polarization, where a polarized society is one that is divided into a small number (larger than two) of opposed groups. The special case of bi-polarization fits applications of the model where it would take a very large group - e.g. at least half of the population of interest - to over-turn a policy or to elect a president or to produce a divided public opinion. For example, one can think of Brexit, the election of Trump, the implementation of a Carbon tax in France or even at the beginning of 2020 divided views about risk reducing measures regarding the coronavirus.

tial opinion vector $\alpha_0 = (-0.9, -0.7, -0.3, 0, 0.45, 0.8, 0.9)$. For $\tau = 0.5$, individuals reach consensus even though individuals 2 and 6 are ideologically-opposed, as shown in the left panel of figure 2.4. However, for the same network structure, if the initial opinion vector is $\alpha_0 = (0, -0.3, -0.7, -0.9, 0.8, 0.9, 0.45)$, then this group of individuals does not reach consensus, as shown in the right panel of figure 2.4. This is because individuals 3 and 4 repulse each other towards long-run opinions $a_{\infty,3} = -1$ and $a_{\infty,4} = 1$. In this case, long-run opinions become $\alpha_{\infty} = (-1, -1, -1, -1, 1, 1, 1)$ which corresponds to a bi-polarized group of individuals.

Consensus

Consensus of opinion depends on the number of expressers in the network, how they are connected (or not) to each other and the distribution of their initial opinions. Namely, consensus is reached whenever there is a unique expresser in the network or several expressers with like-minded neighbors who express, and for any two disjoint sets of connected expressers, individuals are in average like-minded across both sets. In the former case, the initial opinion of the unique expresser fully determines the long-run opinions of the remaining consensual individuals. While in the latter case, long-run opinions of individuals can be different but that difference is at most τ .

Proposition 2.3 *Given a network structure G , an initial opinion vector $\alpha_0 \in [-1, 1]^n$ and the set of expressers $E \subset N$, long-run opinions form a consensus if and only if exactly one of the following statements holds:*

- (i) $|E| = 1$
- (ii) *There exists $\kappa \geq 1$ sets of connected expressers s.t. $E = \bigcup_{k=1}^{\kappa} \mathcal{E}_k, \forall k_i \neq k_j \in \{1, \dots, \kappa\}, |\bar{\alpha}_{0, \mathcal{E}_{k_i}} - \bar{\alpha}_{0, \mathcal{E}_{k_j}}| < \tau$ and for each set $\mathcal{E}_{k_i} \subset E, \forall i \neq j \in \mathcal{E}_{k_i}$ and $j \in \mathcal{E}_{k_i} \cap N_i, |\alpha_{0,i} - \alpha_{0,j}| < \tau$.*

Proof. See Appendix 2.7.5 ■

Corollary 2.1 (unique opinion leader) *Let $\alpha_0 \in [-1, 1]^n$ be a vector of initial opinions. If the network structure G is a star where the central node has a degree $k > 1/\delta^*$ then the long-run opinions a_{∞} form a consensus.*

These necessary and sufficient conditions map long-run opinions to the network structure and the initial distribution of opinions. To see this, recall that choosing to express or hide one's opinion is determined by local popularity, which is a network statistic. Moreover,

in the case where the structure of the network of interpersonal relationships allows the existence of more than one expresser, the distribution of initial opinions along with the structure of connections among expressers become crucial in reaching consensus. In particular, consensus can never be reached if there exists a pair of ideologically-opposed expressers who are directly linked.

Bi-polarization

Unlike consensus, characterizing the initial opinion distribution of expressers within a network isn't sufficient to explain the emergence of a bi-modal long-run opinion distribution; that is, *a central interval that is sparsely populated and left and right intervals that are densely populated*, (Friedkin (2015) [49]). The characterization is challenging because the emergence of a bi-polarized society depends on where expressers are placed in the network, how they interact together and how many people they influence. If extreme opinion groups of expressers do form, the network structure together with the opinion difference threshold τ determine to which extreme opinion group, consensual individuals will belong.

Recall that the long-run opinion of each consensual individual is exactly a convex combination of the opinions of expressers to whom they are directly or indirectly linked by a path of other consensual individuals. For some network structures and expressers' opinion distributions, consensual individuals can receive influence from ideologically-opposed groups of expressers. In which case, consensual individuals remain moderate in the long-run. In figure 2.1, individual 11 holds a long-run opinion of *zero*, when individuals 1 and 2 hold respectively opinions 1 and -1 . In other words, the existence of long-run moderate individuals can block the split of the population into two ideologically-opposed groups.

Intuitively, the network position of moderate consensual individuals allows them to receive influence from different opinion groups. They are initially linked to both influence groups and they don't express an opinion themselves. Such individuals could be interpreted as intermediaries or neutral TV hosts or moderators, who act as buffers against opinion polarization.

Nevertheless, the absence of moderate consensual individuals isn't sufficient for long-run opinions to be polarized. In particular, one needs to account for the size of both opinion groups. The split of expressers into two extreme opinion groups isn't sufficient to cause opinion bi-polarization. Both extreme opinion groups need to have influence over an equally large number of consensual individuals, so that the society becomes divided. In other words, for consensual individuals to belong to one of both extreme opinion groups, they need to receive *enough* influence from the members of one group so that they hold similar extreme opinions in the long-run. Finally, for two extreme opinion groups to form,

at least one pair of ideologically-opposed expressers need to interact together or there must exist at least two individuals initially at each end of the opinion spectrum. Recall that the influence of each expresser $i \in E \subset N$ is summarized by their spectral influence s_i in definition 2.4. Define the set of expressers that hold long-run opinion 1 and -1 respectively by: $E^+ = \{i \in E : \alpha_{i,\infty} = 1\}$ and $E^- = \{i \in E : \alpha_{i,\infty} = -1\}$.

Lemma 2.2 *Suppose that $E = E^+ \cup E^- \subset N$, $E^+ \neq \emptyset$ and $E^- \neq \emptyset$. If long-run opinions are bi-polarized then*

- (i) $|\sum_{i \in E^+} s_i - \sum_{i \in E^-} s_i| < \epsilon$,
- (ii) $\exists k \in N$ s.t. $1 - \tau/2 \leq \sum_{i \in E^+} \mathcal{G}_{ki} \leq \tau/2$.

Proof. See Appendix 2.7.6. ■

2.5. Network topology and opinion patterns

In this section I explore the model through simulations. The objective is to relate the topology of the network to long-run polarization of opinions. Polarization of opinions is measured simply by looking at the variance in final opinions. Clearly, the maximum level of polarization is 1 and the minimum level is 0. To focus on the network topology, I generate a large number of initial opinions vectors (1000) distributed uniformly at random over $[-1, 1]$ for the same network structure with n individuals, then I compute the average level of polarization of the (1000) final opinion vectors obtained.

Formally, I study the evolution of opinions of a large set of individuals in scale-free networks, as it is a general family of networks where the topology allows for the co-existence of expressers and consensual individuals. I provide a number of (inter-related) network statistics that measure inequality in the degree distribution such as assortativity, the Gini coefficient of the degree distribution and per capita average degree. I also look at network statistics that measure connectivity such as neighborhood connectivity and average path length. I also look at whether expressers are densely or sparsely connected⁹, by identifying the number of connected sets of expressers in Definition 2.3. The description of the network statistics is in appendix 2.7.7.

The degree distribution within scale-free networks follows a power law. Due to this inequality in the degree distribution, expressers and consensual individuals co-exist. Barabási

⁹Density or sparsity of connections among expressers means whether they are isolated or interact with other expressers. In a graph with k expressers that have no expressing neighbors, the subgraph of G restricted to expressers will have exactly k components. One can think of some δ^* -core and count the number of components to account for sparsity or density.

and Albert (1999) [13] explain the emergence of such networks by the following two mechanisms: the network expands because new individuals (vertices) keep getting added to an existing network and they attach preferentially to individuals who are already well connected. The mechanism leading to the formation of links is out of the scope of this paper. The objective of the simulations is to make obvious the impact of the interaction structure - location in the network and initial opinion - of expressers on the long-run opinion pattern of the whole group of individuals. To that end, I generate scale-free networks and work with them as a snapshot at a given point in time where no new individuals are added and no new links are created. Each network topology is generated by providing an initial number of hubs h and a number of nodes n_c a newly added node connects to.

I start by providing two examples with high and low average polarization levels with 100 individuals, $h = 5$ initial hubs and $n_c = 1$. Then I give aggregate statistics by varying the number of hubs h and the number of connections n_c of each newly added nodes. All figures appear after the conclusion.

2.5.1. Example

I select two structures with the same number of hubs to illustrate the results of the model before moving to aggregate statistics. The first network structure G_1 contains 14 expressers, represented by squares in figure 2.7. Out of those 14 expressers, 11 are linked through a path of expressers, highlighted in red. In other words, the network topology allows for interaction between the majority of expressers. There are a number of empirical papers that support the idea that ideologically-opposed individuals interact together often (e.g. Conover et al. (2011) [39]). The first panel of figure 2.7 displays the initial opinions as colors given by the $[-1, 1]$ colormap on the east side of the figure. The second panel displays the long-run opinions. The third panel shows the evolution of opinions, where dotted lines correspond to the evolution of opinions of consensual individuals and the solid lines correspond to the evolution of opinions of expressers.

The second structure G_2 contains 11 expressers and only two pairs of expressers are linked as shown in figure 2.8. I select those two network structures by generating for the same initial opinions, a large number of scale-free networks with $n = 100$, $h = 5$ and $n_c = 1$. Then I pick out two network structures with a high and low average level polarization. The average level of polarization in network G_1 over the 1000 runs is 0.84, as opposed to only 0.43 for the network structure G_2 . Figures 2.5 and 2.6 show the distribution over the 1000 runs of the size of the group of individuals holding extreme opinion -1 , the group of individuals holding extreme opinion 1 and the group of individuals who hold a moderate opinion. For network G_1 the average size of extreme groups is almost 50%, while in network G_2 this same

share is occupied by the group of moderate individuals. To plot the distribution of the size of the opinion groups, I compute for each of the 1000 runs, the number of individuals within each of the three groups. Then I use this output to plot the histogram and the kernel density.

2.5.2. Agregate statistics

In order to study the impact of a specific topology of scale-free networks on the mean level of polarization, I vary the number of initial hubs h and the number of nodes n_c to which a newly node is added. Doing so allows me to look at network topologies with different levels of connectivity and a wide range of degree distributions. With $n = 200$ individuals, $h \in \{10, 20, 30, 40, 50, 60, 70\}$ and $c_n \in \{1, 2, \dots, 8\}$ I obtain 56 scale-free network topologies¹⁰ and I compute for each the mean level of polarization over 1000 runs with the same 1000 initial opinion vectors. I do so to get information about the topology of the network, that are not related to a specific distribution of a given initial opinion vector. Then I compare across those networks the average level of polarization and provide several network statistics. The network statistics could be divided in two groups: (i) (inter-related) measures of degree inequality (ii) measures of connectivity. Furthermore, I provide the α parameter based on a maximum likelihood estimator, which is the exponent of the power-law that fits best the degree distribution of the considered scale-free network topology. In figures 2.9 to 2.13 the markers correspond to different sizes of initial hubs h and colors correspond to the number n_c of connections a newly added node will have. For all figures, the y -axis corresponds to the level of mean polarization while the x -axis gives the value of the different network statistic that will be considered.

Opinion polarization is expected to be observed for more skewed degree distribution where expressers have a very high degree. Following Newman (2002) [82], *a network is said to show assortative mixing if the nodes in the network that have many connections tend to be connected to other nodes with many connections*. Unsurprisingly, figure 2.9 shows that the mean level of polarization is positively correlated with the α parameter and negatively correlated with the assortativity coefficient of the degree distribution. In particular, in the right panel of figure 2.9 the two topologies with $c_n = 1$ and respectively with $h = 20$ and $h = 30$ show a high level of polarization and disassortative mixing. Furthermore, the right panel of figure 2.11 and left panel of figure 2.12 indicate high levels of polarization for high values of the Gini coefficient of the degree distribution and very low network sparsity computed as the per capita mean degree.

The more expressers interact together, the greater should be the possibilities of disagree-

¹⁰Clearly with exactly the same number of hubs and the same n_c one obtains for different runs different network structures. But the variance in the estimated α parameter is very small, meaning that they display similar degree distributions and levels of connectivities.

ment and consequently opinion polarization. Surprisingly, the left panel of figure 2.10 shows that the lowest levels of polarization are observed for topologies where the number of sets of connected expressers is smaller than 2. This means that many expressers are connected among each other, yet polarization is low. The right panel of both figures 2.10 and 2.11 bring an explanation. For those same network topologies, the average path length and the diameter are very small (respectively smaller than 3 and smaller than 5). Hence, those network topologies with a low level of polarization display high connectivity. Thereby, expressers interact among each other and can fall into disagreement, but consensual individuals are exposed to more influence sources. Intuitively, for those network topologies, expressers don't have a *fan base* which is solely devoted to each of them. This explanation is also supported by figure 2.12 and the right panel of figure 2.13. In the right panel of figure 2.13 low levels of polarization are related to higher neighborhood connectivity. Yet for these high levels of neighborhood connectivity (pink, purple and blue markers), the number of expressers is dispersed over the whole range as shown in figure 2.12.

2.6. Conclusion and the way forward

This paper introduces two novel ingredients to classic opinion formation models with assimilative influence. It relates the opinion updating rule to the network topology, by introducing two types of individuals who either update their opinions à la DeGroot, or update their opinions using a law of motion which incorporates assimilation or distancing. The chosen rule for opinion updating depends on whether the individual is locally popular or not. I show that opinion patterns can be explained by focusing on how influential individuals interact among each other and how they influence the (less popular) masses of users. In particular, individuals who are not popular enough to express but are connected to two extreme opinion groups can obstruct full opinion bi-polarization. Those individuals are neutral TV hosts or journalists. By means of simulations, I show that when popular individuals are densely connected among each other but the overall level of connectivity in the network is relatively low, then the average level of polarization is high. However when *all* individuals are densely connected, consensual individuals have access to more influence sources and the average polarization level is low. This present model can be extended in several ways. The choice of expression can depend on the opinion itself. Namely, even if an individual is popular enough, they may not express when they hold an opinion that is too far from the average of opinions within their social circle (including both popular and unpopular individuals). Another promising direction for a follow up paper, is to endogenize the network structure and allow agents who *hide* to be able to *express* after a certain number of periods depending on the evolution of opinions in their neighborhood.

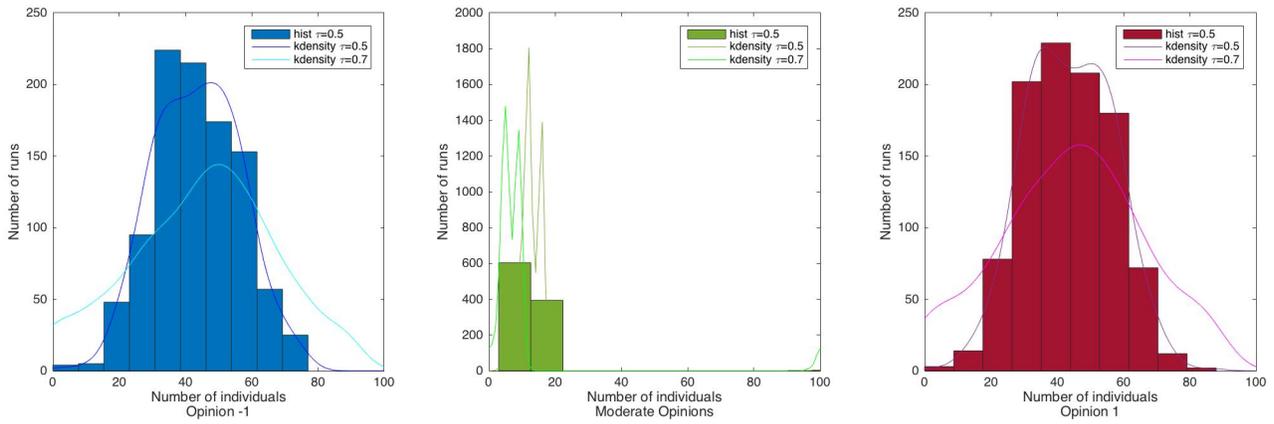


Figure 2.5: Kernel density over 1000 runs of groups of individuals with extreme opinion -1 (blue), 1 (red) and moderate (green) in scale-free network G_1

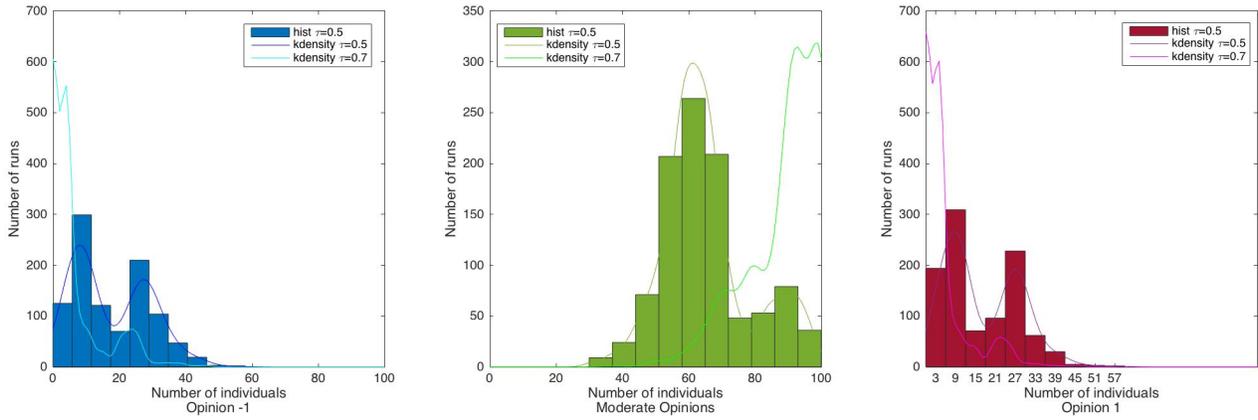


Figure 2.6: Kernel density over 1000 runs groups of individuals with extreme opinion -1 (blue), 1 (red) and moderate (green) in scale-free network G_2

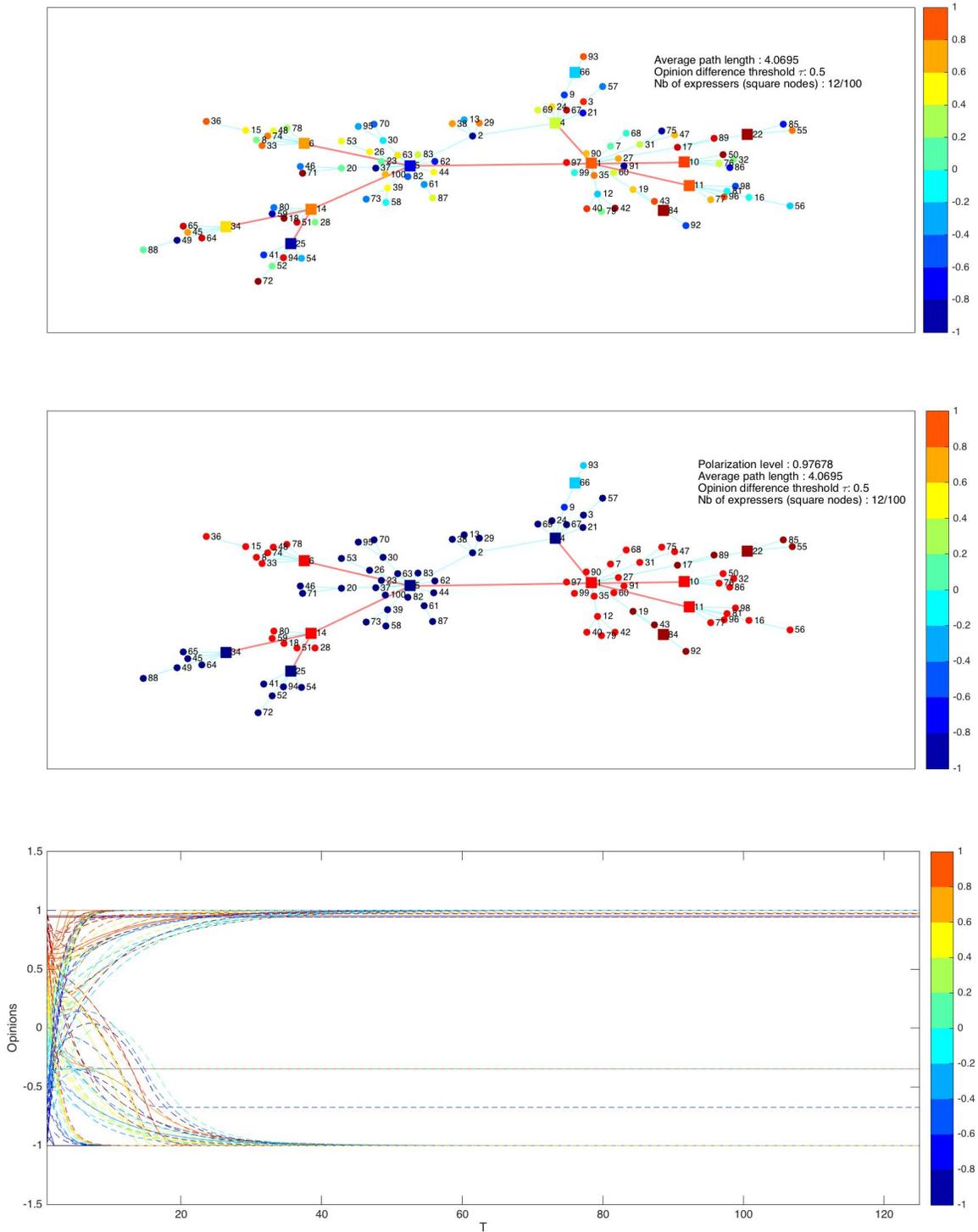


Figure 2.7: Initial opinions, final opinions and evolution of opinions in scale-free network G_1 . Expressers are represented with a square marker. Links between expressers are high-lighted in red.

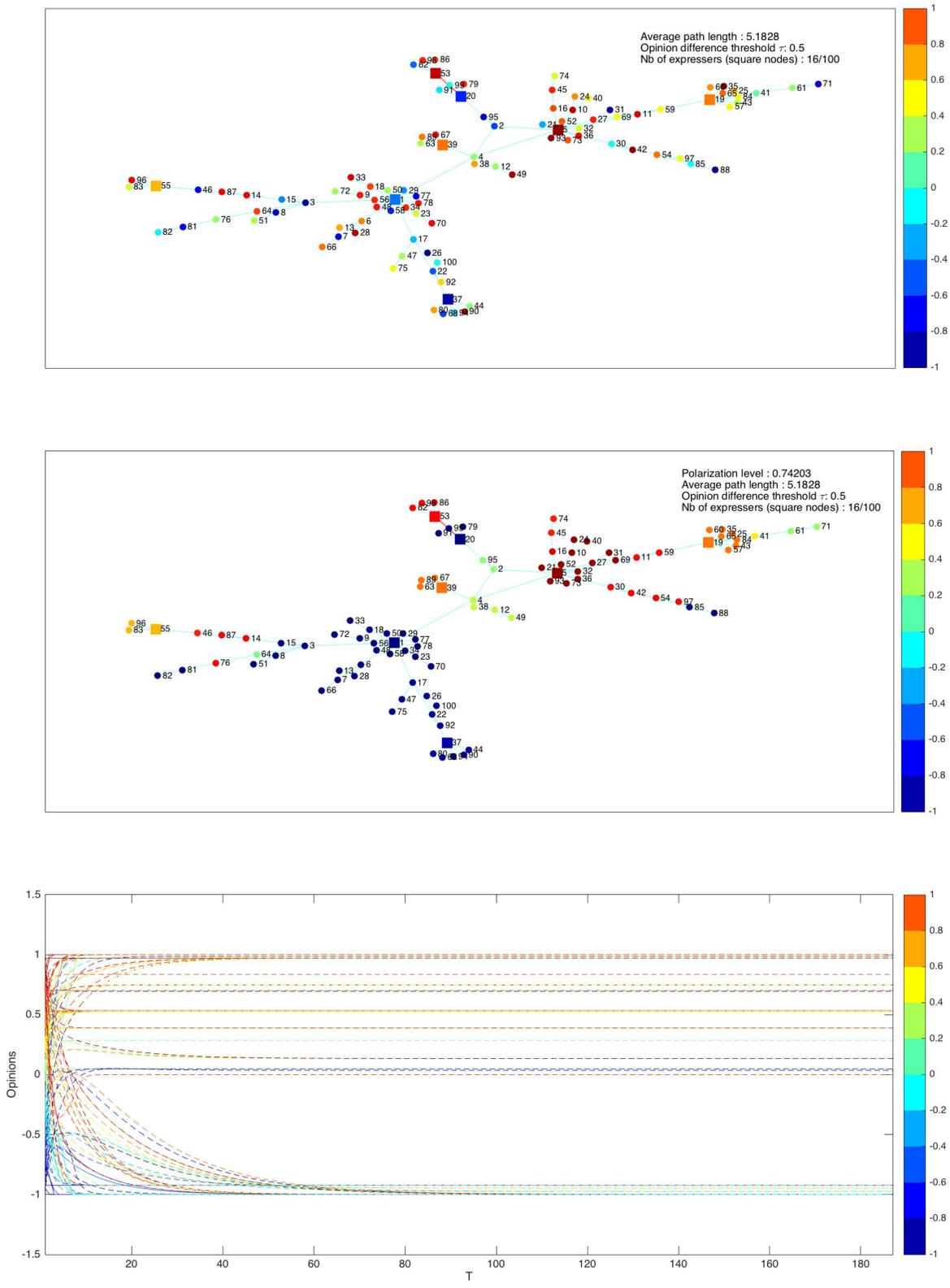


Figure 2.8: Initial and final opinions in the run with the highest level of polarization in scale-free network G_2 . Links between expressers are highlighted in red.

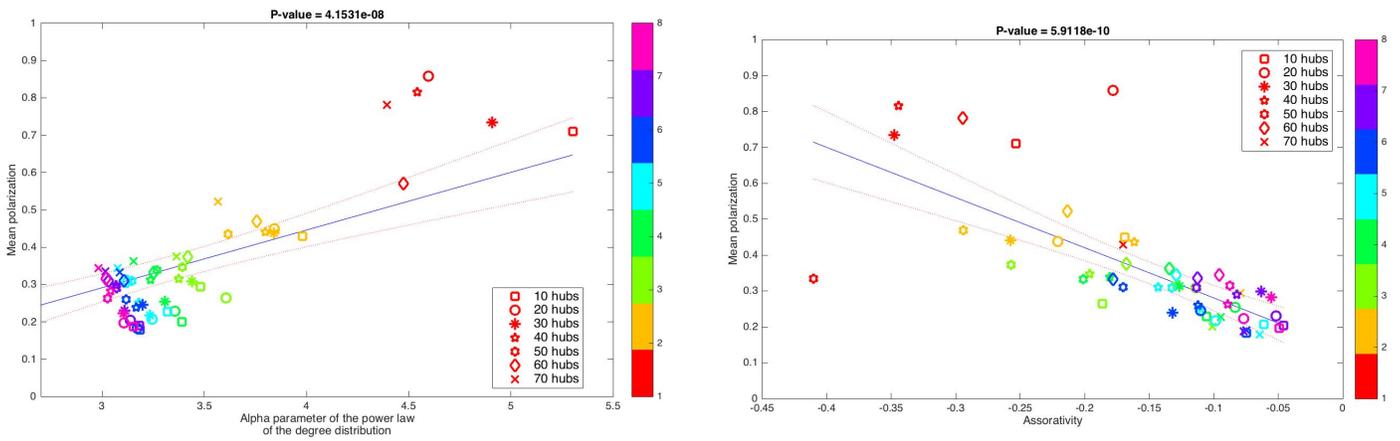


Figure 2.9: Markers correspond to different sizes of initial hubs h and colors correspond to the number n_c of connections a newly added node will have.

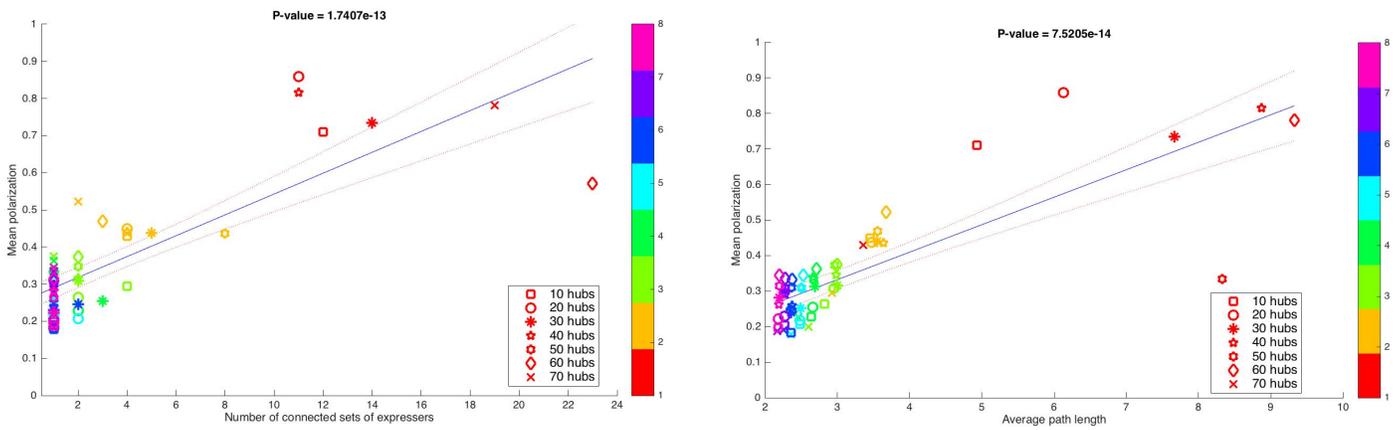


Figure 2.10

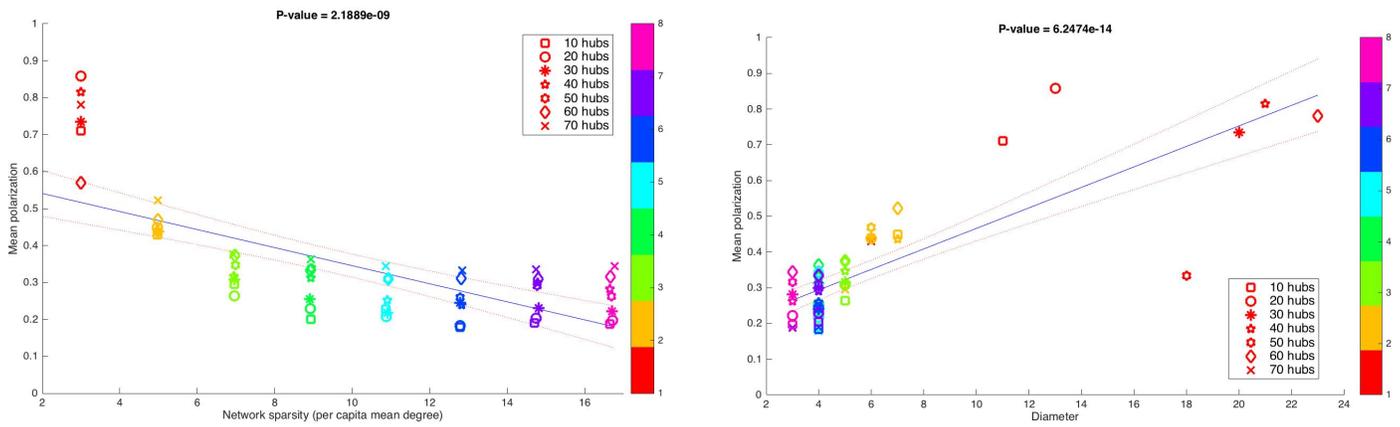


Figure 2.11

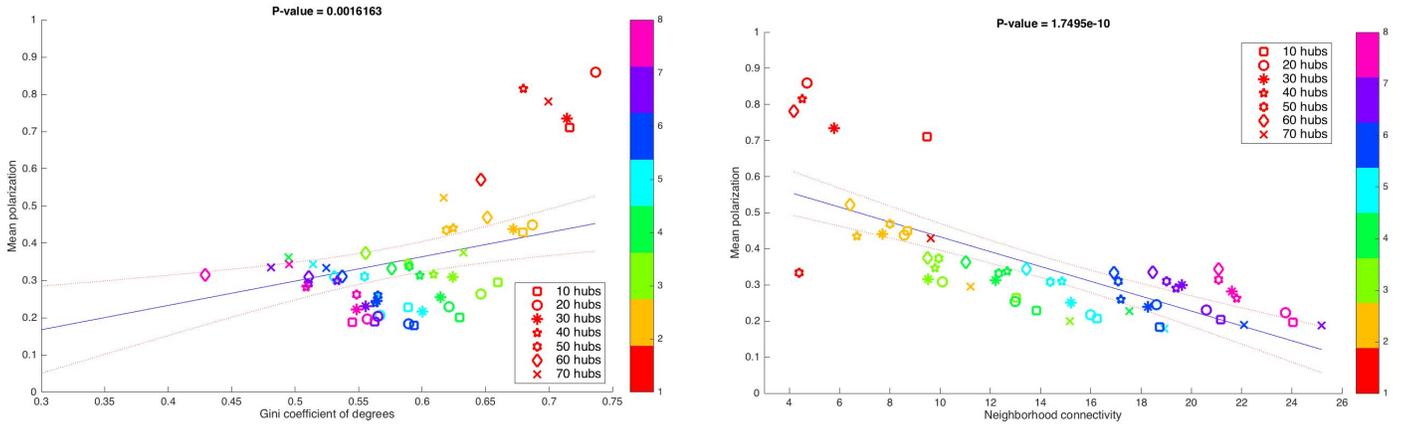


Figure 2.12

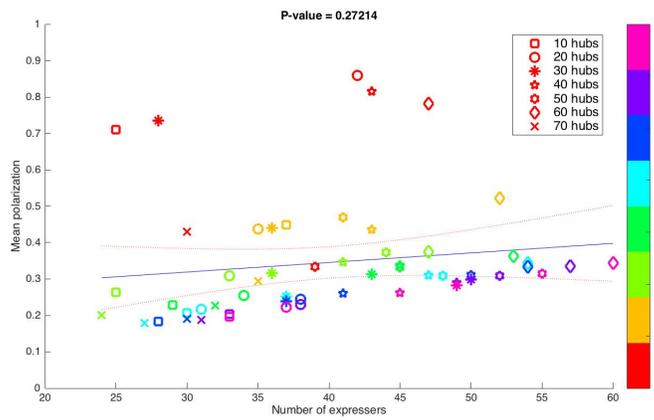


Figure 2.13

2.7. Appendix

2.7.1. Proof of Proposition 2.1

(i) When $|\mathcal{E}| = 1$ it means that individual $i \in \mathcal{E}$ has no direct neighbors who choose to express, hence individual i never updates their initial opinion and their long-run opinion is exactly their initial opinion $\alpha_{0,i}$.

(ii) Suppose without loss of generality that $|\mathcal{E}| = \kappa$. If $|\mathcal{E}| = \kappa > 1$ and all individuals $i \in \mathcal{E}$ are like-minded, that is $\forall i \neq j \in \mathcal{E}, |\alpha_{0,i} - \alpha_{0,j}| < \tau$ then for $\mu \in (0, 1/\kappa)$ the opinions get updated in the following way:

$$\left\{ \begin{array}{l} \alpha_{1,1} = \alpha_{0,1} + \mu \sum_{j \neq 1 \in \mathcal{E}} g_{1j}(\alpha_{0,j} - \alpha_{0,1}), \\ \vdots \\ \alpha_{1,\kappa} = \alpha_{0,\kappa} + \mu \sum_{j \neq \kappa \in \mathcal{E}} g_{\kappa j}(\alpha_{0,j} - \alpha_{0,1}), \end{array} \right. \Leftrightarrow \left\{ \begin{array}{l} \alpha_{1,1} = (1 - \mu d_1(\mathcal{E}))\alpha_{0,1} + \mu \sum_{j \neq 1 \in \mathcal{E}} g_{1j}\alpha_{0,j}, \\ \vdots \\ \alpha_{1,\kappa} = (1 - \mu d_\kappa(\mathcal{E}))\alpha_{0,\kappa} + \mu \sum_{j \neq \kappa \in \mathcal{E}} g_{\kappa j}\alpha_{0,j}, \end{array} \right.$$

where $d_i(\mathcal{E}) = \sum_{j \in \mathcal{E}} g_{ij}$ corresponds to the number of expressers that are in the set of connected expressers \mathcal{E} and are also direct neighbors of individual $i \in \mathcal{E}$. Writing the above system in matrix notation and using induction we get the following relation :

$$\alpha_{t,\mathcal{E}} = M^t \alpha_{0,\mathcal{E}},$$

where $\alpha_{t,\mathcal{E}} = (\alpha_{t,1}, \dots, \alpha_{t,\kappa})^T$, $\alpha_{0,\mathcal{E}} = (\alpha_{0,1}, \dots, \alpha_{0,\kappa})^T$ and M an $\kappa \times \kappa$ symmetric matrix with diagonal entries $m_{ii} = 1 - d_i(\mathcal{E})\mu$ and off diagonal entries $m_{ij} = \mu g_{ij}$, for $j \neq i \in \mathcal{E}$. Hence, M is a symmetric matrix, with non-negative entries and whose columns and rows sum to one. In order to get the long-run opinions we need to compute $\lim_{t \rightarrow \infty} \alpha_{t,\mathcal{E}} = \lim_{t \rightarrow \infty} M^t \alpha_{0,\mathcal{E}}$.

Claim 1 $\lim_{t \rightarrow \infty} M^t$ exists.

This limit exists because all the eigenvalues of the matrix M are smaller or equal to 1. To see this, simply recall that by the Gershgorin Circle Theorem (1931), the eigenvalues of the square matrix M belong to the union of its Gershgorin disks. In the case of the matrix M the Gershgorin disks¹¹ write for each $i \in \mathcal{E}$, $D_i = \{x \in \mathbb{R} : |x - m_{ii}| \leq \sum_{j \neq i} |m_{i,j}|\} = \{x \in \mathbb{R} : |x - (1 - d_i(\mathcal{E})\mu)| \leq d_i(\mathcal{E})\mu\}$. Hence, the upper bound of the eigenvalues of M is given

¹¹All the eigenvalues of M are real because M is a real symmetric matrix.

exactly by $\max_{i \in \mathcal{E}} (1 - d_i(\mathcal{E})\mu) + d_i(\mathcal{E})\mu = 1$. Now I will show that $\lim_{t \rightarrow \infty} \alpha_{t,\mathcal{E}}$ is exactly the average of the initial opinions of individuals $1, \dots, \kappa \in \mathcal{E}$.

Claim 2 Let $\mathbf{1}_{p,q}$ be a matrix of ones of size $p \times q$. $\lim_{t \rightarrow \infty} M^t = \frac{1}{\kappa} \mathbf{1}_{\kappa,1} \mathbf{1}_{1,\kappa}$.

Intuitively, since at each time period every updated opinion of an expresser is a convex combination of the opinions of like-minded neighbors who also express, the long-run opinions converge to the average of initial opinions of the members of the connected set of expressers. Formally, I use theorem 1 in Xiao and Boyd (2004) [99], which states that $\lim_{t \rightarrow \infty} M^t = \frac{1}{\kappa} \mathbf{1}_{\kappa,1} \mathbf{1}_{1,\kappa}$ if and only if (i) the vector $\mathbf{1}$ is a left eigenvector of M associated with the eigenvalue one, (ii) the vector $\mathbf{1}$ is a right eigenvector of M associated with the eigenvalue one, (iii) one is a simple eigenvalue of M . Conditions (i) and (ii) hold for the matrix M because it is symmetric and row stochastic. To see this, one can simply sum the entries over a given row $i \in \mathcal{E}$: $m_{ii} + \sum_{j \neq i \in \mathcal{E}} m_{ij} = 1 - d_i(\mathcal{E})\mu + \sum_{j \neq i \in \mathcal{E}} g_{ij}\mu = 1 - d_i(\mathcal{E})\mu + d_i(\mathcal{E})\mu = 1$. Since the matrix M is symmetric, it is also column stochastic and the vector one is a left and right eigenvector of the matrix M associated with the eigenvalue one. Finally, condition (iii) holds because the matrix M is irreducible with non-negative entries; because the set of individuals in \mathcal{E} is connected and they are all like-minded, in the sense of definition 2.3. Hence the eigenvalue 1 is simple (Perron-Frobenius Theorem).

2.7.2. Proof of Proposition 2.2

Warm-up for the proof and notations. Recall that $\underline{N}_{i,t-1} = \{j \in N_i \cap E \text{ s.t. } |\alpha_{i,t-1} - \alpha_{j,t-1}| < \tau\}$ is the set of expressing neighbors of individual $i \in N$ such that their opinion difference is smaller than τ (like-minded). Similarly, recall that $\overline{N}_{i,t-1} = \{j \in N_i \cap E \text{ s.t. } |\alpha_{i,t-1} - \alpha_{j,t-1}| \geq \tau\}$. Clearly it follows that $\underline{N}_{i,t-1} \cup \overline{N}_{i,t-1} = N_i \cap E$. Without loss of generality, suppose that individual $i \in N$ chooses to express and belongs to the set of connected expressers $\mathcal{E} \subseteq E$ such that $|\mathcal{E}| > 1$ and there exists at least one pair of expressers in \mathcal{E} who are neighbors and initially ideologically-opposed. Each individual $i \in \mathcal{E}$ updates their opinion at each time step according to the law of motion (2.2), given by :

$$\alpha_{i,t} = \alpha_{i,t-1} + \mu \sum_{j \in \underline{N}_{i,t-1}} (\alpha_{j,t-1} - \alpha_{i,t-1}) + \mu \sum_{j \in \overline{N}_{i,t-1}} (\alpha_{i,t-1} - \alpha_{j,t-1}) \quad \text{s.t } \alpha_{i,t} \in [-1, 1] \quad (2.6)$$

$$\Leftrightarrow \alpha_{i,t} = \alpha_{i,t-1} (1 + \mu(|\overline{N}_{i,t-1}| - |\underline{N}_{i,t-1}|)) + \mu \left(\sum_{j \in \underline{N}_{i,t-1}} \alpha_{j,t-1} - \sum_{j \in \overline{N}_{i,t-1}} \alpha_{j,t-1} \right) \quad \text{s.t } \alpha_{i,t} \in [-1, 1]. \quad (2.7)$$

The size of the sets $\overline{N}_{i,t-1}$ and $\underline{N}_{i,t-1}$ can vary between two periods because a like-minded

neighbor at period $t - 1$ can become ideologically-opposed at a subsequent period (see the examples in section 2.4.1). In other words, it's impossible to summarize the above system of equations by one time-invariant matrix because the entries or weights vary with time depending on the opinion differences between connected expressers and whether the opinion of a given individual has reached the upper or lower bound. Hence we need to account for the *flow* of influence through *chains* of expressers within a given set \mathcal{E} . To do so, it's convenient to write the opinion of a given individual i at period t belonging to the connected set of expressers \mathcal{E} as an **affine** combination of the opinions their neighbors j at period $t - 1$:

$$\alpha_{i,t} = \sum_{j \in N_i \cap \mathcal{E}} a_{ij}(t-1) \alpha_{j,t-1}, \quad (2.8)$$

where

$$a_{ij}(t-1) = \begin{cases} \mu & \text{if } i \neq j, |\alpha_{i,t-1} - \alpha_{j,t-1}| < \tau \text{ and } \alpha_{i,t-1} \in (-1, 1) \\ -\mu & \text{if } i \neq j, |\alpha_{i,t-1} - \alpha_{j,t-1}| \geq \tau \text{ and } \alpha_{i,t-1} \in (-1, 1) \\ 1 + \mu(|\overline{N}_{i,t-1}| - |\underline{N}_{i,t-1}|) & \text{if } i = j \text{ and } \alpha_{i,t-1} \in (-1, 1) \\ 0 & \text{if } i \neq j \text{ and } \alpha_{i,t-1} \in \{1 + \epsilon, -1 - \epsilon\} \\ 1 & \text{if } i = j \text{ and } \alpha_{i,t-1} \in \{1 + \epsilon, -1 - \epsilon\} \end{cases}$$

Let $A(t-1)$ be the matrix with entries $a_{ij}(t-1)$ for $i, j \in \mathcal{E}$. It follows that the opinions at period t of the expressers who belong to \mathcal{E} can be written as:

$$\alpha_{\mathcal{E},t} = A(t-1)A(t-2) \dots A(0)\alpha_{\mathcal{E},0} = B(t-1,0)\alpha_{\mathcal{E},0},$$

where $B(t-1,0)$ is the matrix product of $A(t-1)A(t-2) \dots A(0)$. In other words it's a matrix that keeps track of the accumulated (positive and negative) weights between periods $t - 1$ and 0. In particular, the entry $B_{ij}(t, t-1)$ reports the influence of j on i 's opinion between periods t and $t - 1$. Recall that the set of expressers that have at least one ideologically-opposed neighbor is given by:

$$IO(\mathcal{E}) = \{i \in \mathcal{E} : \exists j \in \mathcal{E}, i \in \overline{N}_{j,0}, j \in \overline{N}_{i,0}\}.$$

Part (i). Show that if i is connected indirectly to expressers with ideologically-opposed neighbors through a path of like-minded neighbors then $\alpha_{i,t}$ can be written as a convex combination of the opinions of neighbors for all $t \geq 1$. In this case individual i holds in the long-run a moderate opinion $\alpha_{i,\infty} \in (-1, 1)$. Otherwise, their opinion keeps getting pushed to the upper or lower bound of the opinion interval and necessarily $\alpha_{i,\infty} \in \{-1, 1\}$.

Claim 3 Let $i_1, i_2 \dots i_k \in IO(\mathcal{E})$. If there exists at each $t \geq 0$ at least two paths of expressers with only like-minded neighbors connecting $i \notin IO(\mathcal{E})$ to at least two elements in $\{i_1, i_2 \dots i_k\}$ then $\alpha_{i,\infty} \in \text{conv}(\alpha_{i_1,\infty}, \dots, \alpha_{i_k,\infty})$.

Proof. Suppose that $IO(\mathcal{E}) \neq \emptyset$, that is there exists at least one pair of neighbors in \mathcal{E} that are initially ideologically-opposed. Let s be the time period such that for all $i_k \in IO(\mathcal{E})$, $a_{i_k j}(s) = 0$ for all $j \neq i_k$, $a_{i_k j}(s) = 1$ for $i_k = j$.

(\Rightarrow) Suppose that there exists at each $t \geq 0$ paths of expressers with only like-minded neighbors connecting $i \notin IO(\mathcal{E})$ to at least two individuals $i_k \neq i_j \in IO(\mathcal{E})$. That is there exists $i, j_1, j_2 \dots, j_k \notin IO(\mathcal{E})$ such that for all $t \geq s$, $B_{i,i_k}(t, s) > 0$ and $B_{i,i_j}(t, s) > 0$. Since,

1. the opinion of a given individual $j_k \in \{j_1, j_2, \dots\} \cup \{i\} \notin IO(\mathcal{E})$ at period $t \geq s$ is a **convex** combination of the opinions of their neighbors at period $t - 1$ (because they all have like-minded neighbors):

$$\alpha_{j_k,t} = \sum_{k \in N_i \cap \mathcal{E}} a_{j_k k}(t-1) \alpha_{k,t-1},$$

where $a_{j_k k}$ takes the value μ for any $k \neq j_k$ and $1 - |N_{j_k,t}| \mu$ for $k = j_k$.

2. and at time period $t \geq s$, $\forall i_k \in IO(\mathcal{E})$, $\alpha_{i_k,t} = \alpha_{i_k,s} \in \{-1, 1\}$,
3. it follows that $\forall j_k \in \{j_1, j_2, \dots\} \cup \{i\} \notin IO(\mathcal{E})$, $\alpha_{j_k,t} \in \text{conv}(\alpha_{i_1,s}, \dots, \alpha_{i_k,s})$.
4. Moreover, since $IO(\mathcal{E}) \neq \emptyset$ and $\forall j_k \in \{j_1, j_2, \dots\} \cup \{i\} \notin IO(\mathcal{E})$ are connected (indirectly) to at least two individuals in $IO(\mathcal{E})$ then $\alpha_{j_k,t} \in \text{conv}(\alpha_{i_1,s}, \dots, \alpha_{i_k,s})$ and $\alpha_{i,t} \notin \{-1, 1\}$.
5. $\forall t \geq 0$, $\forall j_k \in \{j_1, j_2, \dots\} \cup \{i\} \notin IO(\mathcal{E})$ have like-minded neighbors at each period, hence the argument extends to the limit.

■

In figure 2.14, I provide an example that gives an intuition for the previous proof. I set δ^* such that all individuals can express for the sake of the example and expressers are hence represented by a square. The colors of each node correspond to their opinion at period T indicated by the color map over $[-1, 1]$ on the east side of each figure. The set of connected expressers is $\mathcal{E} = E = N$ and $IO(\mathcal{E}) = \{1, 9, 10\}$. Individuals $\{2, \dots, 8\}$ all have like-minded neighbors, as indicated by the colors in the left panel of figure 2.14 and they

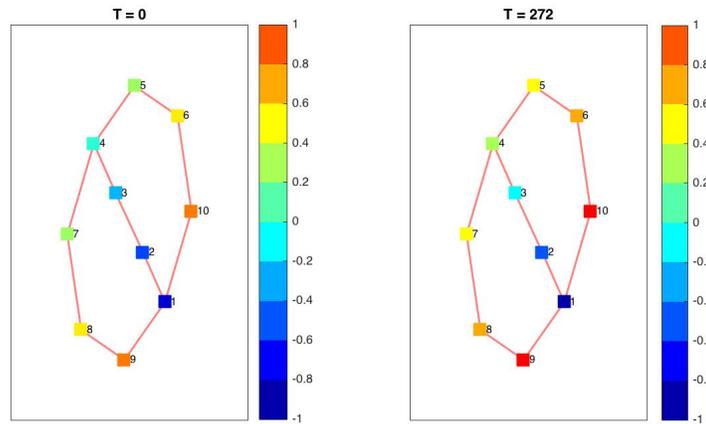


Figure 2.14

are all connected to the expressers in $IO(\mathcal{E})$ by three paths of like-minded expressers. The right panel indicates the long-run opinions and we see that individuals $\{2, \dots, 8\}$ still have like-minded neighbors. In particular, their long-run opinion is a convex combination of the opinions of the individuals in $IO(\mathcal{E}) = \{1, 9, 10\}$. Individual 4 is exactly at distance 3 from each of the three individuals in $IO(\mathcal{E}) = \{1, 9, 10\}$ and their long-run opinion is $\frac{1}{3}\alpha_{1,\infty} + \frac{1}{3}\alpha_{9,\infty} + \frac{1}{3}\alpha_{10,\infty} = \frac{1}{3}(-1) + \frac{1}{3}(1) + \frac{1}{3}(1) = \frac{1}{3}$. Individuals in $IO(\mathcal{E}) = \{1, 9, 10\}$ reach the upper and lower bound of the opinion period just after one period of interaction.

Part (ii). Show that when a given individual i is initially neutral (opinion zero) with no like-minded friends, if she is repulsed by two neighbors who are themselves ideologically-opposed and the negative influence she receives at each period is exactly of the same magnitude, then she remains neutral. The simplest example is a circle with three individuals 1,2,3 where $\alpha_{1,0} = 0$, $\alpha_{2,0} = -0.8$ and $\alpha_{3,0} = 0.8$. Looking at expression (2.7), the opinions get updated in period $t = 1$ as follows:

$$\begin{aligned}\alpha_{1,1} &= 0 + \mu(-(-0.8) - (0.8)) = 0 \\ \alpha_{2,1} &= -0.8 + \mu(-0 - (0.8)) = -0.8 - \mu 0.8 \\ \alpha_{3,1} &= 0.8 + \mu(-0 - (-0.8)) = 0.8 + \mu 0.8\end{aligned}$$

In subsequent periods, the opinion of individual 1 remains zero because $\mu(-\alpha_{2,t} - \alpha_{3,t}) = 0$ for all $t \geq 0$. However, if $\alpha_{2,0} \neq -\alpha_{3,0}$ then the opinion of individual 1 will be pushed to the upper or lower bound after a certain number of periods depending on which of $|\alpha_{2,0}|$ and $|\alpha_{3,0}|$ is larger.

Claim 4 Let $i_1, i_2 \dots i_k \in IO(\mathcal{E})$. If there exists $i \in IO(\mathcal{E})$ such that (i) $\alpha_{i,0} = 0$, (ii) $\forall t \geq 0$

$\underline{N}_{i,t} = \emptyset$, (iii) and $\sum_{j \in \overline{N}_{i,t}} \alpha_{j,t} = 0$ then $\alpha_{i,\infty} = 0$.

Proof. Let $i \in IO(\mathcal{E})$ such that $\alpha_{i,0} = 0$ and $\forall t \geq 0$, $\underline{N}_{i,t} = \emptyset$. Since individual i has no like-minded neighbors, $a_{i,k}(t) = -\mu$ for all $k \in \overline{N}_{i,0} = N_i$ and $\forall t \geq 0$. Their opinion gets updated as follows:

$$\alpha_{i,1} = 0 - \mu \sum_{j \in \overline{N}_{i,0}} \alpha_{j,0}$$

The opinion of i is zero at period 1 if and only if $\sum_{j \in \overline{N}_{i,0}} \alpha_{j,0} = 0$. Since, all the neighbors are in $\overline{N}_{i,0}$ and $\alpha_{i,0} = 0$ then even for an opinion threshold $\tau > \epsilon$, $\alpha_{j,0} \neq 0$. Hence, $\sum_{j \in \overline{N}_{i,0}} \alpha_{j,0} = 0$ if and only if the opinions $\alpha_{j,0}$ for all $j \in \overline{N}_{i,0}$ cancel out. By induction, the argument holds for all time periods.

Notice that if at any given period t this sum doesn't cancel out then the opinion of i will be different than zero in all subsequent period until it reaches the upper or lower bound.

Part (iii). Otherwise if (i) and (ii) do not hold and $IO(\mathcal{E}) \neq \emptyset$, then $\forall i \in N$, $\alpha_{i,\infty} \in \{-1, 1\}$. There are exactly three remaining cases to consider.

- If (i) does not hold then there are two cases to consider:
 1. There exists exactly 1 path of like-minded neighbors connecting $i \notin IO(\mathcal{E})$ to $i_k \in IO(\mathcal{E})$. In this case, by the same argument used to show part (ii), the long-run opinion of $i \notin IO(\mathcal{E})$ is $\alpha_{i,\infty} = \alpha_{i_k,\infty}$.
 2. There does not exist any path of like-minded neighbors connecting $i \notin IO(\mathcal{E})$ to $i_k \in IO(\mathcal{E})$. This is impossible when $|IO(\mathcal{E})| \geq 1$, because this would mean that i has herself an initially ideologically-opposed neighbor.
- If (i), (ii) do not hold and $|IO(\mathcal{E})| \geq 1$ then we need to consider the case where two individuals are directly linked and belong to $IO(\mathcal{E})$. Let $i \neq j \in IO(\mathcal{E})$ be two individuals such that $g_{ij} = 1$ and $|\alpha_{i,0} - \alpha_{j,0}| \geq \tau$ is the smallest element of the set $\{i \in N : \exists j \in N, |\alpha_{i,0} - \alpha_{j,0}| \geq \tau\}$. Without loss of generality suppose that $\alpha_{i,0} > \alpha_{j,0}$ so that $\alpha_{i,0} - \alpha_{j,0} \geq \tau$. I start by showing that :

$$P_1 : \alpha_{i,1} - \alpha_{j,1} > \alpha_{i,0} - \alpha_{j,0} \geq \tau.$$

Recall that: $\alpha_{i,1} = -\mu\alpha_{j,0} + \sum_{k \neq j \in N_i} a_{ik}(0)\alpha_{k,0}$ and $\alpha_{j,1} = -\mu\alpha_{i,0} + \sum_{k \neq i \in N_j} a_{jk}(0)\alpha_{k,0}$. It follows that their difference writes:

$$\begin{aligned}
\alpha_{i,1} - \alpha_{j,1} &= -\mu\alpha_{j,0} + \sum_{k \neq j \in N_i} a_{ik}(0)\alpha_{k,0} + \mu\alpha_{i,0} - \sum_{k \neq i \in N_j} a_{jk}(0)\alpha_{k,0} \\
&= \mu(\alpha_{i,0} - \alpha_{j,0}) + (1 + \mu(|\overline{N}_{i,0}| - |\underline{N}_{i,0}|))\alpha_{i,0} - (1 + \mu(|\overline{N}_{j,0}| - |\underline{N}_{j,0}|))\alpha_{j,0} \\
&+ \sum_{\substack{k \neq j \in N_i \\ k \neq i}} a_{ik}(0)\alpha_{k,0} - \sum_{\substack{k \neq i \in N_j \\ k \neq j}} a_{jk}(0)\alpha_{k,0} \\
&= (\alpha_{i,0} - \alpha_{j,0}) + \mu(\alpha_{i,0} - \alpha_{j,0}) + \mu|\overline{N}_{i,0}|\alpha_{i,0} - \mu \sum_{\substack{k \neq j \in \overline{N}_{i,0} \\ k \neq i}} \alpha_{k,0} - \mu|\underline{N}_{i,0}|\alpha_{i,0} + \mu \sum_{\substack{k \neq j \in \underline{N}_{i,0} \\ k \neq i}} \alpha_{k,0} \\
&- \mu|\overline{N}_{j,0}|\alpha_{j,0} + \mu \sum_{\substack{k \neq i \in \overline{N}_{j,0} \\ k \neq j}} \alpha_{k,0} + \mu|\underline{N}_{j,0}|\alpha_{j,0} - \mu \sum_{\substack{k \neq i \in \underline{N}_{j,0} \\ k \neq j}} \alpha_{k,0} \\
&= (\alpha_{i,0} - \alpha_{j,0}) + \mu(\alpha_{i,0} - \alpha_{j,0}) + \mu \sum_{\substack{k \neq j \in \overline{N}_{i,0} \\ k \neq i}} (\alpha_{i,0} - \alpha_{k,0}) \\
&+ \mu \sum_{\substack{k \neq j \in \underline{N}_{i,0} \\ k \neq i}} (\alpha_{k,0} - \alpha_{i,0}) + \mu \sum_{\substack{k \neq i \in \overline{N}_{j,0} \\ k \neq j}} (\alpha_{k,0} - \alpha_{j,0}) + \mu \sum_{\substack{k \neq i \in \underline{N}_{j,0} \\ k \neq j}} (\alpha_{j,0} - \alpha_{k,0}) = (\alpha_{i,0} - \alpha_{j,0}) + R,
\end{aligned}$$

where

$$\begin{aligned}
R &= \mu(\alpha_{i,0} - \alpha_{j,0}) + \mu \sum_{\substack{k \neq j \in \overline{N}_{i,0} \\ k \neq i}} (\alpha_{i,0} - \alpha_{k,0}) + \mu \sum_{\substack{k \neq j \in \underline{N}_{i,0} \\ k \neq i}} (\alpha_{k,0} - \alpha_{i,0}) + \\
&\mu \sum_{\substack{k \neq i \in \overline{N}_{j,0} \\ k \neq j}} (\alpha_{k,0} - \alpha_{j,0}) + \mu \sum_{\substack{k \neq i \in \underline{N}_{j,0} \\ k \neq j}} (\alpha_{j,0} - \alpha_{k,0}).
\end{aligned}$$

Statement P_1 holds if and only if $R > 0$. Assume by contradiction that $R < 0$. But, R is lower bounded by $\mu\tau + \mu\overline{N}_{i,0}\tau + \mu\underline{N}_{i,0}\tau + \mu\overline{N}_{j,0}\tau + \mu\underline{N}_{j,0}\tau > 0$.

General argument taking into account all cases:

Without loss of generality, let s be the time period such that, for all $t \geq s$, $\overline{N}_{i,t} = \overline{N}_{i,s}$ and $\underline{N}_{i,t} = \underline{N}_{i,s}$ for all $i \in N$. Now, consider the variation of opinion of individual $i \in N$ at period $t \geq s$:

$$\begin{aligned}
\alpha_{t,i} &= \alpha_{t-1,i} + \mu \sum_{j \neq i \in \overline{N}_{i,t-1}} (\alpha_{i,t-1} - \alpha_{j,t-1}) + \mu \sum_{j \neq i \in \underline{N}_{i,t-1}} (\alpha_{i,t-1} - \alpha_{j,t-1}), \\
\Leftrightarrow \alpha_{t,i} - \alpha_{t-1,i} &= \mu \sum_{j \neq i \in \overline{N}_{i,t-1}} (\alpha_{i,t-1} - \alpha_{j,t-1}) + \mu \sum_{j \neq i \in \underline{N}_{i,t-1}} (\alpha_{i,t-1} - \alpha_{j,t-1}).
\end{aligned}$$

Hence we can write the variation in the opinion of $i \in N$ as:

$$\frac{d\hat{\alpha}_i}{dt} = \mu \sum_{j \neq i \in \bar{N}_{i,t-1}} (\alpha_{i,t-1} - \alpha_{j,t-1}) + \mu \sum_{j \neq i \in \underline{N}_{i,t-1}} (\alpha_{i,t-1} - \alpha_{j,t-1}) = \nu_i.$$

For all $t \geq s$, $\frac{d\hat{\alpha}_i}{dt} = 0$ if and only if $\nu_i = 0$. Notice that it is impossible with $\alpha_{i,t-1} \neq 0$ that $\mu \sum_{j \neq i \in \bar{N}_{i,t-1}} (\alpha_{i,t-1} - \alpha_{j,t-1}) = 0$ because $|\alpha_{i,t-1} - \alpha_{j,t-1}| \geq \tau$ for all $j \neq i \in \bar{N}_{i,t-1}$.

$$\nu_i = 0 \Leftrightarrow \mu \sum_{j \neq i \in \bar{N}_{i,t-1}} (\alpha_{i,t-1} - \alpha_{j,t-1}) = -\mu \sum_{j \neq i \in \underline{N}_{i,t-1}} (\alpha_{i,t-1} - \alpha_{j,t-1}) \quad (2.9)$$

$$\Leftrightarrow g_{ij} = 0, \forall j \in \bar{N}_{i,t-1} \text{ and } \alpha_{i,t-1} = \alpha_{j,t-1}, \forall j \in \underline{N}_{i,t-1}, \text{ (consensus)} \quad (2.10)$$

$$\text{or } g_{ij} = 0, \forall j \in \bar{N}_{i,t-1} \text{ and } |\underline{N}_{i,t-1}| \alpha_{i,t-1} = \sum_{j \neq i \in \underline{N}_{i,t-1}} \alpha_{j,t-1}, \text{ (condition (i))} \quad (2.11)$$

$$\text{or } g_{ij} = 0, \forall j \in \underline{N}_{i,t-1} \text{ and } \alpha_{i,t-1} = 0 \text{ and } \sum_{j \neq i \in \bar{N}_{i,t-1}} \alpha_{j,t-1} = 0. \text{ (condition (ii))} \quad (2.12)$$

$$\text{or } \alpha_{i,t-1} = \frac{1}{|N_i|} \left(\sum_{j \neq i \in \bar{N}_{i,t-1}} \alpha_{j,t-1} + \sum_{j \neq i \in \underline{N}_{i,t-1}} \alpha_{j,t-1} \right), \text{ (condition (iii))} \quad (2.13)$$

Otherwise for all $t \geq s$, $|\frac{d\hat{\alpha}_i}{dt}| > \epsilon$ then the opinion of individual $i \in N$ keeps increasing or decreasing. But since opinions are bounded, her long-run opinion reaches 1 or -1 . ■

2.7.3. Proof of Lemma 2.1

Case 1: $|\alpha_{i,0} - \alpha_{j,0}| < \tau$. The law of motion 2.2 rewrites:

$$\begin{cases} \alpha_{i,t} = \alpha_{i,t-1} + \mu(\alpha_{j,t-1} - \alpha_{i,t-1}) \\ \alpha_{j,t} = \alpha_{j,t-1} + \mu(\alpha_{i,t-1} - \alpha_{j,t-1}) \end{cases} \Leftrightarrow \begin{cases} \alpha_{i,t} = (1 - \mu)\alpha_{i,t-1} + \mu\alpha_{j,t-1} \\ \alpha_{j,t} = (1 - \mu)\alpha_{j,t-1} + \mu\alpha_{i,t-1} \end{cases}$$

We can write the above system in matrix notation:

$$\begin{bmatrix} \alpha_{i,t} \\ \alpha_{j,t} \end{bmatrix} = \begin{bmatrix} 1 - \mu & \mu \\ \mu & 1 - \mu \end{bmatrix} \begin{bmatrix} \alpha_{i,t-1} \\ \alpha_{j,t-1} \end{bmatrix} \Leftrightarrow \begin{bmatrix} \alpha_{i,t} \\ \alpha_{j,t} \end{bmatrix} = \overbrace{\begin{bmatrix} 1 - \mu & \mu \\ \mu & 1 - \mu \end{bmatrix}^t}^{=M^t} \begin{bmatrix} \alpha_{i,0} \\ \alpha_{j,0} \end{bmatrix} \text{ (by induction).}$$

Moreover, we can diagonalize the matrix M^t so that we can compute the limit easily:

$$M^t = \begin{bmatrix} 1 - \mu & \mu \\ \mu & 1 - \mu \end{bmatrix}^t = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1^t & 0 \\ 0 & (1 - 2\mu)^t \end{bmatrix} \begin{bmatrix} 1/2 & 1/2 \\ -1/2 & 1/2 \end{bmatrix}.$$

For $\mu \in (0, 1/2)$, $\lim_{t \rightarrow \infty} (1 - 2\mu)^t = 0$. Notice that this is equivalent to upper bounding the distance between opinions at a given period t and the limiting opinions by the second highest eigenvalue.¹² It follows that when the opinions of i and j are close enough then they converge exactly to their average:

$$\alpha_\infty^a = \lim_{t \rightarrow \infty} \begin{bmatrix} \alpha_{i,t} \\ \alpha_{j,t} \end{bmatrix} = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix} \begin{bmatrix} \alpha_{i,0} \\ \alpha_{j,0} \end{bmatrix} = \begin{bmatrix} \frac{\alpha_{i,0} + \alpha_{j,0}}{2} \\ \frac{\alpha_{i,0} + \alpha_{j,0}}{2} \end{bmatrix}.$$

For $\epsilon > 0$, the time t_a it takes to convergence is: $t_a \geq \frac{\log(\epsilon)}{\log(1-2\mu)}$.

Case 2: $|\alpha_{i,0} - \alpha_{j,0}| \geq \tau$. The law of motion (2.2) rewrites:

$$\begin{cases} \alpha_{i,t} = \alpha_{i,t-1} + \mu(\alpha_{i,t-1} - \alpha_{j,t-1}) \\ \alpha_{j,t} = \alpha_{j,t-1} + \mu(\alpha_{j,t-1} - \alpha_{i,t-1}) \end{cases} \Leftrightarrow \begin{cases} \alpha_{i,t} = (1 + \mu)\alpha_{i,t-1} - \mu\alpha_{j,t-1} \\ \alpha_{j,t} = (1 + \mu)\alpha_{j,t-1} - \mu\alpha_{i,t-1} \end{cases}.$$

We can write the above system in matrix notation:

$$\begin{bmatrix} \alpha_{i,t} \\ \alpha_{j,t} \end{bmatrix} = \begin{bmatrix} 1 + \mu & -\mu \\ -\mu & 1 + \mu \end{bmatrix} \begin{bmatrix} \alpha_{i,t-1} \\ \alpha_{j,t-1} \end{bmatrix} \Leftrightarrow \begin{bmatrix} \alpha_{i,t} \\ \alpha_{j,t} \end{bmatrix} = \overbrace{\begin{bmatrix} 1 + \mu & -\mu \\ -\mu & 1 + \mu \end{bmatrix}^t}^{=M^t} \begin{bmatrix} \alpha_{i,0} \\ \alpha_{j,0} \end{bmatrix} \quad (\text{by induction}).$$

Moreover, we can diagonalize the matrix M^t :

$$M^t = \begin{bmatrix} 1 + \mu & -\mu \\ -\mu & 1 + \mu \end{bmatrix}^t = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1^t & 0 \\ 0 & (1 + 2\mu)^t \end{bmatrix} \begin{bmatrix} 1/2 & 1/2 \\ -1/2 & 1/2 \end{bmatrix}.$$

The limit opinions of i and j are:

$$\alpha_\infty^r \lim_{t \rightarrow \infty} \begin{bmatrix} \alpha_{i,t} \\ \alpha_{j,t} \end{bmatrix} = \lim_{t \rightarrow \infty} \frac{1}{2} \begin{bmatrix} 1 + (1 + 2\mu)^t & 1 - (1 + 2\mu)^t \\ 1 - (1 + 2\mu)^t & 1 + (1 + 2\mu)^t \end{bmatrix} \begin{bmatrix} \alpha_{i,0} \\ \alpha_{j,0} \end{bmatrix} = \lim_{t \rightarrow \infty} \frac{1}{2} \begin{bmatrix} \alpha_{i,0} + \alpha_{j,0} + (\alpha_{i,0} - \alpha_{j,0})(1 + 2\mu)^t \\ \alpha_{i,0} + \alpha_{j,0} + (\alpha_{j,0} - \alpha_{i,0})(1 + 2\mu)^t \end{bmatrix}.$$

For any positive μ this limit explodes. However, recall that opinions have an upper 1 and lower bound -1 . It follows that when the opinions of i and j are faraway they diverge until they reach the upper and lower limit of opinions. Moreover, there exists a time t for a given

¹²For more details on this topic in linear algebra See Silva, Silva and Fernandes (2016) [91].

$\mu > 0$ such that that we remain within the permitted bounds. To find this time t given μ , we must solve:

$$\begin{cases} \frac{1}{2}(\alpha_{i,0} + \alpha_{j,0} + (\alpha_{i,0} - \alpha_{j,0})(1 + 2\mu)^t) = 1 & \text{if } \alpha_{i,0} > \alpha_{j,0} \\ \frac{1}{2}(\alpha_{i,0} + \alpha_{j,0} + (\alpha_{i,0} - \alpha_{j,0})(1 + 2\mu)^t) = -1 & \text{if } \alpha_{i,0} < \alpha_{j,0}. \end{cases}$$

Given μ , we get the following t_r (for integer values take the floor function):

$$t_r = \begin{cases} \frac{\log\left(\frac{2-\alpha_{i,0}-\alpha_{j,0}}{\alpha_{i,0}-\alpha_{j,0}}\right)}{\log(1+2\mu)} & \text{if } 1 \geq \alpha_{i,0} > \alpha_{j,0} \geq -1 \\ \frac{\log\left(\frac{-2-\alpha_{i,0}-\alpha_{j,0}}{\alpha_{i,0}-\alpha_{j,0}}\right)}{\log(1+2\mu)} & \text{if } -1 \leq \alpha_{i,0} < \alpha_{j,0} \leq 1. \end{cases}$$

For very small ϵ and $\mu \in (0, 1/2)$, it takes a very large number of periods to reach consensus while to reach the bounds 1 and -1 the individuals take a finite number of time periods. In other words, $t_r < t_a$ because we can always find a small enough ϵ such that the inequality holds. Formally, we solve the inequality $t_a > t_r$ for $\epsilon > 0$, for the case where $\alpha_{i,0} > \alpha_{j,0}$ (similarly for the other case) and t_a at its lower bound:

$$\frac{\log(\epsilon)}{\log(1-2\mu)} > \frac{\log\left(\frac{2-\alpha_{i,0}-\alpha_{j,0}}{\alpha_{i,0}-\alpha_{j,0}}\right)}{\log(1+2\mu)} \Leftrightarrow \epsilon < \exp\left(\frac{\log\left(\frac{2-\alpha_{i,0}-\alpha_{j,0}}{\alpha_{i,0}-\alpha_{j,0}}\right) \log(1-2\mu)}{\log(1+2\mu)}\right).$$

2.7.4. Proof of Theorem 2.1

Part (i) convergence: let λ be an eigenvalue of the matrix \tilde{G} . Recall that the algebraic multiplicity of λ is the number of times it is repeated as a root of the characteristic polynomial and the geometric multiplicity of λ is the maximum number of linearly independent eigenvectors associated with λ . An eigenvalue is semi-simple if its algebraic multiplicity is equal to its geometric multiplicity (definitions p.510, chapter 7, Meyer (2000) [78]). For $\tilde{G} \in \mathbb{R}^{n \times n}$, $\lim_{t \rightarrow \infty} \tilde{G}^t$ exists if and only if $\rho(\tilde{G}) < 1$ (the spectral radius) or else $\rho(\tilde{G}) = 1$ where $\lambda = 1$ is the only eigenvalue on the unit circle and $\lambda = 1$ is semi-simple (see *Limits of Powers* page 630, chapter 7, in Meyer (2000) [78]). Moreover, for every stochastic matrix, the spectral radius is 1 and it is semi-simple (p.696, Chapter 8 in Meyer (2000) [78] or see Corollary 2, page 2214, in Ding and Rhee (2011) [65]). Since, matrix \tilde{G} is a stochastic matrix, it has a spectral radius of 1 and it is semi-simple. Therefore, \tilde{G} is a convergent matrix.

Part (ii) spectral projector: when $\lim_{t \rightarrow \infty} \tilde{G}^t$ exists, it is equal to the spectral projector associated with eigenvalue 1 (again see p.630, chapter 7, in Meyer (2000) [78]).

Reminder from p.629 Meyer (2000) [78]. Recall that a row stochastic matrix A can be de-

composed using its Jordan form J :

$$J = \begin{bmatrix} I_{p \times p} & \mathbf{0} \\ \mathbf{0} & \mathbf{K} \end{bmatrix}$$

where $I_{p \times p}$ is the identity matrix of size p , with p the algebraic multiplicity of the eigenvalue 1 and \mathbf{K} a diagonal matrix with entries corresponding to remaining eigenvalues which are strictly smaller than 1. Hence, $\tilde{G}_\theta^t = PJ^tP^{-1}$. Now write $P = (P_1, P_2)$ where P_1 are the columns that correspond to the eigenvectors associated with the eigenvalues 1 and P_2 are the columns that correspond the eigenvectors associated with the remaining eigenvalues which are strictly smaller than 1. Similarly $P^{-1} = Q = (Q_1; Q_2)$ with Q_1 the lines associated with the eigenvalues 1. Since K^t vanishes when t is large because all the diagonal entries are strictly smaller than one, $\lim_{t \rightarrow \infty} \tilde{G}_\theta^t = P_1Q_1$ which is the spectral projector of the eigenvalue 1.

Part (iii). The multiplicity of the eigenvalue 1 is equal to the number of essential classes.

Recall that from Seneta (1981) [89]: *we say that i leads to j and write $i \rightarrow j$ if there exists an integer $m \geq 1$ such that $t_{ij}^m > 0$ (chain between i and j). We say that i and j communicate if $i \rightarrow j$ and $j \rightarrow i$ and write in this case $i \leftrightarrow j$. The index i is called essential when : $i \rightarrow j$ implies $i \leftrightarrow j$ and there is at least one j such that $i \rightarrow j$. It is therefore clear that all essential indices (if any) can be subdivided into essential classes in such a way that all the indices belonging to one class communicate, but cannot lead to an index outside the class.*

The matrix \tilde{G} can contain several essential classes that are either: (i) singletons, when an expresser has reached the upper or lower bound of the opinion interval and is no longer updating their opinion (one self-loop), or (ii) contain more than one expresser, this occurs when individuals within a connected set of expressers are like-minded and keep updating their opinions until they reach consensus. Each sub-matrix of \tilde{G} corresponding to an essential class is row stochastic, because (a) there are no outgoing edges from the members of the essential class to members outside the class by definition and (b) the matrix \tilde{G} is row stochastic. Furthermore, a sub-matrix corresponding to a single self-communicating class is irreducible. Hence, each sub-matrix corresponding to an essential class is an irreducible aperiodic (because of self-loops) stochastic sub-matrix and by the Perron-Frobenius theorem of non-negative matrices, each such sub-matrix has an associated eigenvalue 1 that is simple.

Finally, the matrix \tilde{G} can be interpreted as an n -state Markov chain. Form Seneta (1981) we further know that *if an n -state MC contains at least two essential classes of states, then any*

weighted linear combination of the stationary distribution vectors corresponding to each such class, each appropriately augmented by zeros to give an $(n \times 1)$ vector, is a stationary distribution of the chain.

2.7.5. Proof of Proposition 2.3

Case (i): $|E| = 1$. Suppose that $E = \{i\}$ and that the initial opinion of this individual is $\alpha_{0,i}$. Then using proposition 2.1, this expresser remains stubborn forever because she does not have neighbors who also express. Hence they never updates their opinion and $\alpha_{\infty,i} = \alpha_{0,i}$. Moreover from theorem 2.1 the long-run opinion of consensual individuals is a convex combination of opinions of expressers. Since there is only one expresser then all consensual individuals have long-run opinion of $\alpha_{0,i}$ and consensus prevails.

Case (ii): $|E| > 1$. Long-run opinions form consensus when $\forall i \neq j \in N, |\alpha_{\infty,i} - \alpha_{\infty,j}| < \tau$.

1. Let $E = \bigcup_{k=1}^{\kappa} \mathcal{E}_k$ be the set of expressers such that $\kappa \geq 1$.
2. For $\kappa = 1$ there is a unique set of connected expressers $E = \mathcal{E}_1$ and long-run opinions form a consensus if and only if $\forall i \neq j \in \mathcal{E}_1$, such that $g_{ij} = 1, |\alpha_{0,i} - \alpha_{0,j}| < \tau$. Since each member within \mathcal{E}_1 has like-minded neighbors, the members of \mathcal{E}_1 converge to the average of their initial opinions as shown in proposition 2.1. Moreover, using theorem 2.1, the opinions of consensual individuals are convex combinations of the opinions of expressers. Since here there is only one set of connected expressers, the opinion of each consensual individual is exactly the average of opinions of the members of the set of connected expressers \mathcal{E}_1 .
3. For $\kappa > 1$, without loss of generality suppose that the union of the two connected set of expressers \mathcal{E}_1 and \mathcal{E}_2 is equal to E . Long-run opinions form consensus if and only if (i) $\forall i \neq j \in \mathcal{E}_k$ for $k \in \{1, 2\}$, such that $g_{ij} = 1, |\alpha_{0,i} - \alpha_{0,j}| < \tau$ (ii) $|\bar{\alpha}_{0,\mathcal{E}_1} - \bar{\alpha}_{0,\mathcal{E}_2}| < \tau$ where $\bar{\alpha}_{0,\mathcal{E}_k}$ is the average of initial opinions within the set \mathcal{E}_k for $k \in \{1, 2\}$. Since within each of both sets all members have like-minded neighbors within \mathcal{E}_1 and \mathcal{E}_2 opinions of expressers converge respectively to $\bar{\alpha}_{0,\mathcal{E}_1}$ and $\bar{\alpha}_{0,\mathcal{E}_2}$. Moreover, consensus can prevail if and only if $|\bar{\alpha}_{0,\mathcal{E}_1} - \bar{\alpha}_{0,\mathcal{E}_2}| < \tau$ because then any long-run opinion of a consensual individual $i \in N \setminus E$ is at a distance of at most τ from any other long-run opinion of

other individuals in the network. Formally,

$$\begin{aligned}
\alpha_{\infty_i} &= \sum_{j \in \mathcal{E}_1 \cup \mathcal{E}_2} \mathcal{G}_{ij} \alpha_{j,\infty} = \sum_{j_1 \in \mathcal{E}_1} \mathcal{G}_{ij_1} \bar{\alpha}_{0,\mathcal{E}_1} + \sum_{j_2 \in \mathcal{E}_2} \mathcal{G}_{ij_2} \bar{\alpha}_{0,\mathcal{E}_2} \\
&= \bar{\alpha}_{0,\mathcal{E}_1} \sum_{j_1 \in \mathcal{E}_1} \mathcal{G}_{ij_1} + \bar{\alpha}_{0,\mathcal{E}_2} \sum_{j_2 \in \mathcal{E}_2} \mathcal{G}_{ij_2} \\
&= \bar{\alpha}_{0,\mathcal{E}_1} \sum_{j_1 \in \mathcal{E}_1} \mathcal{G}_{ij_1} + \bar{\alpha}_{0,\mathcal{E}_2} (1 - \sum_{j_1 \in \mathcal{E}_1} \mathcal{G}_{ij_1}) \\
&= \bar{\alpha}_{0,\mathcal{E}_2} + (\bar{\alpha}_{0,\mathcal{E}_1} - \bar{\alpha}_{0,\mathcal{E}_2}) \sum_{j_1 \in \mathcal{E}_1} \mathcal{G}_{ij_1}
\end{aligned}$$

Hence for any expresser in \mathcal{E}_2 the opinion difference with a given consensual individual $i \in N \setminus E$ is at most τ (similarly for any expresser in \mathcal{E}_1):

$$|\alpha_{\infty_i} - \bar{\alpha}_{0,\mathcal{E}_2}| = |(\bar{\alpha}_{0,\mathcal{E}_1} - \bar{\alpha}_{0,\mathcal{E}_2}) \sum_{j_1 \in \mathcal{E}_1} \mathcal{G}_{ij_1}| \leq |(\bar{\alpha}_{0,\mathcal{E}_1} - \bar{\alpha}_{0,\mathcal{E}_2})| < \tau$$

Furthermore, for two consensual individuals $i \neq j \in N \setminus E$ their long-run is at most τ because:

$$\begin{aligned}
|\alpha_{\infty_i} - \alpha_{\infty_j}| &= |(\bar{\alpha}_{0,\mathcal{E}_1} - \bar{\alpha}_{0,\mathcal{E}_2}) (\sum_{j_1 \in \mathcal{E}_1} \mathcal{G}_{ij_1} - \sum_{j_1 \in \mathcal{E}_1} \mathcal{G}_{jj_1})| \leq |(\bar{\alpha}_{0,\mathcal{E}_1} - \bar{\alpha}_{0,\mathcal{E}_2})| (1 - \sum_{j_1 \in \mathcal{E}_1} \mathcal{G}_{jj_1}) \\
&\leq |\bar{\alpha}_{0,\mathcal{E}_1} - \bar{\alpha}_{0,\mathcal{E}_2}| \\
&< \tau
\end{aligned}$$

The arguments easily extend for more than 2 sets of connected expressers. To see this, think of three sets of connected expressers \mathcal{E}_1 , \mathcal{E}_2 and \mathcal{E}_3 where the long-run opinions within each set are respectively $\bar{\alpha}_{0,\mathcal{E}_1}$, $\bar{\alpha}_{0,\mathcal{E}_2}$ and $\bar{\alpha}_{0,\mathcal{E}_3}$. For all $i \in N \setminus E$, $\alpha_{i,\infty} \in \text{conv}(\bar{\alpha}_{0,\mathcal{E}_1}, \bar{\alpha}_{0,\mathcal{E}_2}, \bar{\alpha}_{0,\mathcal{E}_3})$ and $|\bar{\alpha}_{0,\mathcal{E}_2} - \bar{\alpha}_{0,\mathcal{E}_1}| < \tau$, $|\bar{\alpha}_{0,\mathcal{E}_3} - \bar{\alpha}_{0,\mathcal{E}_1}| < \tau$, $|\bar{\alpha}_{0,\mathcal{E}_2} - \bar{\alpha}_{0,\mathcal{E}_3}| < \tau$,

it follows that :

$$\begin{aligned}
|\alpha_{i,\infty} - \bar{\alpha}_{0,\mathcal{E}_1}| &= \left| \sum_{j_1 \in \mathcal{E}_1} \mathcal{G}_{ij_1} \bar{\alpha}_{0,\mathcal{E}_1} + \sum_{j_2 \in \mathcal{E}_2} \mathcal{G}_{ij_2} \bar{\alpha}_{0,\mathcal{E}_2} + \sum_{j_3 \in \mathcal{E}_3} \mathcal{G}_{ij_3} \bar{\alpha}_{0,\mathcal{E}_3} \right. \\
&\quad \left. - \overbrace{\left(\sum_{j_1 \in \mathcal{E}_1} \mathcal{G}_{ij_1} + \sum_{j_2 \in \mathcal{E}_2} \mathcal{G}_{ij_2} + \sum_{j_3 \in \mathcal{E}_3} \mathcal{G}_{ij_3} \right)}^{=1} \bar{\alpha}_{0,\mathcal{E}_1} \right| \\
&= \left| \sum_{j_2 \in \mathcal{E}_2} \mathcal{G}_{ij_2} \bar{\alpha}_{0,\mathcal{E}_2} + \sum_{j_3 \in \mathcal{E}_3} \mathcal{G}_{ij_3} \bar{\alpha}_{0,\mathcal{E}_3} - \sum_{j_2 \in \mathcal{E}_2} \mathcal{G}_{ij_2} \bar{\alpha}_{0,\mathcal{E}_1} - \sum_{j_3 \in \mathcal{E}_3} \mathcal{G}_{ij_3} \bar{\alpha}_{0,\mathcal{E}_1} \right| \\
&= \left| \sum_{j_2 \in \mathcal{E}_2} \mathcal{G}_{ij_2} (\bar{\alpha}_{0,\mathcal{E}_2} - \bar{\alpha}_{0,\mathcal{E}_1}) + \sum_{j_3 \in \mathcal{E}_3} \mathcal{G}_{ij_3} (\bar{\alpha}_{0,\mathcal{E}_3} - \bar{\alpha}_{0,\mathcal{E}_1}) \right| \\
&< \sum_{j_2 \in \mathcal{E}_2} \mathcal{G}_{ij_2} \tau + \sum_{j_3 \in \mathcal{E}_3} \mathcal{G}_{ij_3} \tau \\
&\leq \tau
\end{aligned}$$

2.7.6. Proof of Lemma 2.2

Let $k \in N$ be a consensual individual. It follows from theorem 2.1 that:

$$\alpha_{k,\infty} = \sum_{i \in E} \mathcal{G}_{k,i} \alpha_{i,\infty} = \sum_{i \in E^+} \mathcal{G}_{k,i} - \sum_{i \in E^-} \mathcal{G}_{k,i}$$

Moreover, recall that for all $k \in N$, $\sum_{i \in E} \mathcal{G}_{k,i} = \sum_{i \in E^+} \mathcal{G}_{k,i} + \sum_{i \in E^-} \mathcal{G}_{k,i} = 1$. Individual k doesn't hold the extreme opinion of the members of E^+ , if and only if, for $i \in E^+$:

$$\begin{aligned}
|\alpha_{k,\infty} - \alpha_{i,\infty}| \geq \tau &\Leftrightarrow \left| \left(2 \sum_{i \in E^+} \mathcal{G}_{ki} - 1 \right) - 1 \right| \geq \tau \\
&\Leftrightarrow 1 - \sum_{i \in E^+} \mathcal{G}_{ki} \geq \frac{\tau}{2} \\
&\Leftrightarrow 1 - \frac{\tau}{2} \geq \sum_{i \in E^+} \mathcal{G}_{ki}
\end{aligned}$$

Similarly, individual k doesn't hold the extreme opinion of the members of E^- , if and only if, for $i \in E^-$:

$$\begin{aligned}
|\alpha_{k,\infty} - \alpha_{i,\infty}| \geq \tau &\Leftrightarrow \left| \left(2 \sum_{i \in E^+} \mathcal{G}_{ki} - 1 \right) - (-1) \right| \geq \tau \\
&\Leftrightarrow \sum_{i \in E^+} \mathcal{G}_{ki} \geq \frac{\tau}{2}
\end{aligned}$$

Hence, if if long-run opinions are bi-polarized then necessarily all individuals belong to

either of both extreme opinion groups and there does not exist an individual $k \in N$ such that:

$$1 - \tau/2 \leq \sum_{i \in E^+} \mathcal{G}_{ki} \leq \tau/2$$

Now recall that from theorem 2.1 that the limit of \tilde{G}^t when t is large, exists and is given by \mathcal{G} . Since \tilde{G} is row stochastic, we can show by induction that \tilde{G}^t is also row stochastic. The row stochasticity of \tilde{G}^t is a linear condition, hence continuous so it is preserved by the limits. It follows that :

$$\sum_{j \in N} \sum_{i \in N} \mathcal{G}_{ij} = |N| \quad (2.14)$$

Let \mathbf{a}_∞ be a bi-polarized long-run opinion vector (see definition 2.6) and suppose by contradiction that

$$\left| \sum_{i \in E^+} (\mathcal{G}' \mathbf{1})_i - \sum_{i \in E^-} (\mathcal{G}' \mathbf{1})_i \right| > \epsilon \quad (2.15)$$

Equation (2.15) can be rewritten as:

$$\sum_{i \in E^+} (\mathcal{G}' \mathbf{1})_i - \sum_{i \in E^-} (\mathcal{G}' \mathbf{1})_i = \nu \text{ for } |\nu| > 1$$

Moreover, $\sum_{i \in E^+} (\mathcal{G}' \mathbf{1})_i = \sum_{j \in N} \sum_{i \in E^+} \mathcal{G}_{ji}$ and for all $j \in N$, $\sum_{i \in E^+} \mathcal{G}_{ji} + \sum_{i \in E^-} \mathcal{G}_{ji} = 1$. Hence:

$$\begin{aligned} \sum_{i \in E^+} (\mathcal{G}' \mathbf{1})_i - \sum_{i \in E^-} (\mathcal{G}' \mathbf{1})_i = \nu &\Leftrightarrow \sum_{j \in N} \sum_{i \in E^+} \mathcal{G}_{ji} - \sum_{j \in N} \sum_{i \in E^-} \mathcal{G}_{ji} = \nu \\ &\Leftrightarrow \sum_{j \in N} \left(\sum_{i \in E^+} \mathcal{G}_{ji} - \sum_{i \in E^-} \mathcal{G}_{ji} \right) = \nu \\ &\Leftrightarrow \sum_{j \in N} \left(1 - 2 \sum_{i \in E^-} \mathcal{G}_{ji} \right) = \nu \\ &\Leftrightarrow \frac{|N| - \nu}{2} = \sum_{j \in N} \sum_{i \in E^-} \mathcal{G}_{ji} = \sum_{j \in N} \sum_{i \in N^-} \mathcal{G}_{ji} \end{aligned}$$

Since $|\nu| > 1$ and there does not exist moderate individuals, the size of the group N^- is strictly smaller than $|N|/2$ which contradicts the definition of a bi-polarized long-run opinion vector.

2.7.7. Network Statistics

Consider a graph G , with N vertices and M edges.

- Assortativity or assortative mixing in Newman (2002) [82], assortativity of an observed network is given by :

$$r = \frac{M^{-1} \sum_i j_i k_i - [M^{-1} \sum_i \frac{1}{2}(j_i + k_i)]^2}{M^{-1} \sum_i \frac{1}{2}(j_i^2 + k_i^2) - [M^{-1} \sum_i \frac{1}{2}(j_i + k_i)]^2},$$

where j_i and k_i are the degrees of the vertices at the ends of the i th edge with $i = 1, \dots, M$.

- Neighborhood connectivity: average degree in the neighborhood of a given node $i \in N$

$$\frac{1}{d_i} \sum_{j \in N_i} d_j,$$

where d_i is the degree of node i and N_i is the neighborhood.

- Per capita average degree:

$$\frac{\sum_{i \in N} d_i}{N},$$

where d_i is the degree of node i .

- Average path length :

$$l = \frac{1}{N(N-1)} \sum_{i \neq j} d(v_i, v_j),$$

where $d(v_i, v_j)$ is the shortest path between nodes v_1 and v_2 (computed here with Dijkstra).

2.7.8. Eigenvector centrality

For the simulation I have used the local popularity measure defined as the number of friends over the average number of friends of friends. I have argued in the introduction that a local measure is suitable for an opinion formation model, as an individual can have locally high impact over her own friends, even though she may not be globally influential in the network. Nevertheless, the convergence results go through for any other centrality measure. Figures 2.15 and 2.16 show the final opinions for the same initial opinion and network structure but two different centrality measures, respectively eigenvector centrality and local popularity. We have used as an expression threshold $\delta^* = 1/n$ when taking the eigenvector centrality measure so that in any regular network, individuals with no different levels of expertise express. We see that when using the eigenvector centrality measure, locally popular individuals do not get to express. Furthermore there is a lot of interaction and disagreement within the hub containing the agent with the highest eigenvector centrality and her friends.

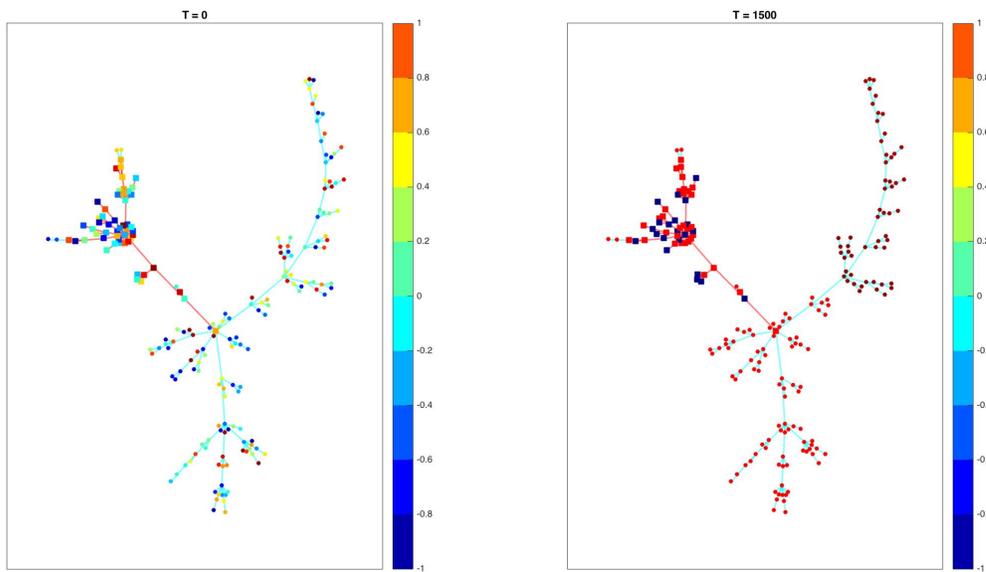


Figure 2.15: Eigenvector centrality. Squares are expressers, mean polarization 0.47

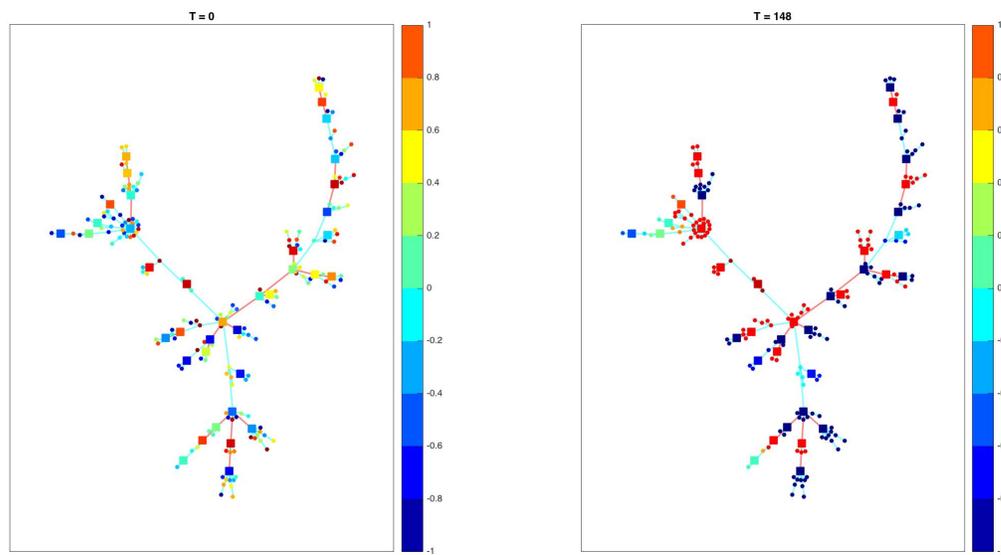


Figure 2.16: Local popularity. Squares are expressers, mean polarization 0.78

3. Strategic cultural migration with peer effects

(joint with Alexia Lochmann¹)

Abstract. This paper examines the role played by cultural identity and social contacts in self-selection into migration. On June 23rd, 1939 in Berlin, Benito Mussolini and Adolf Hitler signed an agreement that gave the German speaking South Tyrolean population in northern Italy an option: to leave their homeland (move to the Third Reich) and keep their culture, or to stay in their homeland and abandon their culture. We formalize a model where strategic households interact together and affect each others' intentions-to-emigrate. Each household has a cultural type and can have two different sorts of interactions with other households: *(i)* interactions within their social circles and *(ii)* interactions with their geographic neighbors who own property. Households have an incentive to match the emigration decision of social contacts in order to avoid the loss of social capital, and mis-match the migration decision of their geographic neighbors to seize the opportunity of buying their property if they migrate. Relying on the recently digitized micro level household data we provide empirical evidence for our model. Using a *2SLS* approach to address endogeneity in peer effects, we find that a stronger cultural connection towards the German culture and the emigration decision of the social acquaintances significantly increases the intentions-to-emigrate. We see that personal property as well as neighborhood property, on the other hand, significantly decrease the intentions-to-emigrate.

Keywords: *migration choice, cultural identity, European history, peer effects.*

JEL Classification: J15, Z10, Z13, N34, C72, L14.

¹Institut Convergences Migrations and Free University of Bolzano. This research project benefits from the from the support of the Archivio di Stato di Bolzano, the Free University of Bolzano, Labex OSE and institut Convergences Migrations. We are particularly indebted to Harald Toniatti and Pietro Vezzani.

3.1. Introduction

In the current public debate, as well as in the literature on international migration, a core question is self-selection into migration. It follows that the phenomenon received a great deal of attention in the economic literature. Centered around the *Roy* model, this literature agrees on wage differentials as major drivers of the decision *whether* to move, and *where* to move. However, the decision if and where to emigrate goes beyond the sole consideration of economic variables. Individuals often have in their homeland many types of relationships, with neighbors, friends, colleagues and family, that impact differently their migration decisions. The feeling of safety and a future in both the home and the destination country, perceived economic opportunities in both countries, family and friends networks, as well as a cultural factors, play decisive roles. These factors are crucial, yet understudied, often because of the difficulty to observe and measure them.

When individuals are engaged in different types of relationships in their homeland, to what extent do these multiple interactions affect their emigration intentions? Does cultural attachment play a role in self-selection into migration? Our paper answers these questions by considering the inter-play between cultural, social and economic components of migration decisions. To that end, we formalize an emigration game with strategic households, where each has a cultural type. We study the behavior of households when interacting in two contexts: the social network which accounts for peer pressure and the geographic network which accounts for economic opportunities of expanding one's property. We further provide original empirical evidence based on the recent digitization of historical household level data from South Tyrol, the northernmost province of Italy. This dataset provides ground for a rich research agenda allowing for the analysis of a case study of the role of cultural identity in the decision making process of migration.

Dalen and Henkens (2007) [41] argue that to discover the roots of selection into migration, one should start in the source country, before selection has taken place.¹ On that account, the academic contribution of this paper is framed by a story around the definition of "home", about roots that tie places to people, but also about cultural roots, social roots, ideological roots and economic roots. On June 23rd, 1939 in Berlin, Benito Mussolini and Adolf Hitler signed an agreement that gave the German speaking South Tyrolean population an option: to leave their homeland and keep their culture, or to stay in their homeland and

¹This argument stems from the Theory of Planned Behavior (see Armitage and Conner (2001) [7]). Many papers that depart from economic considerations as the sole determinant of migration, focus on emigration intentions as they reveal different mechanisms in the decision making process for self-selection into migration (e.g. Gubhaju and De Jong (2009) [56] or Caron (2020) [35]).

abandon their culture², by accepting full assimilation to Italian culture and language. The idea was that “Migration might succeed where assimilation had failed”, Alcock (2001) [5]. The agreement was designed to facilitate Mussolini’s effort to fully italianize South Tyrol, and to advance Hitler’s plans to assemble all Germans in one empire. The percentage of people who opted into leaving the homeland, by filling an official form at their municipality, lies somewhere between eighty and ninety percent.³ In the end, about half of those households ended up effectively emigrating. Anecdotes point towards several explanations: peer pressure pushed households with a strong Germanic culture to effectively migrate, but some families ended up staying because economic opportunities were opening-up due the possibility of purchasing the neighbor’s house.

To study the cultural, social and economic components of self-selection into migration, we formalize a game with strategic households that report their intentions-to-emigrate. Each household has a cultural type, which is their degree of attachment to the germanic culture. Households interact with other households within their social circles, including friends, work colleagues and neighbors. The social circle of each household represents their social capital. Furthermore, to account for economic factors we consider property ownership.⁴ Households can interact with their geographic neighbors. The properties of neighbors who decide to migrate are potential economic opportunities that a given household can seize. Hence, each household can have two types of interactions: (i) with acquaintances in the social network and (ii) with neighbors in the geographic network. Households optimally select an action which maximizes their net migration gains when taking into account own cultural type, own property and interactions with peers in both networks.

This paper makes several contributions. First, we characterize the optimal behavior of households when playing the emigration game. We show that when only considering interactions in the social network (*friendship game*), households intentions-to-emigrate are strategic complements. To be more precise, for a given household, if most of their social contacts intend to emigrate, then the household is also more likely to emigrate otherwise they lose their social capital. Furthermore, each household’s place of residence is represented by a node in a geographic network. Properties have values ranging from 0 to 1, where 0 means that household does not own their property. Independently of interactions in the social network, households have a lower incentive to emigrate, when all their neighbors in the geographic network have the intention to emigrate. This is because they can purchase the

²By the time almost all of the population was German speaking.

³Different entities who handled this agreement had interest in publishing different numbers, see for example Alexander et al. (1993) [6].

⁴Given that households belong to an agricultural society in the late 1930’s property ownership is the most relevant economic component of migration decisions.

neighbor's house and enhance their economic well-being. In the geographic network, migration intentions are strategic substitutes (*chicken game*). When combining both types of interactions, we show that households impact each other's emigration intentions depending on the linkage patterns in both the social and geographic networks. Two individuals who are linked in both the social and the geographic network (e.g. a geographic neighbor that is also a friendly work colleague), affect positively each other's migration decisions, but the higher the value of the neighbor's property, the lower the effect. We study an exhaustive example to show how network positions play out in the net migration gains. A given individual who is influential in the social network, has negative migration gains whenever her social contacts have low cultural types entailing relatively low intentions-to-emigrate, even if the individual herself has relatively strong germanic culture.

Second, we provide empirical evidence of the above explained mechanisms, based on the recent digitization of historical household level data. We look at the case of South Tyrol, whose Institutional Autonomy is often mentioned as a successful example in the search for solutions to problems in culturally divided communities. The breakthrough in creating a peaceful environment was founded through the understanding of an *extended* identity in contrast to a *changed* identity often aimed at through assimilation policies. The information from the data set allows for definition and creation of novel proxies for cultural identity.

To proxy migration intentions, we rely on the application date as an outcome variable, where the application period ranges from fall 1939 until early 1945. We then normalize this variable between zero and one, one reflecting the earliest application date and zero the latest. We develop different proxies for cultural identity, more precisely "inherited" and "acquired" cultural preferences, and show their relevance as well as test for their robustness. The concept of "inherited" cultural preferences stems from sociology literature and was recently taken up by economists (see, e.g. Fryer Jr et al. (2004) [68]; Rubinstein et al. (2013) [86], Biavaschi et al. (2017) [19]). This literature proxies parents' cultural preferences which are reflected in their children's names. The first name of an adult can, hence, be interpreted as their *inherited* cultural preferences, i.e. the preferences of their parents at the time they named them. Furthermore, we add a proxy of "acquired" cultural preferences over the course of lifetime, which derives from past migration history to German-speaking countries.

We add personal property and neighborhood property as economic components of the emigration decision making process. In particular to define neighborhood property we look at the addresses of households in the data set and map them to a geographic network. We further build a social network, where the social contacts are individuals who share the same occupation within the same village, family members and geographic neighbors. Finally, we add a set of control variables such as age, gender and marital status. We discuss this set

of cultural, economic and social network related variables and their role in the emigration intentions first in a simple *OLS* setting, and then in a *2SLS* setting relying on the methodology proposed by Gaviria and Raphael (2001) [52] to address possible endogeneity concerns related to peer effects. We use the average background characteristics of the social circles to instrument for their emigration intentions.

We find evidence that a stronger cultural connection towards the German culture significantly anticipates the application day, showing, hence, a strong intention-to-emigrate. The emigration decision of the social networks also significantly anticipates the application day. We see that personal property as well as neighborhood property, on the other hand, significantly delay the application day, hence showing a weaker intention-to-emigrate. Using data on the farming profession, and on police records and illnesses, we can rule out alternative explanations for emigrating, based on some historical evidence that suggests that part of the leaving decision might have been forced by the participating governments. We also add a robustness test proposed by Oster (2019) [84], which further provides confidence for the significance and robustness of the results we find.

Related literature. This paper is related to several strands of literatures. First, it contributes to the literature related to the drivers of self-selection into migration. Many seminal papers (e.g. Borjas (1987) [31], Chiswick (1999) [38], Chiquiar and Gordon (2005) [37], Bauer et al. (2005) [14], McKenzie and Rapoport (2010) [76], Moraga (2011) [79], Borjas et al. (2019) [30]) study economic push-factors, such as differences in income, and migration networks as the main drivers for selection into migration. The most common framework cited and worked around in the field of self-selection into migration is the *Roy Model*. This workhorse model was first introduced by Borjas (1987) [31]. It analyzes the economic drivers of migration: under which conditions does an individual decide to move? The main finding is that individuals decide to move when the earnings for their skill level are higher abroad than at home. Our paper extends this literature by adding cultural and social factors to the economic considerations of agents. It is not evident to quantify the magnitude of the impact of each of cultural and social on the emigration decision. The real world large scale migration experiment we examine tries to disentangle the cultural preferences components and quantify their magnitude. Furthermore, we depart from this literature by focusing on the cost of migration in terms of loss of social capital and economic opportunities in the homeland. Hence this cost depends on household migration intentions and is not a constant which is exogenous to the model and uniform across agents.

Second, this paper builds on an emerging literature which incorporates origin and destination networks in the study of migration. Munshi (2020) [80] argues that the Roy model cannot explain a number of stylized facts that include higher mobility of individuals to-

wards destinations where others from their origin have moved before. McKenzie and Rapoport (2010) are among the first papers that introduce migration networks to the question around self-selection into migration. The authors argue that migration networks lower the costs of migrating, hence facilitate migration. Most recently, Rapoport et al. (2017) [85] study social mixing in the destination country. They develop a framework to determine the effects of migration on cultural change. The authors aim at finding answers to the question whether migration spurs cultural convergence or divergence. The authors argue that cross-country interactions in a globalized world initiate not only an economic but also a cultural exchange. Both theoretically and empirically, the authors find evidence for bilateral cultural convergence. Whereas Rapoport et al. (2017) [85] explain cultural change as a result of increased human mobility, in this research paper we argue that cultural factors are also an important parameter in the emigration decision itself. Blumenstock, Chi, Tan (2019) [24] further explore the determinants of migration decision by focusing on the specific topology of the network. They try to disentangle whether the social network is used as a social support or for information diffusion. They use location data from mobile phone activity in Rwanda over a five-year period for roughly one million individuals, in order to build a detailed picture of their social network before and after migration. Their work confirms the hypothesis according to which people are more likely to migrate to destinations where they have more social contacts and are less likely to leave their original place of residence when they belong to larger social networks. By considering household ties with respect to a geographic network and a social network, we extend the growing literature on migration networks. This literature typically considers one single type of interaction modeled by one network. Yet, influence between households' migration decisions can depend on the particular type of a tie they have among each other.

Third, this paper is related the theoretical and empirical literature that studies peer effects in networks. Our model is embedded in the theoretical literature that studies games on networks; see Bramoullé and Kranton (2015) [32] for a survey. In their seminal paper, Ballester, Calvo-Armengol and Zenou (2008) [10] introduce a class of games with linear quadratic payoff functions. In their model, actions are efforts. There are local complementarities, that is the actions of two directly connected players are strategic complements. Furthermore, they introduce a global substitutability effect, where the increase of the overall level of effort of all players reduces the level of effort of a given player.⁵ The authors show that optimal actions are proportional to an eigenvector centrality measure. Our paper extends this litera-

⁵Their model is framed in terms of crime. The more the local circle of friends of a given individual, engages in crime the more likely that individual will engage in crime. However, if the global level of crime becomes very high in their city then it pays less for a given individual to engage in crime because there are less houses to steal...

ture by formalizing a game where players' actions are both *local* strategic complements and *local* strategic substitutes. We do so by modeling interactions of households in two networks with the same set of nodes but with different linkage patterns: the geographic network with households properties where migration decisions are local strategic substitutes and the social network where migration decisions are local strategic complements.

There exists many centrality measures for one-layer networks that capture different aspects depending on the type of interaction considered between agents (see Wasserman and Faust (1994) [97] or Jackson (2008) [61]). Nevertheless, the study of centrality measures for multi-layered networks or multiplex networks is still at its very stages. The difficulty lies in finding a relevant definition for centrality when agents can have multiple types of relations. Two notable approaches are: (i) balance theory where ties are signed (see Cartwright and Harary (1956) [36] and Bonacich and Lloyd (2004) [28]) and (ii) block-modeling analysis, where the adjacency matrix is sorted in a way such that zero-blocks appear (see Boorman and White (1976) [29]). But collapsing multiple layers in one network with positive and negative weights for links, oversees the importance or influence that a given agent can have in one of the layered networks, representing each a different context.

Finally, this paper is related to the recent literature around cultural identity and its influence on preferences and human behavior. Cultural identity is tricky to capture, yet has been shown to influence preferences and behavior significantly (see e.g. Akerlof and Kranton (2000) [4]; Hoff and Pandey (2006) [59], Benjamin, Choi and Strickland (2010) [16], Benjamin, Choi and Fisher (2016) [15], Bisin et al., (2016) [20] and Fouka (2020) [47]). In the context of occupational mobility, Munshi and Wilson (2011) [81] examine the role played by *local* identity. They provide evidence from historical data about migration in the Midwest in the 19th century. In their paper identity is transmitted by an institution: the church. They rely on the participation of early immigrants in different church communities (e.g. Catholic or Lutheran) as an indicator of identity. In this analysis we rely on two potential indicators to speak about cultural identity: we use the first name to account for inherited cultural preferences and the migration history to Austria or Germany to account for personally acquired cultural preferences.

Outline. The remainder of the paper is organized as follows. Section 3.2 provides an overview of the historical context of this paper. In section 3.3, we introduce our model of strategic migration with peer effects. In section 3.4, we provide empirical evidence, describe the data and the empirical strategy. Section 3.5 concludes. All figures and tables are placed after the conclusion.

3.2. Historical context

At school and in school essays, the “noi” was simply: “we, the Italians”, and it was forbidden to mean anything else. (...) It was a foreign “We”, that had annexed ours, yet unable to extinguish it. A selfish, jealous “We”, that, like God the Father, tolerated no other “We” beside himself, in himself, and among himself. It was the “We” of a totalitarian national community. This all was “Us” back then. A confusing human landscape, reflection of a confused time.

(translated from: Gatterer, C., 2003, p. 17 [51])

The history of South Tyrol in the 20th century evolves around the struggle of the German and Ladin speakers for recognition and preservation of their cultural identity, and the need for a bridge between the province and the Italian State, Alcock (2001) [5]. While initially the question evolved around being re-annexed to Austria, once this possibility was ruled out, the population started fighting for autonomy within the Italian State. This autonomy aimed at creating an environment in which the three linguistic groups (German, Ladin and Italian speakers) could live in harmony among each other and with the Italian State, while preserving and promoting their own culture. The struggle was long and persistent, not always peaceful, and took the form of several different treaties at the national and international level. On the path towards being one of the wealthiest regions in Europe today, where the three cultures and languages coexist in respect and harmony, numerous lessons were learned with regard to the creation of an extended sense of identity.

With the end of World War I, Italy received from Austria the southernmost part of Tyrol. Fascist policy implied the assimilation of all cultural minorities of the Italian territory. Ettore Tolomei, an official from the province of Trento, was responsible for doing so in South Tyrol. Without any transitory path, Italian was announced to be the only official language in South Tyrol - in all public as well as private establishments and institutions. The German language was abolished, written as well as oral. Whoever did not speak Italian were dismissed from their post, and Italians were sent to replace them. Street signs, town names and last names were italianized. The only teaching language in school was Italian.⁶ Nonetheless, the German and Ladin population did not bend in front of the Italian State, even though they were treated like foreigners in their own homeland. A famous example are the so-called “Catacomb” schools: underground schools where children were taught German.

This is why, almost 20 years later, Fascist Italy had still not succeeded in assimilating the South Tyroleans. Adolf Hitler and Benito Mussolini met in Berlin in 1939 and decided

⁶Only religious institutions were allowed to teach in German.

to force the heads of households of the German speaking population in Italy to choose whether they would want to remain in their homeland but abandon their culture and accept *Italianization*, or leave their homeland but follow their language and culture. Under relentless propaganda from Fascist and Nazi organizations, who respectively defined opters (people who opted to leave) and stayers as traitors. More than 80% of the population decided to leave, by 1939. In the end, however, only around half of the opters left, with (initially) no return option. The "opters" were relocated in different parts of Germany, mostly southern Germany then (Austria nowadays). Figure 3.4 shows the net migration in different parts of South Tyrol from 1940 to 1955, and we observe a large negative balance in 1940, which gradually increases to a positive balance in the later years. The only outlier is the capital town of the region: Bolzano. We know from historical evidence that a high share of Italians moved to Bolzano, in order to build new industries, rather than the remote villages. Figure 3.5 shows net migration variation across regions in South Tyrol, in 1940. Here again we observe Bolzano as outlier.

Despite the efforts by the South Tyroleans to return South Tyrol to Austria after WW II, it remained a part of Italy. However, the result of negotiations was the Gruber-DeGasperi Agreement signed on September 5th, 1956, which assigned the first legal autonomy to the province. Austria had, and still has, the right to act as the 'minority's' protecting power in case of violation of the clauses.

3.2.1. Migration decision: The procedure

The decision making process of the population eligible for opting was two-fold. The first decision was whether or not to opt into leaving, and the second was whether to effectively leave. The procedure was designed as follows. Before the 31st of December 1939 - so only six months after the agreement was signed - each head of household had to decide for the family whether to leave to Germany or to stay in South Tyrol (Italy).⁷ The head of household went to the local municipality to fill out a form: a red form for leaving, a white form for staying (*Verkaufte Heimat*, 1989). Individuals eligible to the option could also decide not to opt at all, which was equal to opting for staying in Italy. Once the deadline was over, as second step every opter had to go to the local office of the ADERSt, where they had to officially renounce to Italian citizenship and request German citizenship (hereafter called "Moving request files"). In the time to follow this procedure, the authorities proceeded to the evaluation of the value of the property and assets of the opters who requested to leave.

⁷"Head of household" was the male head of household, by the time. The decision was, hence, taken by him for the entire family. Exceptions were adult unmarried women, divorced women and widowed women.

In our analysis, we use the ADERSt moving request files. These files were carefully handled and well organized, partly because they were used for statistical and control reasons regarding the Option in South Tyrol in the Third Reich (Alexander et al. 1993 [6]). In our data we have two pieces of information which reflect the first and second decision, respectively. We know the exact date of when the opters officially renounced to Italian citizenship at the ADERSt office and requested German citizenship, and we know whether or not the opter effectively left. The first piece of information shows the willingness-to-leave or intentions-to-emigrate, whereas the second piece of information shows the actual migration status. Figure 3.6 summarizes this step-wise emigration procedure, with the corresponding time period and important drivers for the respective decision.

3.3. The model

Following the historical event, N households participated in the Option-Agreement by filling out application forms at their local municipality, yet only around half effectively left South Tyrol. The present paper attempts to explain the contribution of cultural, social and economic factors in the decision making process of emigration. To that end, we formalize an emigration game between strategic households that report an intention-to-emigrate. Each household has a cultural type and can interact in two different contexts with other households: the social network and the geographic network.

Based on the predictions of the model, we provide in section 3.4 empirical evidence where the intention-to-emigrate is proxied by the application date. We normalize this variable between zero and one, one reflecting the earliest application date and zero the latest.

3.3.1. Set-up

Let $N = \{1, \dots, n\}$ be a set of households. Each household $i \in N$ has a cultural type $\tau_i \in [0, 1]$. A household with cultural type $\tau_i = 1$ has fully Germanic culture, while $\tau_i = \epsilon$ means that they are german speaking but with a very low level of cultural attachment towards the Germanic culture. The decision variable of each household is denoted by y_i and takes values in \mathbb{R} . Actions are interpreted as a report of the intention-to-emigrate. The vector of actions is denoted by $\mathbf{y} = (y_1, \dots, y_n) \in \mathbb{R}^n$.

The social network. Each household is represented by a node i in a social network, represented by the adjacency matrix S with typical entries $s_{ij} \in \{0, 1\}$. The existence of a social tie between two households $i \neq j \in N$ is denoted by $s_{ij} = 1$ and $s_{ij} = 0$ otherwise. The social circle of each household and their number of acquaintances are respectively denoted by $N_i(S) = \{j \in N, s_{ij} = 1\}$ and $d_i(S) = |N_i(S)|$.

Geographic network. The place of residence of household $i \in N$ corresponds to node i in a geographic network, represented by the adjacency matrix G with typical entries $g_{ij} \in \{0, 1\}$. When two households $i \neq j \in N$ are neighbors then $g_{ij} = 1$, otherwise $g_{ij} = 0$. We denote by $N_i(G) = \{j \in N, g_{ij} = 1\}$ the neighborhood of individual $i \in N$. The local connectivity of each household within the geographic network G is simply their number of neighbors denoted by $d_i(G) = |N_i(G)|$. Finally, each location in the geographic network G has a value $v_i \in \mathbb{R}_+$, where $v_i = \epsilon$, for small $\epsilon > 0$, means that individual $i \in N$ does not own their place of residence and any $v_i > \epsilon$ indicates that the household owns the place of residence which is valued at v_i . For simplicity we make the following assumption. It corresponds to a normalization of the values of property.

Assumption 3 *The value of a property at location $i \in N$ is $v_i \in [0, 1]$.*

3.3.2. The game

The network structures S and G and the cultural types are assumed to be common knowledge across households. Each Household simultaneously chooses an action $y_i \in \mathbb{R}$ in order to maximize their net migration gain, taking into account both social and economic externalities. This total payoff is the sum of payoffs of two games played simultaneously on both networks. We summarize below both games, provide the payoff function for each and the best-response of households to their neighbors' actions. The equilibrium concept we use for this complete information game is Nash equilibrium.

Friendship game. For a network S , given cultural types $\tau_i \in [0, 1]$, each household $i \in N$ chooses $y_i \in \mathbb{R}$ that minimizes the following quadratic cost function:

$$c_{i,s}(y_i, y_{-i}, S) = (1 - \alpha)(y_i - \tau_i)^2 + \alpha \sum_{j \in N} s_{ij}(y_i - y_j)^2 \quad (3.1)$$

where the parameter $\alpha \in (0, 1)$ is exogenous and corresponds to the weight a household gives to matching own cultural type. The best-response y_i^* of household $i \in N$ to other households' actions in the friendship game is given by:

$$y_i = \frac{(1 - \alpha)\tau_i + \alpha \sum_{j \in N} s_{ij}y_j}{(1 - \alpha) + \alpha \sum_{j \in N} s_{ij}} \quad (3.2)$$

Independently of the economic factors, households are positively influenced by the migration intentions within their social circles. If most of the social acquaintances of a given household, intend to emigrate the household is more likely to emigrate, otherwise by staying behind they lose their social capital. That is, reports of emigration intentions are strate-

gic complements. Hence there is a compromise between matching own cultural type and matching neighbors actions to avoid the loss of social capital.

Lemma 3.1 For $\alpha \in (0, 1)$, the friendship game has a unique Nash equilibrium given by:

$$\mathbf{y}^f = (I - \alpha\beta S)^{-1}(1 - \alpha)\beta D_\tau \mathbf{1}_n$$

where $\beta = ((1 - \alpha)I + \alpha D_S)^{-1}$ and $D_S = \text{diag}(\sum_{j \in N} s_{1j}, \dots, \sum_{j \in N} s_{nj})$, $D_\tau = \text{diag}(\tau_1, \dots, \tau_n)$ and $\mathbf{1}_n$ a column vector of ones with n rows.

Proof. See Appendix 3.6.1. ■

Whenever the cultural types are not homogenous then households never end up with exactly the same emigration intentions. In this case, peer effects can increase or decrease the emigration intentions and the way households are affected by neighbors will depend on their network position in S . Clearly, with a connected network S and a uniform level of cultural attachment, say $\tau_i \equiv 0.5$ for all $i \in N$, reports of emigration intentions are identical across households. That is, $\forall i \in N$, $y_i \equiv 0.5$ when $\tau_i \equiv 0.5$.

Chicken game. Let $\delta \in (0, 1)$ be a discount factor for the value of property, which accounts for depreciation. The payoff for a given household $i \in N$ in the game played on the geographic network G , when choosing action $y_i \in \mathbb{R}$ is given by:

$$v_{i,g}(y_i, y_{-i}, G) = \frac{\tau_i}{1 + v_i} y_i - \frac{1}{2} y_i^2 - \delta \sum_{j \neq i \in N} \frac{g_{ij} v_j}{d_i(G)} y_j y_i. \quad (3.3)$$

The above payoff function can be interpreted as the net economic value of emigration. The higher the value of own property and the lower the gain from migrating. When the value of own property is close to zero and the household is geographically isolated or the values of neighbors properties are zero, then the gain from migration only depends on cultural type. The cost of migrating is represented by an average cost of *forgone* economic opportunities of purchasing neighbors' properties (if any). Independently of the social interactions, households are more likely to want to remain in the homeland, if ownership opportunities open-up because their neighbors have the intention to emigrate. Hence, actions are strategic substitutes in the geographic network. Given the neighbors' actions, each household $i \in N$ best-responds in the following way:

$$\hat{y}_i = \frac{\tau_i}{1 + v_i} - \frac{\delta}{d_i(G)} \sum_{j \neq i \in N} g_{ij} v_j y_j^*. \quad (3.4)$$

Notice that it is possible that the optimal decision \hat{y}_i becomes negative for certain combinations of $(\tau_i, v_i) \in [0, 1] \times [0, 1]$ and geographic network structures G . This occurs typically for households with low cultural types and high connectivity in G .

3.3.3. Equilibrium outcomes

Now we study emigration intentions of households when playing both games simultaneously. Using (3.1) and (3.3), the total payoff of a given household in the migration game is:

$$\pi_i(y_i, \mathbf{y}_{-i}; S, G) = v_{i,g}(y_i, \mathbf{y}_{-i}; G) - c_{i,s}(y_i, \mathbf{y}_{-i}; S). \quad (3.5)$$

This total payoff function is the difference between the economic value of migration and the loss of social capital. Recall that for all $i \in N$, $(\tau_i, v_i) \in [0, 1]^2$ and $\alpha \in (0, 1)$. Each household best-responds to neighbors' actions as follows:

$$y_i^* = \frac{1}{\kappa_i} \frac{\tau_i(1 + 2(1 - \alpha)(1 + v_i))}{(1 + v_i)} + \frac{1}{\kappa_i} \sum_{j \neq i \in N} (2\alpha s_{ij} - \frac{\delta g_{ij} v_j}{d_i(G)}) y_j^*, \quad (3.6)$$

where $\kappa_i = 1 + 2(1 - \alpha) + 2\alpha \sum_{j \neq i \in N} s_{ij}$. From equation (3.6), intentions-to-emigrate are fully determined by the cultural type, for any household who is isolated in both networks and has no property. Recall that the α parameter is the weight each household places on matching other households' migration decisions in the social network. For values of α that are not too small, when two households are connected in both the social and the geographic network, then they positively influence each other's decision making process. But this effect becomes weaker for higher property values. Hence, the positive influence⁸ exerted by household j on household i in the social network, is diminished by the negative influence exerted by the same household in the geographic network. Precisely, the second term in the right hand side of equation (3.6) shows the impact of linkage patterns in both networks on the migration decision of a given household. This impact could be positive, negative or zero if the two considered households are disconnected in both networks.

Proposition 3.1 *Consider N households embedded in social network S and geographic network G .*

- (i) *If household $i \in N$ is socially and geographically isolated, and has no property that is $\sum_{i \neq j \in N} s_{ij} = \sum_{i \neq j \in N} g_{ij} = v_i = 0$, then $y_i^* = \tau_i$.*

⁸In the sense that migration decisions go in the same direction for a linked pair of households i and j , so if y_j increases then y_i increases and vice versa.

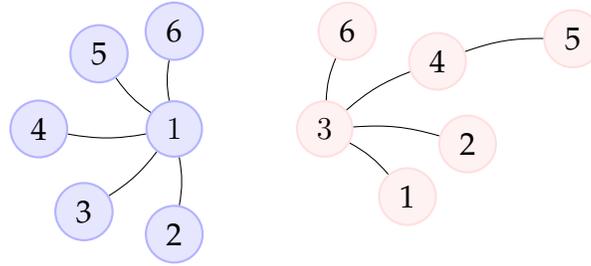


Figure 3.1: Geographic network G (left panel) and social network S (right panel).

(ii) If for all households $i \in N$, $\tau_i \geq \epsilon$ for ϵ and each household $i \in N$ has a least one neighbor in both networks S and G , then optimal migration decisions are :

$$Y^* = (I - \sigma M)^{-1} \sigma X D_\tau \mathbf{1}_n, \quad (3.7)$$

where $X = (I + D_v)^{-1} (I + 2(1 - \alpha)(I + D_v))$, $M = 2\alpha S - \delta D_G^{-1} G D_v$, $\sigma = (I + 2(1 - \alpha)I + 2\alpha D_S)^{-1}$ and D_v, D_τ, D_G, D_S are diagonal matrices with respectively diagonal terms property values, cultural types, degrees in the geographic network and degrees in the social network.

Proof. See appendix 3.6.2. ■

The convergent matrix σM in equation (3.7) is a directed and weighted adjacency matrix, where the weights could be either positive or negative. The entry on row i and column j is the impact that the emigration intentions of household j has on the emigration intentions of household i , when the two types of interactions on both networks are taken into account. In particular, we are interested in understanding the importance of the migration decisions of other households on the decision of a given household, when taking into account interactions on both networks.⁹

To give some insight about the role of centrality, we provide an example of two households who are *central* in only one of both networks. We compare their net migration gains given by (3.5) for neighborhood property values and cultural types of neighbors that are either low or high.

Example: der Kapitalist und der Priester. Consider the geographic network G on the left panel of figure 3.1 and the social network S on the right panel. Assume that the weight assigned to matching neighbors' action is $\alpha = 1/2$, the property value discount factor is $\delta = 0.9$

⁹This exercise is relevant for the design of migration policies, because it helps in addressing the question of how a planner can incentivize agents when local externalities are at play.

and that there are six households. The absence of links in the geographic network can indicate for example that there is a mountain separating properties 2 and 3. Household 3 (the Priest) is the most central¹⁰ in the social network S , while household 1 (the capitalist) is the most central in the geographic network G . We look at the net migration gains given by (3.5) of households 1 and 3, when all households take the optimal actions. In particular, we look at how this net migration gain varies as a function of own cultural type and different level of own property values, when cultural types and neighborhood property values of the remaining households are either high or low. The top right panel of figure 3.2 shows that when cultural types of social contacts are low and their neighborhood property is high, even when the priest has a very high cultural attachment the net migration gain is negative. Hence, the effect of matching the actions of social contacts is higher than the effect of mismatching their actions, for a very central player in the social network. Conversely, the top right panel of figure 3.3 shows that due to the high centrality of household 1 in the geographic network, when cultural types of neighbors are low (driving down migration intentions) then the net migration gains are positive for almost all levels of cultural attachment and all property values. Here, the substitutability of actions is at play and even though the social contacts of household 1 have a low level of cultural attachment, household 1 can have a positive migration gain for strictly positive levels of cultural attachment and own property value that is not at the maximum level.

3.4. Empirical evidence

We provide supporting evidence for our model, about self-selection into migration during a historical migration choice experiment between Nazi Germany and Fascist Italy. This historical episode allows us to have unique insight in the emigration decision making process. Intentions-to-emigrate are proxied by the application date for emigration of each opter, where the application period ranges from fall 1939 until early 1945. The rich data set allows us to build several indicators to account for the level of cultural attachment to the Germanic culture such as “inherited” cultural ties, using the Germanic-ness of the first name of the opter. Furthermore, we are able to identify the neighborhood property of each household by matching the addresses to a geographic map. Following the findings of our model, we expect cultural types and emigration intentions of social contacts to have a positive contribution on the outcome variable, while own property and neighborhood property to have a negative contribution.

¹⁰In the sense of centrality measure for one-layer networks, such as degree centrality, closeness centrality, Betweenness centrality and Bonacich centrality.

3.4.1. Data

The main data source for this project is the detailed documentation of the Option Agreement procedure, held by the National Archive of the Province of Bolzano. The Archive holds four sets of files related to the Option Agreement. The first set of files are the so-called “option requests”. These are the files withheld by the Italian authority designated to handle the Option. There exist approximately 140 000 option requestss from 1939, that provide socio-demographic and economic information on the opter and their family, as well as information on the timing of departure and destination. For each option request there is a corresponding personal file (second set of files) by the German immigration office responsible for the Option (“ADERSt: Deutsche Anund Rueckwanderungsstelle”). These files provide, again, detailed socio-demographic and economic information on the opter. Whereas the documentation in each file varies greatly, these ADERSt-files contain one document that is common throughout all files: the so-called “Abwanderungsantrag”, or “Request to leave”. It summarizes, on one form, the most important socio-demographic characteristics of the opter and their spouse and children as well as their parents. We have information regarding birth date, birth place and detailed address of the opter as well as their entire family, the opter’s learnt profession and the current profession, family status, religion, military status, health status, police records and migration history. From the ADERSt files we can furthermore trace information regarding the financial status of the opter (Did they have any significant assets or property? What is the value of the property?), the migration status (Did the person leave or not leave?) and the return migration status (Did the person or their child come back after 1948?). The third set of files, again linked through name and record number of the individual, are the financial statements for each opter that allowed identifying the exact value of all the physical and financial assets of the opter. The monetary equivalent to the assets, determined by two separate commissions (an Italian and a German one), was transferred to the individual, who agreed to sell everything they had and move to the Third Reich, with (initially) no return option. Finally, there is a fourth set of files. There are around 75 000 Back-Option Requests, from 1948, when the South Tyroleans were granted the legal right to move back. Children of opters had to prove their family status, in order to be allowed to come back to Italy.

The requests are held in boxes which we randomly drew, stratified by commune and representative for the leaving request filing date. We digitized a 2% sample, which gave clean files for 2338 “opters”, where each opter is the head-of-household.¹¹ See table 3.1 in

¹¹The files of the office of Bolzano are missing, it is unclear whether they were destroyed, lost or are locked away somewhere untraceable.

the appendix for more information. The digitization process was a highly sensitive procedure, since all the original files are still stored in the National Archive of Bolzano and cannot leave the archive. Several months of a daily, careful scanning with two research assistants was necessary in order to collect scans for the 2% sample. Afterwards, we tried several OCR¹² techniques to speed up the data entry procedure, however, due to the age of the files which produce low quality scans it was of no help. At the end, the data was encoded manually.

From the sample we successfully collected and digitized, we drew a first set of descriptive statistics, shown in Table 3.2. In our sample, 57% of all opters ended up effectively migrating - this corresponds to the approximations in the historic literature. We see that the age gap between migrants and non-migrants is relatively small (36 vs 41 years), with the migrant population being slightly younger. The percentage of women who migrated is also lower than women who did not migrate (27 vs. 35%). Previous military service seems to be similar across migrants and non-migrants, on average around 44% of all people in the sample have carried out previous military service. Furthermore, we see that individuals with a police record tend to be more likely to emigrate (8% vs 3%), and that individuals with illnesses migrate less frequently (10% vs 18%). A high percentage of individuals in our sample have previously migrated (this includes migration inside of Italy and Germany and Austria, as well as outside these three countries) - individuals who decided not to emigrate were even more likely to have migrated before (83% vs 97%). In terms of children in the household, families with children tended to be less likely to emigrate (22% vs. 34%). In terms of family status, opters who were single, were relatively more likely to emigrate than married, widowed or divorced opters. Furthermore, capitalists ("notable property owners") and opters of no income ("out of the labor market") were less likely to emigrate with respect to workers who were employed. Whereas in this section we give a general description of the data, in the next section we describe the variables we use and indicators we create, more in detail.

3.4.2. Definition of proxy variables

We examine the selection-into-migration of agents who have opted into leaving and are at the bulk of taking the decision of whether to leave their home country for good or not. Relying on the theoretical model developed in the previous section, in this section we define our proxy variables.

¹²Optical character recognition.

First, we define the outcome variable, which proxies the migration intention. We rely on the application for emigration of each opter. More precisely, we rely on the application date, where the application period ranges from fall 1939 until early 1945. We then normalize this variable between zero and one, one reflecting the earliest application date and zero the latest. The average value of variable "time of application for emigration" is 0.857, which corresponds to the 14th June 1940, with a standard deviation of 0.151, or 300 days. Furthermore, we introduce a robustness specification, where we look at the actual emigration status of the opters, a binary variable equal to one if the opter ended up effectively emigrating and zero otherwise. The mean emigration probability is 0.572 with a standard deviation of 0.495.

In the following sub-sections we introduce our explanatory variables, most notably our definition of the variable capturing culture, and the social and geographical networks.

Cultural type: degree of attachment to the German culture

In this analysis we rely on three variables that reflect potential cultural ties to Germany, introducing a fourth one for a sub-sample of opters with children. The variables explained in the following paragraphs aim at proxying *inherited* and *acquired* cultural preferences.

The first variable is opter i 's first name. Here we follow the literature of first names as a reflection of cultural preferences. This concept stems from sociology literature and was recently taken up by economists (see, e.g. Fryer Jr et al. (2004) [68]; Rubinstein et al. (2013) [86], Biavaschi et al. (2017) [19]). However, this literature proxies parents' cultural preferences which are reflected in their children's names. In our case, we rely on the the first names of adults. The first name of an adult, can be interpreted as their *inherited* cultural preferences, i.e. the preferences of their parents at the time they named them. In their lifetime, a person's cultural identity might have been subject to changes. To take this into account, the second variable we look into is opter i 's migration history to Austria or Germany, to account for personally *acquired* cultural preferences over lifetime.

Whereas it is reasonable to assume that all these indicators reflect cultural preferences to a certain extent, they might reflect a number of other factors as well, for example economic factors. Past migration history, for example, has likely affected opter i 's economic ties to Germany as well. The results, in particular the magnitude of the effects we find, are, hence, to be interpreted with caution.

Our first indicator of cultural ties to Germany is the first name of each opter i , the inher-

ited cultural preferences. Unlike other countries like the US, for Germany and Austria there is no official registry of the most popular first names in the early 20th Century. We therefore rely on other indicators of "German-ness" of a first name. We rely on anecdotal evidence that Germanic first names - in particular in the early 20th Century - were a strong reflection of attachment to the German ethnicity. There is little scientific evidence for this phenomenon, until Jesus Casquete in 2016 published an article in the *European Journal of Cultural and Political Sociology*, about "The importance of being Horst" which discusses the encouragement of Nazis to give Germanic first names to newborn children (especially boys). We use a list of Wikipedia listing all German first names with Germanic roots and use the Stata algorithm *reclink*¹³ to determine the proximity of the names of the opters in our sample to the Germanic names in the list. From this data, we create an index by quintiles. However, as we see in figure 3.7, there is little variation and frequency in the values in between the strict zero and one. For the sake of precision of our estimates, we prefer using a binary variable for our analysis: taking as zero all non-fully Germanic names, and one all fully Germanic names.

An example for this procedure is the following. Suppose we have three first names: Josef, Michael and Siglinde. According to the Stata "reclink" algorithm, the first name "Josef" has a zero percent correspondence to any Germanic name in our list, so the index will be zero, on a possible range between zero and one. The first name "Michael" has a similarity index equal to 0.5231, the similarity could stem from the Germanic first name "Emich". "Siglinde" has a similarity index of 1, since "Siglinde" is a Germanic first name. From this continuous variable, we take as zeroes all non-fully Germanic names, and ones all fully Germanic names. The "Germanic first name index" of Josef will be equal to 0, so will Michael's index, and Siglinde's index will be equal to 1.

Furthermore, we proxy *acquired* cultural preferences. We look at the migration history of each opter, more specifically, whether the opter has previously migrated to either Austria or Germany. We, thus, have a dummy variable equal to one if opter i has previously migrated to Austria or Germany, and equal to zero otherwise. Approximately 12% of all opters in our sample have previously migrated to Germany or Austria, with a standard deviation of 0.34.

From the "inherited", and "acquired" cultural preferences, i.e. the opter's first name, and their previous migration, we create a simple average, which gives us a variable between zero and one, with an average of 0.196 and a standard deviation of 0.279.

¹³The "reclink" command allows for a fuzzy merge. It uses record linkage methods and matches observations between two datasets where no perfect key fields exist.

Own property

The next indicator we introduce shows the property status. Here fore, we rely on two variables present directly on the AderSt records. First, the property status of the opter themselves, a binary variable equal to 1 if the opter has property and 0 otherwise, and second, the family property. Here, again, we rely on a variable equal to 1 if the opter's family has property and 0 otherwise. In the early 20th Century in the region, property was usually passed on directly through heritage, hence, family property is here considered future personal property. We, hence call "Personal property" a binary variable equal to 1 if the opter either has personal property or family property, or both, and 0 otherwise, with a mean of 0.464 and a standard deviation of 0.499.

Geographic network and social network

In addition to accounting for their own property, opters tended to consider the opportunity of buying close-by property, knowing that their neighbor opted for leaving. The next indicator we introduce shows, hence the neighborhood property status. Here fore, we create a binary variable equal to 1 if there is at least one neighbor who owns property, 0 otherwise. The mean of the variable "Neighborhood property" is 0.535, with a standard deviation of 0.498. Hence, we proxy the geographic network with opter i 's geographic neighbors property.

Furthermore we proxy the social network with family, geographic neighbors and peers from the same social group: "Capitalist", "Worker", and "No income". We create an index which shows the average emigration intentions of the social group of opter i in the village of opter i .

Control Variables

Our set of control variables includes the following variables: age, a dummy equal to one if the opter is female, and zero otherwise, a dummy equal to one if the opter has previously served in the military and zero otherwise, a dummy equal to one if the opter has a police record and zero otherwise, a dummy equal to one if the opter has declared illnesses and zero otherwise, dummy equal to one if the opter has previously migrated and zero otherwise, and finally a dummy equal to one if the opter has children and zero otherwise.

In addition, in order to account for possible delays in application deadlines caused by the administration officers, we add officer fixed effects to our model. Partly through image recognition algorithms, partly through manual image recognition, we identify common signatures of the officers who handled the ADERSt leaving request. In our sample, we identified a total of 88 different signatures, i.e. 88 different officers.

3.4.3. Empirical Strategy

We use the collected data described in section 3.4.1, in order to carefully proxy the variables we aim at estimating. In our model there are two channels through which peer effects are present: the social network and the geographic network. Our model can be written in the following form:

$$y_i = \beta_0 + \underbrace{\beta_1 \tau_i}_{\text{cultural type}} + \underbrace{\beta_2 v_i}_{\text{own property}} + \beta_3 \overbrace{\frac{1}{d_i(S)} \sum_{j \in N} s_{ij} y_j}_{\text{social network}} + \beta_4 \overbrace{\frac{1}{d_i(G)} \sum_{j \in N} g_{ij} v_j}_{\text{geographic network}} + \epsilon_i \quad (3.8)$$

where β_3 captures the social effect and β_4 captures the economic effect. From equation (3.8), we see that the migration decision of household i is affected by the average migration decisions of their neighbors in the social or/and the economic network. Conversely, the migration decision of i will affect the average migration decisions. To see this, pick any neighbor j of household i . The migration decision of j is affected by the mean decisions of their neighbors, namely household i . Hence household i affects the decision of j and j affects the decision of i through the average of decisions. Therefore, the error terms will be correlated with the mean decisions in equation (3.8) and using *OLS* can yield biased estimates.

To overcome the possible bias, we rely on the literature in the estimation of peer effects. In particular, we follow the estimation strategy proposed by Gaviria and Raphael (2001) [52]. In their setting, the authors assume that contextual effects are non-existent, hence the average background characteristics of individual i 's peers provide a sound instrument for the average peer behavior. In our case, we pick up their strategy, following the assumption that household i 's decision is affected by the average characteristics of their peers only through the peers' emigration intentions. There is one exception in the characteristics, which is the neighborhood property, that provides an economic incentive not to emigrate. We account for this characteristic separately in our model, see equation 3.8.

We carry out a *2SLS* approach in addition to our possibly biased *OLS* approach, relying on the average characteristics of household i 's peers to instrument the average peer emigration decision, namely cultural type, age, gender, previous military service, police record, illnesses, previous migration and children in the household, with a mean of 0.350 and a standard deviation of 0.161 between zero and one.

3.4.4. Results

In this section we present the results of equation 3.8, relying on the variables introduced in section 3.4.2. The regression table 3.3 shows the results from a regression of a set of ex-

planatory variables on the outcome variable “time of application for emigration”, which is a normalized variable between 0 and 1 of the application date, 1 being the earliest application date, 0 the latest. “Cultural type” reflects the cultural type of an individual. It is composed of a “Germanic first name” index and migration history to Germany or Austria, to account for “inherited” and “acquired” cultural preference. “Social network decision” reflects the emigration decision of the social networks, which include neighbors and individuals from the same social class (capitalist, worker, no income) in the same village. “Personal property” is a binary variable equal to one if the person or their family owns property, and zero otherwise. “Neighborhood property” is a binary variable indicating whether the individual’s neighbors own property. Columns 1 and 4 show results from specifications without the inclusion of a set of control variables, columns 2 and 5 show the results with inclusion of officer fixed effects, and columns 3 and 6 include officer fixed effects and a set of the control variables (age, female, previous military service, illnesses, previous migration and children in the household). Columns 1 to 3 show the results from an OLS specification. Columns 4 to 6 show the results of a 2SLS regression, relying on the average covariates of the social network as instrument (covariates cultural type, age, female, previous military service, illnesses, previous migration and children in the household), following the strategy introduced by Gavrila and Raphael (2001) [52]. All regressions are reported with robust standard errors. The bottom panel shows the Sanderson-Windmeijer (2016) [88] first-stage chi-squared and F statistics, which are tests of underidentification and weak identification of endogenous regressors.

Across all specifications, we see that a stronger cultural connection towards the German culture significantly anticipates the application day, showing, hence, a strong intention to migrate. The emigration decision of the social networks also significantly anticipates the application day. We see that personal property as well as neighborhood property, on the other hand, significantly delay the application day, hence showing a weaker intention to migrate. These results are in line with the propositions drawn from the model in section 3.3.

More precisely, we see that opters of a more German cultural type have a between 0.016 and 0.04 earlier application date, hence increasing the application time variable from the average of 0.86 to 0.90. These results might appear small in magnitude, but they are large when translated to days. A value of 0.04 of our indicator equals 79 days. This means that opters of a more German cultural type applied for emigration up to 79 days earlier than the average. Furthermore, we find that opters with property apply later, by between 0.011 and 0.04, and opters who have property in their neighborhood that they could possibly buy, apply between 0.08 and 0.013 later.

Looking at the average emigration decision of the social network, we find a strong effect of the migration decision of the group of peers on a given opter i 's application date. In particular in the 2SLS specifications, opters apply between 0.09 and 0.32 earlier, or up to two standard deviations (600 days) earlier.

Alternative explanations and robustness

After discussing our main results, we investigate the nature of the selection-into-migration in this peculiar setting: Was the emigration decision entirely voluntary, or was it partly determined by the participating government? There is some historical evidence which claims that some people were deported, whereas for others the leaving time was artificially delayed (Eisinger et al., 1989; Alexander et al., 1993). This means that, whether or not opters ended up moving or not, was not entirely their decision, but was partly decided by the German and Italian authorities. Initially, there were no boundaries as to whom could enter the German territory, but there is some evidence, see Alexander et al. 1993, p. 50, for example, that suggests that the German authorities in their idea of reuniting all Germans in one empire, followed the idea of reuniting those Germans who were of "unambiguous German race, healthy and with a morally and politically flawless character" (roughly translated from "eindeutig deutscher Volkszugehörigkeit, erbggesund, moralisch und politisch einwandfreien Charakters". On the other hand, Italian authorities had another idea of whom they mostly wanted to "get rid of": theoretically, everybody except for the farmers in the mountains whom they knew they could not replace with Italians. Practically, they wanted to send away the morally and physically questionable individuals first: people in prisons and people in hospitals and especially psychiatric clinics.¹⁴ There is little evidence on whether and how this happened practically, and how the individuals were selected in the end - according to German or Italian ideas - or if they were selected at all. To look into this possible alternative explanation, we check whether we find differential effects when accounting for those population groups, which according to Alexander et al. (1993) [6] were potentially "sampled in" by the Italian Government and "sampled out" by the German Government - i.e. the most vulnerable parts of the population in one specification, and the farmers in another specification.

Here fore we add two specifications, relying on a sample that excludes vulnerable individuals in one specification, and farmers in another specification. We define as 'vulnerable individuals' opters with a police record or an illness. We present two specifications, table 3.4 for the case of vulnerable individuals and table 3.5 for the case of farmers. We find that in

¹⁴Mentally ill individuals were deported to Germany and killed there; see Eisterer et al. (1989) [44], Alexander et al. (1993) [6].

the case of excluding vulnerable individuals, we find our main effects to be consistent with the effects we find in our main specification in table 3.3. In the case of excluding farmers, we find similar magnitude, yet less significance for some of the indicators. We find strong effects for the cultural variable and the social networks variable, however less significant effects for the property variables. This result is intuitive, since we exclude from the sample the farmers, i.e. the individuals who owned most property in early 20th Century South Tyrol and who would likely have been the most willing to extend their property through acquisition of neighborhood property.

Furthermore, we present our main results from table 3.3 using an alternative, binary, measure of the emigration decision, one if the opter effectively emigrated, zero otherwise. We present the results in table 3.6, and find that the results largely correspond. This result gives us further confidence in our main results.

Finally, we carry out a further robustness test, following Oster (2019) [84], and relying on the *psacalc* [83] command. Oster (2019) [84] introduces a novel approach to tackle omitted variable bias. Sometimes it is argued that a set of control variables can properly capture an omitted variable. However, this is difficult to prove and often the set of control variables do not perfectly capture an omitted variable. The methodology of Oster (2019) [84] and her routine *psacalc* [83] gives a parameter *delta* that shows by how much the unobservables would have to be more important to produce a zero treatment effect. The *delta* we find for each specification reflects the significance of our results and proves further robustness. In particular, we find for all our significant variables in the main specification, a *delta* value larger than two each. This threshold value suggests that the unobservables would have to be at least twice as important as the observables to produce a treatment effect of zero, for each of the variables we test. The results to the Oster (2019) test are represented in table 3.7.

3.5. Conclusion

This paper studies the mechanisms by which cultural, social and economic factors affect the intentions-to-emigrate. In particular it focuses on intentions rather than the binary decision to leave or stay. Reporting intentions reveals how individuals identify to a culture and how they are affected by peers because they are non-binding; while the binary decision of leaving or staying entails actual behavior hence overseeing migration motivations. We show that intentions-to-emigrate are positively affected by social contacts, but this effect is weakened when a given social contact is also a geographic neighbor. We further provide supporting empirical evidence for our model, thanks to a historical migration choice exper-

iment between Nazi Germany and Fascist Italy in 1939, following the loss of institutional power of the region's cultural and linguistic majority to a minority.

We are the first to digitize personal records of this Migration Option Agreement, which provide us a rich data base to gain insight into mechanisms of the migration decision that have yet to be examined in the academic literature. In line with the results from our theoretical model, our empirical evidence confirms the relevance and direction of cultural and social alongside economic factors in the emigration decision. Starting from this representative yet small sample, we are digitizing more data to improve statistical power and precision. This additional power can help explore further venues, for example understanding if we observe heterogeneity patterns in decision making.

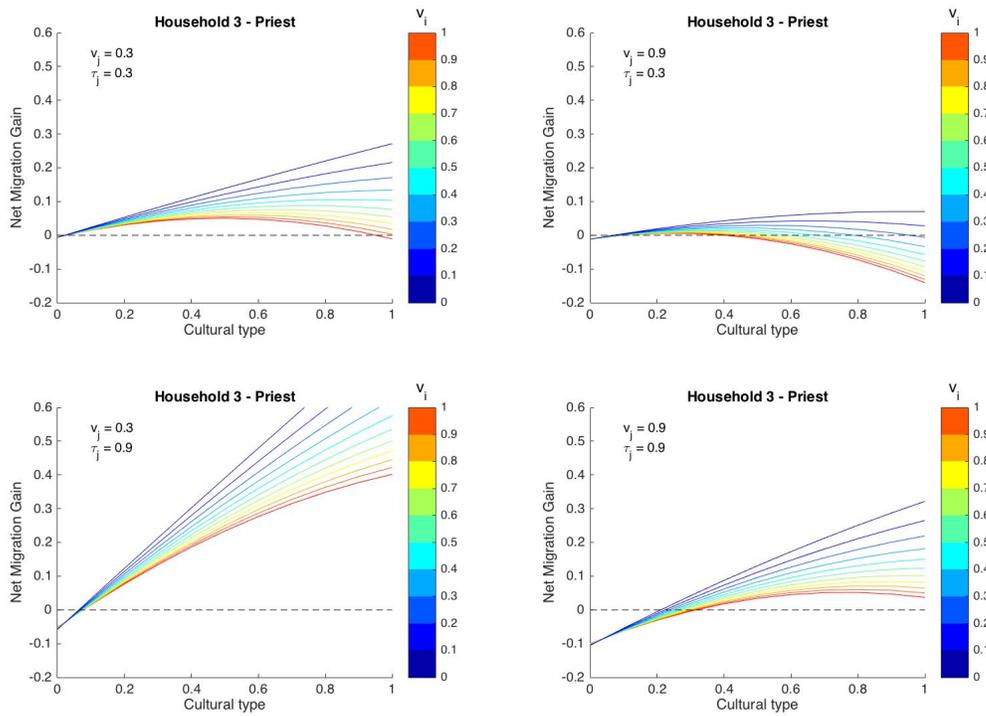


Figure 3.2: Net Migration cost of household 3 (Priest) for different levels of neighborhood property and cultural types of social contacts.

Table 3.1: Determination of the 2% sample (with ADERSt data Bolzano).

Responsible local office	Option Identifiers	German Opters	3%	2%
Zweigst. Meran	200000-239160	20001-23347	1175	783
Zweigst. Brixen	300000-334500	30001-30898	1035	690
Zweigst. Bruneck	400000-440021	40001-40889	1321	880
Hauptst. Opt.ber. im Dt Reich	700001-723105		693	426
N of identifiers = 141930				
N of potential observations 3% = 4258				
N of potential observations 2% = 2839				

Data source: National Archive of the Province of Bolzano

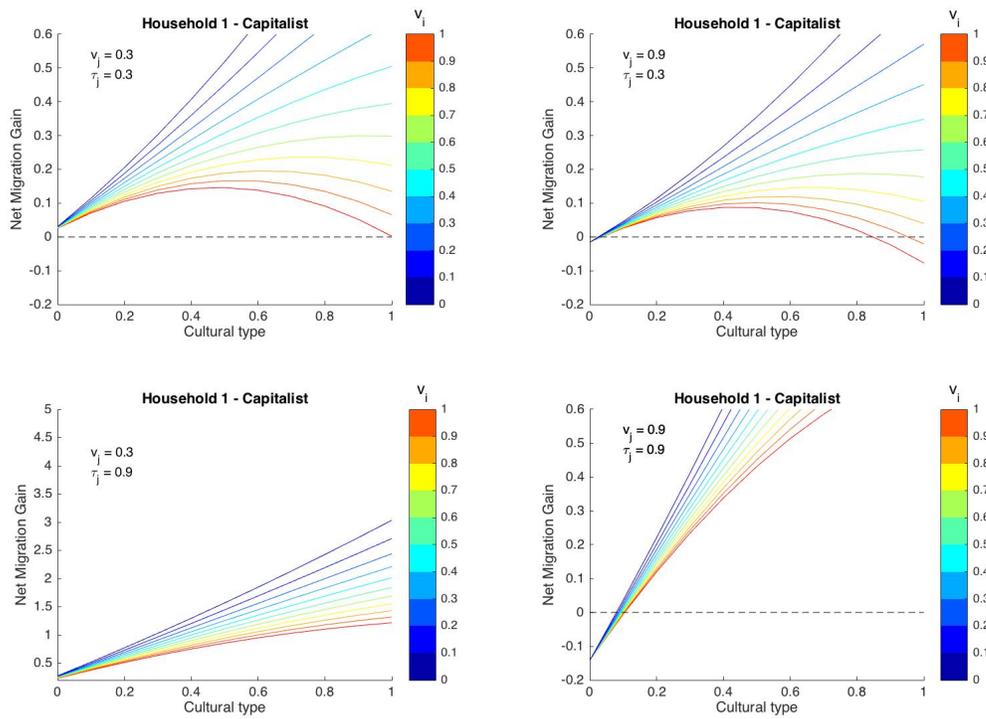
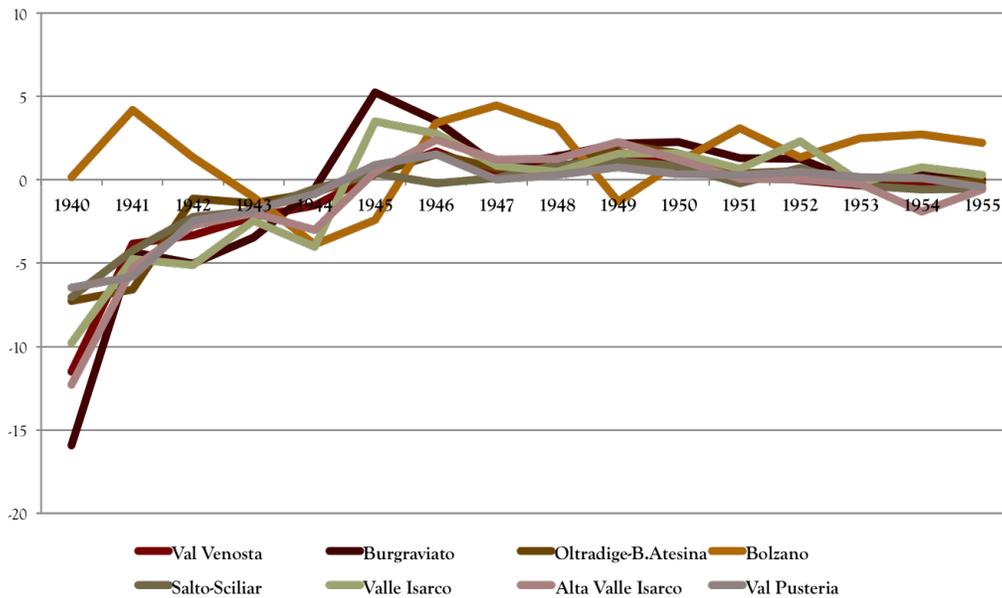


Figure 3.3: Net Migration cost of household 1 (Capitalist) for different levels of neighborhood property and cultural types of social contacts.

Figure 3.4: Net migration in South Tyrol from 1940-1955.



Data source: ASTAT - Institute for Statistics of the Province of Bolzano.

Figure 3.5: Net migration in percentage of population in 1940.



Data source: ASTAT - Institute for Statistics of the Province of Bolzano.

Figure 3.6: Step-wise emigration procedure.

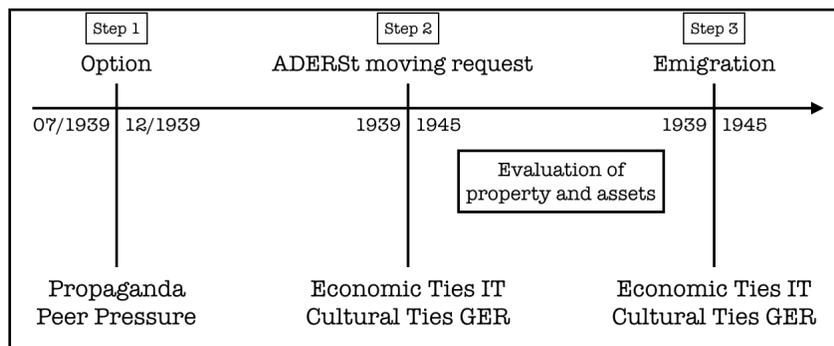


Table 3.2: Descriptive statistics of a sample of 2338 heads-of-households, by leaving status.

Variable	Emigrated: 57%					Did not emigrate: 43%				
	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
Age	1328	36.40512	17.03187	15	88	992	41.81452	15.94427	13	82
Female	1336	.2754491	.4469076	0	1	1001	.3836164	.4865093	0	1
Prev. military service	1332	0.452703	.4979449	0	1	986	0.432049	.4956125	0	1
Police record	1329	0.082017	.2744932	0	1	986	0.030426	.1718435	0	1
Illness	1329	0.108352	.3109416	0	1	986	0.186613	.3897979	0	1
Prev. migration	1337	.8332087	.3729289	0	1	1001	.968032	.1760029	0	1
Children in HH	1336	.2193114	.413935	0	1	1001	.3396603	.4738306	0	1
Divorced	1333	.0247562	.1554395	0	1	999	.013013	.1133867	0	1
Married	1333	.2175544	.4127375	0	1	999	.3233233	.4679792	0	1
Single	1333	.6496624	.4772547	0	1	999	.5295295	.4993772	0	1
Widowed	1333	.108027	.310531	0	1	999	.1341341	.3409671	0	1
Out of labor market	1295	.1389961	.3460761	0	1	962	.1507277	.3579693	0	1
Notable property	1295	.0996139	.2996002	0	1	962	.3378378	.473219	0	1
Currently working	1295	.76139	.4263985	0	1	962	.5114345	.5001292	0	1
Current job = learnt job	1315	0.701901	.4575972	0	1	979	0.797753	.4018809	0	1
Property owner	1229	.1334418	.34019	0	1	928	.419181	.4936911	0	1

Data source: National Archive of the Province of Bolzano

Table 3.3: Strategic cultural emigration and peer effects.

	Time of application for emigration					
	OLS specifications			2SLS specifications		
	(1)	(2)	(3)	(4)	(5)	(6)
Cultural type	0.040*** [0.012]	0.024*** [0.008]	0.021*** [0.008]	0.016 [0.012]	0.017** [0.008]	0.016** [0.008]
Social network decision	0.072*** [0.011]	0.017*** [0.006]	0.017*** [0.006]	0.328*** [0.032]	0.126*** [0.022]	0.093*** [0.021]
Personal property	-0.039*** [0.006]	-0.016*** [0.004]	-0.015*** [0.004]	-0.019*** [0.007]	-0.011** [0.004]	-0.011*** [0.004]
Neighborhood property	-0.013** [0.006]	-0.012*** [0.004]	-0.010*** [0.004]	-0.008 [0.007]	-0.010** [0.004]	-0.009** [0.004]
<i>Mean value of "application"</i>	0.858	0.859	0.856	0.858	0.859	0.856
Officer fixed effects	NO	YES	YES	NO	YES	YES
Control variables	NO	NO	YES	NO	NO	YES
Observations	2,327	2,197	2,167	2,327	2,197	2,167
F-statistic				216.54	163.86	179.13
F-statistic (p-value)				0.000	0.000	0.000
SW Chi-sq				217.00	168.85	185.11
SW Chi-sq (p-value)				0.000	0.000	0.000

Standard errors in brackets

*** p<0.01, ** p<0.05, * p<0.1

The regression table shows the results from a regression of a set of explanatory variables on the outcome variable "time of application for emigration", which is a normalized variable between 0 and 1 of the application date, 1 being the earliest application date, 0 the latest. "Cultural type" reflects the cultural type of an individual. It is composed of a "Germanic first name" index and migration history to Germany or Austria, to account for "inherited" and "acquired" cultural preference. "Social network decision" reflects the emigration decision of the social networks, which include neighbors and individuals from the same social class (capitalist, worker, no income) in the same village. "Personal property" is a binary variable equal to one if the person or their family owns property, and zero otherwise. "Neighborhood property" is a binary variable indicating whether the individual's neighbors own property. Columns 1 and 4 show results from specifications without the inclusion of a set of control variables, columns 2 and 5 show the results with inclusion of officer fixed effects, and columns 3 and 6 include officer fixed effects and a set of the control variables (age, female, previous military service, illnesses, previous migration and children in the household). Columns 1 to 3 show the results from an OLS specification. Columns 4 to 6 show the results of a 2SLS regression, relying on the average covariates of the social network as instrument (covariates cultural type, age, female, previous military service, illnesses, previous migration and children in the household). All regressions are reported with robust standard errors. The bottom panel shows the Sanderson-Windmeijer (SW) first-stage chi-squared and F statistics, which are tests of underidentification and weak identification of endogenous regressors.

Table 3.4: Strategic cultural emigration and peer effects: excluding vulnerable population.

	Time of application for emigration: excluding people with police records or illnesses					
	OLS specifications		2SLS specifications			
	(1)	(2)	(3)	(4)	(5)	(6)
Cultural type	0.051*** [0.013]	0.034*** [0.008]	0.029*** [0.009]	0.029** [0.013]	0.027*** [0.008]	0.024*** [0.009]
Social network decision	0.087*** [0.014]	0.022*** [0.007]	0.023*** [0.007]	0.327*** [0.035]	0.138*** [0.025]	0.101*** [0.022]
Personal property	-0.035*** [0.006]	-0.015*** [0.005]	-0.014*** [0.005]	-0.018** [0.008]	-0.011** [0.005]	-0.011** [0.005]
Neighborhood property	-0.014** [0.006]	-0.013*** [0.004]	-0.010** [0.004]	-0.009 [0.007]	-0.011*** [0.004]	-0.009** [0.004]
<i>Mean value of "application"</i>	0.865	0.866	0.866	0.865	0.866	0.866
Officer fixed effects	NO	YES	YES	NO	YES	YES
Control variables	NO	NO	YES	NO	NO	YES
Observations	1,881	1,776	1,748	1,881	1,776	1,748
F-statistic				171.04	128.4	138.95
F-statistic (p-value)				0.000	0.000	0.000
SW Chi-sq				171.50	132.81	144.06
SW Chi-sq (p-value)				0.000	0.000	0.000

Standard errors in brackets

*** p<0.01, ** p<0.05, * p<0.1

The regression table shows the results from a regression of a set of explanatory variables on the outcome variable "time of application for emigration", which is a normalized variable between 0 and 1 of the application date, 1 being the earliest application date, 0 the latest. "Cultural type" reflects the cultural type of an individual. It is composed of a "Germanic first name" index and migration history to Germany or Austria, to account for "inherited" and "acquired" cultural preference. "Social network decision" reflects the emigration decision of the social networks, which include neighbors and individuals from the same social class (capitalist, worker, no income) in the same village. "Personal property" is a binary variable equal to one if the person or their family owns property, and zero otherwise. "Neighborhood property" is a binary variable indicating whether the individual's neighbors own property. Columns 1 and 4 show results from specifications without the inclusion of a set of control variables, columns 2 and 5 show the results with inclusion of officer fixed effects, and columns 3 and 6 include officer fixed effects and a set of the control variables (age, female, previous military service, illnesses, previous migration and children in the household). Columns 1 to 3 show the results from an OLS specification. Columns 4 to 6 show the results of a 2SLS regression, relying on the average covariates of the social network as instrument (covariates cultural type, age, female, previous military service, illnesses, previous migration and children in the household). All regressions are reported with robust standard errors. The bottom panel shows the Sanderson-Windmeijer (SW) first-stage chi-squared and F statistics, which are tests of underidentification and weak identification of endogenous regressors.

Table 3.5: Strategic cultural emigration and peer effects: excluding farmers.

	Time of application for emigration: excluding farmers			2SLS specifications		
	OLS specifications					
	(1)	(2)	(3)	(4)	(5)	(6)
Cultural type	0.045*** [0.013]	0.030*** [0.009]	0.026*** [0.009]	0.011 [0.013]	0.019** [0.009]	0.019** [0.009]
Social network decision	0.098*** [0.017]	0.014 [0.010]	0.012 [0.010]	0.444*** [0.050]	0.157*** [0.034]	0.104*** [0.031]
Personal property	-0.028*** [0.007]	-0.005 [0.005]	-0.005 [0.005]	-0.012 [0.008]	-0.002 [0.005]	-0.003 [0.005]
Neighborhood property	-0.005 [0.007]	-0.008* [0.004]	-0.008* [0.004]	-0.002 [0.008]	-0.006 [0.004]	-0.007 [0.004]
<i>Mean value of "application"</i>	0.868	0.870	0.870	0.868	0.870	0.870
Officer fixed effects	NO	YES	YES	NO	YES	YES
Control variables	NO	NO	YES	NO	NO	YES
Observations	1,886	1,770	1,742	1,886	1,770	1,742
F-statistic				125.78	99.34	101.6
F-statistic (p-value)				0.000	0.000	0.000
SW Chi-sq				126.12	103.01	105.73
SW Chi-sq (p-value)				0.000	0.000	0.000

Standard errors in brackets

*** p < 0.01, ** p < 0.05, * p < 0.1

The regression table shows the results from a regression of a set of explanatory variables on the outcome variable "time of application for emigration", which is a normalized variable between 0 and 1 of the application date, 1 being the earliest application date, 0 the latest. "Cultural type" reflects the cultural type of an individual. It is composed of a "Germanic first name" index and migration history to Germany or Austria, to account for "inherited" and "acquired" cultural preference. "Social network decision" reflects the emigration decision of the social networks, which include neighbors and individuals from the same social class (capitalist, worker, no income) in the same village. "Personal property" is a binary variable equal to one if the person or their family owns property, and zero otherwise. "Neighborhood property" is a binary variable indicating whether the individual's neighbors own property. Columns 1 and 4 show results from specifications without the inclusion of a set of control variables, columns 2 and 5 show the results with inclusion of officer fixed effects, and columns 3 and 6 include officer fixed effects and a set of the control variables (age, female, previous military service, illnesses, previous migration and children in the household). Columns 1 to 3 show the results from an OLS specification. Columns 4 to 6 show the results of a 2SLS regression, relying on the average covariates of the social network as instrument (covariates cultural type, age, female, previous military service, illnesses, previous migration and children in the household). All regressions are reported with robust standard errors. The bottom panel shows the Sanderson-Windmeijer (SW) first-stage chi-squared and F statistics, which are tests of underidentification and weak identification of endogenous regressors.

Table 3.6: Strategic cultural emigration and peer effects: binary emigration variable.

	Emigration status					
	OLS specifications			2SLS specifications		
	(1)	(2)	(3)	(4)	(5)	(6)
Cultural type	0.202*** [0.032]	0.087*** [0.030]	0.089*** [0.030]	0.099*** [0.037]	0.042 [0.032]	0.044 [0.032]
Social network decision	0.545*** [0.015]	0.316*** [0.021]	0.306*** [0.021]	1.650*** [0.099]	0.990*** [0.094]	0.942*** [0.087]
Personal property	-0.196*** [0.020]	-0.075*** [0.019]	-0.076*** [0.019]	-0.110*** [0.024]	-0.046** [0.021]	-0.050** [0.021]
Neighborhood property	-0.083*** [0.019]	-0.057*** [0.017]	-0.059*** [0.017]	-0.062*** [0.023]	-0.045** [0.019]	-0.048*** [0.018]
<i>Mean value of "emigration"</i>	0.571	0.565	0.568	0.571	0.565	0.568
Officer fixed effects	NO	YES	YES	NO	YES	YES
Control variables	NO	NO	YES	NO	NO	YES
Observations	2,338	2,202	2,171	2,327	2,197	2,167
F-statistic				217.57	163.96	179.34
F-statistic (p-value)				0.000	0.000	0.000
SW Chi-sq				218.04	168.83	185.41
SW Chi-sq (p-value)				0.000	0.000	0.000

Standard errors in brackets

*** p<0.01, ** p<0.05, * p<0.1

The regression table shows the results from a regression of a set of explanatory variables on the outcome variable "application status", a binary variable equal to 1 if the agent effectively emigrated, 0 otherwise. "Cultural type" reflects the cultural type of an individual. It is composed of a "Germanic first name" index and migration history to Germany or Austria, to account for "inherited" and "acquired" cultural preference. "Social network decision" reflects the emigration decision of the social networks, which include neighbors and individuals from the same social class (capitalist, worker, no income) in the same village. "Personal property" is a binary variable equal to one if the person or their family owns property, and zero otherwise. "Neighborhood property" is a binary variable indicating whether the individual's neighbors own property. Columns 1 and 4 show results from specifications without the inclusion of a set of control variables, columns 2 and 5 show the results with inclusion of officer fixed effects, and columns 3 and 6 include officer fixed effects and a set of the control variables (age, female, previous military service, illnesses, previous migration and children in the household). Columns 1 to 3 show the results from an OLS specification. Columns 4 to 6 show the results of a 2SLS regression, relying on the average covariates of the social network as instrument (covariates cultural type, age, female, previous military service, illnesses, previous migration and children in the household). All regressions are reported with robust standard errors. The bottom panel shows the Sanderson-Windmeijer (SW) first-stage chi-squared and F statistics, which are tests of underidentification and weak identification of endogenous regressors.

Table 3.7: Oster (2019) test.

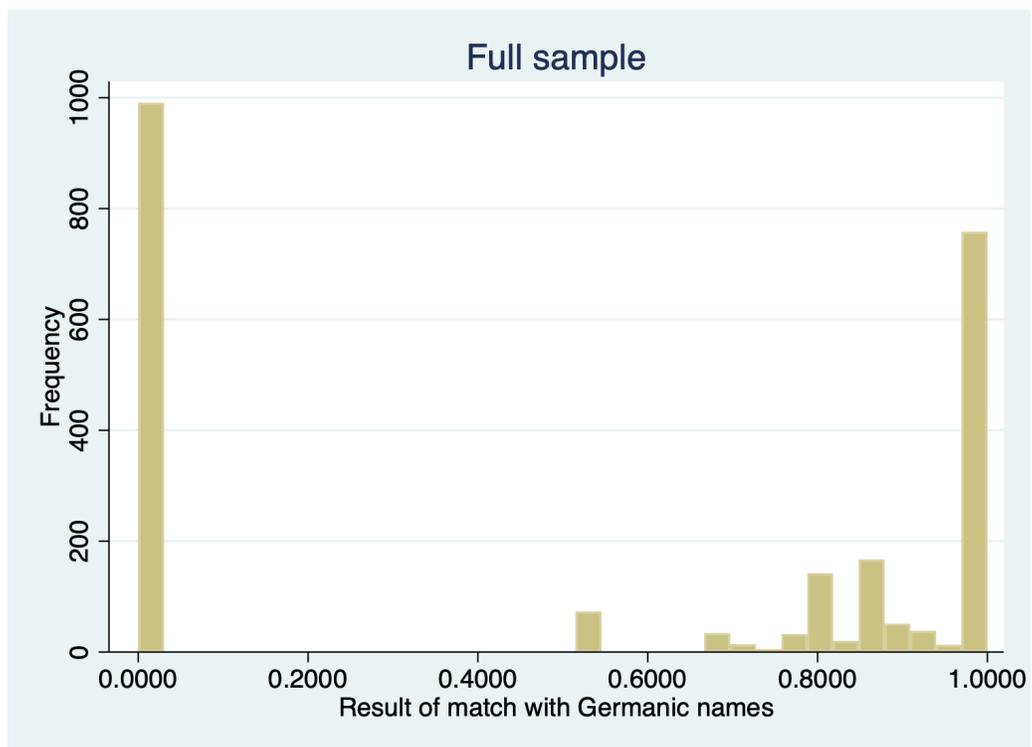
	Time of application for emigration		
	OLS specifications		
	(1)	(2)	(2)
Cultural type	0.040***	0.024***	0.021***
	[0.012]	[0.008]	[0.008]
delta	12.327	7.039	5.323
Social network decision	0.072***	0.017***	0.017***
	[0.011]	[0.006]	[0.006]
delta	6.068	2.215	2.135
Personal property	-0.039***	-0.016***	-0.015***
	[0.006]	[0.004]	[0.004]
delta	4.425	2.642	2.209
Neighborhood property	-0.013**	-0.012***	-0.010***
	[0.006]	[0.004]	[0.004]
delta	5.272	4.708	3.544
<i>Mean value of "application"</i>	0.858	0.859	0.856
Officer fixed effects	NO	YES	YES
Control variables	NO	NO	YES
Observations	2,327	2,197	2,167

Standard errors in brackets

*** p<0.01, ** p<0.05, * p<0.1

The regression table shows the results from a regression of a set of explanatory variables on the outcome variable "time of application for emigration", which is a normalized variable between 0 and 1 of the application date, 1 being the earliest application date, 0 the latest. "Cultural type" reflects the cultural type of an individual. It is composed of a "Germanic first name" index and migration history to Germany or Austria, to account for "inherited" and "acquired" cultural preference. "Social network decision" reflects the emigration decision of the social networks, which include neighbors and individuals from the same social class (capitalist, worker, no income) in the same village. "Personal property" is a binary variable equal to one if the person or their family owns property, and zero otherwise. "Neighborhood property" is a binary variable indicating whether the individual's neighbors own property. All regressions are reported with robust standard errors. *Delta* is calculated using the *psacalc* command by Oster (2013), assuming $R_{max} = 1.1 \cdot R^2$ and $\beta = 0$.

Figure 3.7: Histogram of the match variable: first names and Germanic names



3.6. Appendix

3.6.1. Proof of Lemma 3.1

From standard linear algebra, we know that a system of linear equation has a unique solution if and only if the determinant is not zero. Furthermore, the determinant of a matrix is not zero if and only if the matrix is invertible. The best-response functions (3.2) of households form a system of linear equations which could be written in matrix notation as:

$$\begin{aligned}
 & ((1 - \alpha)I + \alpha D_S)Y = (1 - \alpha)D_\tau \mathbf{1} + \alpha SY \\
 & \Leftrightarrow Y = (1 - \alpha)((1 - \alpha)I + \alpha D_S)^{-1}D_\tau \mathbf{1} + \alpha((1 - \alpha)I + \alpha D_S)^{-1}SY \\
 & \Leftrightarrow (I - \alpha((1 - \alpha)I + \alpha D_S)^{-1}S)Y = (1 - \alpha)((1 - \alpha)I + \alpha D_S)^{-1}D_\tau \mathbf{1} \\
 & \Leftrightarrow (I - \alpha\beta S)Y = (1 - \alpha)\beta D_\tau \mathbf{1},
 \end{aligned}$$

where $\beta = ((1 - \alpha)I + \alpha D_S)^{-1}$. If $\alpha\rho(\beta S) < 1$ with $\rho(\beta S)$ being the spectral radius of the matrix βS , then the matrix $(I - \alpha\beta S)$ is invertible and the above system has a unique solution. The condition $\alpha\rho(\beta S) < 1$ trivially holds for any $\alpha \in (0, 1)$, by the Gershgorin circle theorem. The diagonal terms of the matrix $\alpha\beta S$ are zero and for any given row i , the sum of the off-diagonal terms is exactly $\frac{\alpha d_i(S)}{1 - \alpha + \alpha d_i(S)}$ which is strictly smaller than 1 for any $\alpha \in (0, 1)$. Finally, notice that the rows of the matrix $(1 - \alpha)(I - \alpha\beta S)^{-1}\beta$ sum to 1. Hence, for all $i \in N$, if $\tau_i \equiv \tau$ then $y_i^* = \tau$.

3.6.2. Proof of proposition 3.1

The action space for each household is the interval $Y = [0, 1]$, hence it is convex and compact. Furthermore, the payoff function π_i is concave in each household's strategy y_i . Hence by the Kakutani fixed point theorem there exists a Nash equilibrium. We use the payoff function (3.5) to get the best-response of each player to the actions of other players. We obtain a system of linear equations. We provide a sufficient condition on the parameters of the model

so that the system has a unique solution. Formally,

$$\begin{aligned}
\frac{\partial \pi_i(y_i, \mathbf{y}_{-i}; S, G)}{\partial y_i} = 0 &\Leftrightarrow \frac{\tau_i}{1 + v_i} - y_i - \frac{\delta}{d_i(G)} \sum_{j \neq i \in N} g_{ij} v_j y_j - 2(1 - \alpha)(y_i - \tau_i) - 2\alpha \sum_{j \neq i \in N} s_{ij}(y_i - y_j) \\
&= 0 \\
&\Leftrightarrow \frac{\tau_i(1 + 2(1 - \alpha)(1 + v_i))}{(1 + v_i)} + \sum_{j \neq i \in N} (2\alpha s_{ij} - \frac{\delta g_{ij} v_j}{d_i(G)}) y_j \\
&= (1 + 2(1 - \alpha) + 2\alpha \sum_{j \neq i \in N} s_{ij}) y_i.
\end{aligned}$$

Let $D_v = \text{diag}(v_1, \dots, v_n)$, $D_\tau = \text{diag}(\tau_1, \dots, \tau_n)$, $D_S = \text{diag}(\sum_{j \in N} s_{1j}, \dots, \sum_{j \in N} s_{nj})$, $Y = (y_1, \dots, y_n)^T$, $D_G = \text{diag}(\sum_{j \in N} g_{1j}, \dots, \sum_{j \in N} g_{nj})$, $\mathbf{1}_n$ a column vector with n rows of ones. The above system, in matrix notation writes:

$$\begin{aligned}
X\mathbf{1} + MY &= NY \\
\Leftrightarrow N^{-1}X\mathbf{1}_n + N^{-1}MY &= Y \\
\Leftrightarrow N^{-1}X\mathbf{1}_n &= (I - N^{-1}M)Y,
\end{aligned}$$

where $X = (I + D_v)^{-1}D_\tau(I + 2(1 - \alpha)(I + D_v))$, $M = 2\alpha S - \delta D_G^{-1}GD_v$, $N = I + 2(1 - \alpha)I + 2\alpha D_S$. A solution exists and is unique if the matrix $(I - N^{-1}M)$ is invertible. We prove that the matrix $(I - N^{-1}M)$ is indeed invertible using the following two theorems.

Theorem 3.1 (5.6.12, p.348, **Horn and Johnson (1985)**) *Let $A \in M_n$ be given. Then $\lim_{k \rightarrow \infty} A^k = 0$ if and only if $\rho(A) < 1$.*

Theorem 3.2 (Gershgorin circle theorem) *For a square $n \times n$ matrix A , let $C_i = \{c \in \mathbb{C} : |c - a_{ii}| \leq r_i\}$ be the Gershgorin disc for row i where $r_i = \sum_{j \neq i, j=1}^n |a_{ij}|$. Every eigenvalue $\lambda \in \mathbb{C}$ of the square matrix $A \in \mathbb{R}^{n \times n}$ lies in at least one of the Gershgorin discs C_i . The possible range of the eigenvalues is defined by the outer borders of the union of all discs: $C = \bigcup_{i=1}^n C_i$.*

Recall that the matrix $N^{-1}M = (I + 2(1 - \alpha)I + 2\alpha D_S)^{-1}(2\alpha S - \delta D_G^{-1}GD_v)$. The diagonal terms are zero, hence all the Gershgorin discs are centered around zero. Let $\rho(N^{-1}M)$ be the largest absolute value of the eigenvalues (spectral radius) of $N^{-1}M$. The radius of each Gershgorin disc associated with row i is given by the sum over columns of the off-diagonal

terms. Hence, the radius of the largest Gershgorin disc is :

$$\bar{r} = \max_{i \in N} \left\{ \frac{1}{1 + 2(1 - \alpha) + 2\alpha d_i(S)} \sum_{j \neq i} \left| 2\alpha s_{ij} - \frac{\delta g_{ij} v_j}{d_i(G)} \right| \right\}. \quad (3.9)$$

The matrix $N^{-1}M$ is convergent if and only if $\rho(N^{-1}M) < 1$. Moreover by the Gershgorin disc theorem, $\rho(N^{-1}M) \leq \bar{r}$. We will now show that $\bar{r} < 1$ to be able to conclude.

Recall that for all $i \neq j \in N$, $v_i \in [0, 1]$, $\alpha \in (0, 1)$, $s_{ij} \in \{0, 1\}$, $g_{ij} \in \{0, 1\}$, $d_i(G) \geq 1$ and $d_i(S) \geq 1$. The inequality $\rho(N^{-1}M) \leq \bar{r}$ must hold for all the possible values of the parameters and all network structures S and G .

If $v_j = 0$ for all $j \in N$, then for all $\alpha \in (0, 1)$ and network structures S and G :

$$\bar{r} \leq \max_{i \in N} \left\{ \frac{2\alpha d_i(S)}{1 + 2(1 - \alpha) + 2\alpha d_i(S)} \right\} < 1.$$

The denominator takes the highest value for very small α :

$$\bar{r} \xrightarrow{\alpha \rightarrow 0} \frac{1}{3} \max_{i \in N} \left\{ \sum_{j \neq i} \frac{\delta g_{ij} v_j}{d_i(G)} \right\} \leq \frac{1}{3} \max_{i \in N} \left\{ \sum_{j \neq i} \frac{g_{ij}}{d_i(G)} \right\} < 1$$

3.6.3. Beyond the Italian Border

From the currently digitized files, we cannot follow the opters beyond the Italian border. This means, we do not know where they moved. The corresponding files on the other side of the border are stored in the *Tiroler Landesarchiv* (the Archives of Tyrol) in Innsbruck (Austria). There, opters can be traced through their unique identifier number. However, this data is currently not digitized. Nevertheless, Historical evidence¹⁵ provides us with a macro-level overview of the destination of the opters, presented in Figure 3.8.

Furthermore, digitizing “Tiroler Landesarchiv” in Innsbruck can inform us about whether *peers* moved to the same destination. This information would extend the literature that studies the role of migration networks in the choice of the destination country. For example, Simone Bertoli and Ilse Ruyssen (2018) [18] rely on individual-level data to trace an individual’s direct connections to the migrant networks in different countries and their choice concerning the preferred country of destination. The authors find distance-one connections to be an important driver in the choice of destinations with a similar level of attractiveness.

Finally, to study the question of whether peers remained in the same region we can also

¹⁵Table 9 on page 99 of the book “Heimatlos. Die Umsiedlung der Südtiroler”, Wien, 1993, by Helmut Alexander, Stefan Lechner and Adolf Leidlmair

Region (in German)	Country	Total opters	Percentage
Nordtirol	Austria	38,500	51.5
Vorarlberg	Austria	5,700	7.6
Salzburg	Austria	3,900	5.2
Oberösterreich	Austria	4,500	6.3
Steiermark, Kärnten mit Osttirol	Austria	5,600	7.5
Wien, Niederösterreich, Burgenland	Austria	1,800	2.4
Deutschland in Grenzen von 1937	Germany	10,800	14.5
Andere Gebiete	Other	3,700	5.0
<i>Total</i>		<i>74,500</i>	<i>100.0</i>

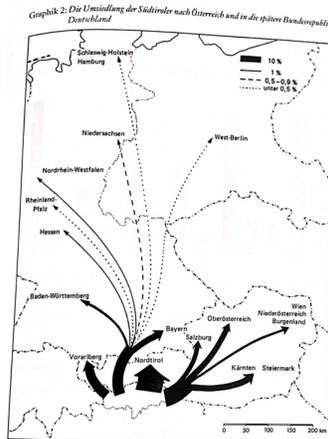


Figure 3.8

look for information about households who decided to migrate but came back after few years. The opters and their children were allowed to migrate back to Italy officially only starting from 1948. They had to renounce to German citizenship and re-request Italian citizenship. According to historical evidence (see Alexander et al., (1993) [6]), about half of all initial opters returned to South Tyrol. The return opters can be traced through the same initially created unique identifiers. However, since the digitization that we carried out is the first attempt to digitize any South Tyrol Option files, the “return option” files are not yet digitized either. Historians Eva Pfanzelter from the University of Innsbruck, and Andrea di Michele from the Free University of Bolzano are starting a project related to the return option, involving an extensive digitization project of the corresponding files. We hope that this data can help us understand the migratory trajectory of the opters who, in the end, returned.

Bibliography

- [1] Daron Acemoglu, Giacomo Como, Fabio Fagnani, and Asuman Ozdaglar. Opinion fluctuations and disagreement in social networks. *Mathematics of Operations Research*, 38(1), 2013.
- [2] Daron Acemoglu, Munther Dahleh, Ilan Lobel, and Asuman Ozdaglar. Bayesian learning in social networks. *Review of Economic Studies*, 78:1201–1236, 2011.
- [3] Daron Acemoglu and Asuman Ozdaglar. Opinion dynamics and learning in social networks. *Dyn Games Appl*, 1:3–49, 2011.
- [4] G. A. Akerlof and R. E. Kranton. *Identity Economics: How Our Identities Shape Our Work, Wages, and Well-Being*. Princeton University Press, 2010.
- [5] A.E. Alcock. *The History of the South Tyrol Question*. London, Michael Joseph, 1970.
- [6] H. Alexander, S. Lechner, and A. Leidlmaier. *Heimatlos. Die Umsiedlung der Südtiroler*. Wien, Deuticke Verlag, 1993.
- [7] C. Armmitage and M. Conner. Efficacy of the theory of planned behaviour: A meta-analytic review. *British journal of social psychology*, 40(4), 2001.
- [8] Baris Ata, Alexandre Belloni, and Ozan Candogan. Latent agents in networks: Estimation and pricing. *arXiv preprint arXiv:1808.04878*, 2018.
- [9] Venkatesh Bala and Sanjeev Goyal. Learning from neighbours. *Review of Economic Studies*, 65:595–621, 1998.
- [10] Coralio Ballester, Antoni Calvó-Armengol, and Yves Zenou. Who’s who in networks. wanted: The key player. *Econometrica*, 74(5):1403–1417, 2006.
- [11] Abhijit Banerjee, Emily Breza, Arun g. Chandrasekhar, and Markus Mobius. Naive learning with uninformed agents. *mimeo*, 2018.
- [12] S. Banisch and E. Olbrich. Opinion polarization by learning from social feedback. *The Journal of Mathematical Sociology*, 43(2):76–103, 2019.

- [13] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286(5439), 1999.
- [14] T. Bauer, G.S. Epstein, and I. N. Gang. Enclaves, language and the location choice of immigrants. *Journal of population economics*, 18(4):649–662, 2005.
- [15] D. J. Benjamin, J.J. Choi, and G. Fisher. Religious identity and economic behavior. *Review of economics and statistics*, 98(4):617–637, 2016.
- [16] D.J. Benjamin, J.J. Choi, and A.J. Strickland. Social identity and preferences. *American Economic Review*, 100(4):1913–1928, 2010.
- [17] Edward L. Bernays. *Propaganda*. 1928.
- [18] S. Bertoli and I. Ruysen. Networks and migrants? intended destination. *Journal of Economic Geography*, 18(4):705–728, 2018.
- [19] C. Biavaschi, C. Giulietti, and Z. Siddique. The economic payoff of name americanization. *Journal of Labor Economics*, 35(4), 2017.
- [20] A. Bisin, A. Patacchini, T. Verdier, and Y. Zenou. Bend it like bechham: Ethnic identity and integration. *European Economic Review*, 90, 2016.
- [21] Alberto Bisin, Eleonora Patacchini, Thierry Verdier, and Yves Zenou. Bend it like bechham: Ethnic identity and integration. *European Economic Review*, 90:146–164, 2016.
- [22] Francis Bloch et al. Targeting and pricing in social networks. *The Oxford Handbook of the Economics of Networks*, pages 776–791, 2016.
- [23] Francis Bloch and Nicolas Quérou. Pricing in social networks. *Games and economic behavior*, 80:243–261, 2013.
- [24] J. Blumenstock, G. Chi, and X. Tan. Migration and the value of social networks. *mimeo*, 2019.
- [25] Ugo Bolletta and Paolo Pin. Polarization when people choose their peers. *mimeo*, 2019.
- [26] Gustave Le Bon. *The Crowd: A Study of the Popular Mind (La Psychologie des foules)*. 1895.
- [27] Phillip Bonacich. Factoring and weighting approaches to status scores and clique identification. *Journal of mathematical sociology*, 2(1):113–120, 1972.
- [28] Phillip Bonacich and Paulette Lloyd. Calculating status with negative relations. *Social Networks*, 26:331–338, 2004.

- [29] Scott A. Boorman and Harrison C. White. Social structure from multiple networks. ii
tole structures. *The American Journal of Sociology*, 81(6):1384–1446, 1976.
- [30] G. J. Borjas, I. Kauppinen, and P. Poutvaara. Self-selection of emigrants: Theory and
evidence on stochastic dominance in observable and unobservable characteristics. *The
economic journal*, 129(617):143–171, 2019.
- [31] G.J. Borjas. Self-selection and the earning of immigrants. *American Economic Review*,
77(4):531–553, 1987.
- [32] Y. Bramoullé and R. Kranton. *Games Played on Networks*, chapter 5. The oxford hand-
book of the Economics of Networks, 2016.
- [33] Ozan Candogan, Kostas Bimpikis, and Asuman Ozdaglar. Optimal pricing in net-
works with externalities. *Operations Research*, 60(4):883–905, 2012.
- [34] Dominique Cardon. *Culture Numérique*. Presses de Sciences Po, 2019.
- [35] Louise Caron. An intergenerational perspective on (re)migration: Return and on-
ward mobility intentions across immigrant generations. *International Migration Review*,
54(3), 2020.
- [36] Dorwin Cartwright and Frank Harary. Structural balance: a generalization of heider’s
theory. *Psychological Review*, 63(5):277–293, 1956.
- [37] D. Chiquiar and H. H. Gordon. International migration, self- selection, and the dis-
tribution of wages: Evidence from mexico and the united states. *Journal of Political
Economy*, 113(2):239–281, 2005.
- [38] B. Chiswick. Are immigrants favorably self-selected? *American Economic Review: Pa-
pers and Proceedings*, 89(2):181–185, 1999.
- [39] M.D. Conover, J. Ratkiewicz, M. Francisco, B.Gonçalves, A. Flammini, and F. Menczer.
Political polarization on twitter. *Proceedings of the Fifth International AAAI conference on
weblogs and social media*, 2011.
- [40] Jacques Cremer and Richard P McLean. Full extraction of the surplus in bayesian
and dominant strategy auctions. *Econometrica: Journal of the Econometric Society*, pages
1247–1257, 1988.
- [41] H.O. Van Dalen and K. Henkens. Longing for the good life: understanding emigration
from a high-income country. *Population and Development Review*, 33(1), 2007.

- [42] Morris H. DeGroot. Reaching a consensus. *Journal of the American Statistical Association*, 69(345):118–121, 1974.
- [43] Gabrielle Demange. Optimal targeting strategies in a network under complementarities. *Games and Economic Behavior*, 105:84–103, 2017.
- [44] K. Eisterer and R. Steininger. *Die Option. Südtirol zwischen Faschismus und Nationalsozialismus*. Innsbruck, Haymon-Verlag, 1989.
- [45] Itay P Fainmesser and Andrea Galeotti. Pricing network effects. *The Review of Economic Studies*, 83(1):165–198, 2015.
- [46] Andreas Flache, Michael Mas, Thomas Feliciani, Edmund Chattoe-Brown, Guillaume Deffuant, Sylvie Huet, and Jan Lorenz. Models of social influence: Towards the next frontiers. *Journal of Artificial Societies and Social Simulation*, 2017.
- [47] V. Fouka. Backlash: The unintended effects of language prohibition in us schools after world war i. *The Review of Economic Studies*, 2020.
- [48] Noah Friedkin and Eugene Johnsen. Social influence and opinions. *Journal of mathematical sociology*, 15(3-4):193–205, 1990.
- [49] Noah E. Friedkin. The problem of social control and coordination of complex systems in sociology: a look at the community cleavage problem. *IEEE control systems magazine*, June, 2015.
- [50] Andrea Galeotti, Benjamin Golub, and Sanjeev Goyal. Targeting interventions in networks. *arXiv preprint arXiv:1710.06026*, 2017.
- [51] C. Gatterer. *Schöne Welt, böse Leut: Kindheit in Südtirol*. Folio Verlag, 2015.
- [52] Alejandro Gaviria and Steven Raphael. School-based peer effects and juvenile behavior. *The review of Economics and Statistics*, 83(2):257–268, 2001.
- [53] Benjamin Golub and Matthew Jackson. How homophily affects the speed of learning and best-response dynamics. *The Quarterly Journal of Economics*, pages 1287–1338, 2012.
- [54] Benjamin Golub and Evan Sadler. *Learning in Social Networks*, chapter 19. The Oxford Handbook of the Economics of Networks, 2016.
- [55] Michel Grabisch, Antoine Mandel, Agnieszka Rusinowska, and Emily Tanimura. Strategic influence in social networks. *Mathematics of Operations Research*, 43, 2018.

- [56] B. Gubhaju and F. De Jong. Individual versus household migration decision rules: Gender and marital status differences in intentions to migrate in south africa. *International Migration*, 47(1), 2009.
- [57] F. Harary. A criterion for unanimity in french's theory of social power. *Studies in Social Power*, 1959.
- [58] Rainer Hegselmann and Ulrich Krause. Opinion dynamics and bounded confidence models, analysis, and simulationb. *Journal of Artificial Societies and Social Simulation*, 5(3), 2002.
- [59] K. Hoff and P. Pandey. Discrimination, social identity, and durable inequalities. *American Economic Review*, 96(2):206–211, 2006.
- [60] International Foundation for Autonomous Agents and Multiagent Systems. *Uncharted but not Uninfluenced: Influence Maximization with an uncertain network*, 2017.
- [61] M. O. Jackson. *Social and Economic Networks*. Princeton University Press, 2008.
- [62] Matthew O Jackson and Asher Wolinsky. A strategic model of social and economic networks. *Journal of Economic Theory*, 71(1):44–74, 1996.
- [63] Ali Jadbabaie and Ali Kakhbod. Optimal contracting in networks. *Journal of Economic Theory*, 183:1094–1153, 2019.
- [64] Wander Jager and Frédéric Amblard. Uniformity, bipolarization and pluriformity captured as generic stylized behavior with an agent-based simulation model of attitude change. *Computational and Mathematical Organization Theory*, 10:295–303, 2004.
- [65] J.Ding and Noah H. Rhee. On the equality of algebraic and geometric multiplicities of matrix eigenvalues. *Applied Mathematics letters*, 24, 2011.
- [66] James R. Larson Jr., Caryn Christensen, Ann S. Abbott, and Timothy M. Franz. Diagnosing groups: Charting the flow of information in medical decision-making teams. *Journal of Personality and social psychology*, 71(2):315–330, 1996.
- [67] J.R. P. French Jr. A formal theory of social power. *Psychol. Rev*, 63, 1956.
- [68] R.G. Fryer Jr and S. D. Levitt. The causes and consequences of distinctively black names. *The Quarterly Journal of Economics*, 119(3):767–805, 2004.
- [69] Elihu Katz and Paul Lazarsfeld. *Personal Influence: The Part Played by People in the Flow of Mass Communications*. 1955.

- [70] Leo Katz. A new status index derived from sociometric analysis. *Psychometrika*, 18(1):39–43, 1953.
- [71] Rachel Kranton and Deborah Minehart. A theory of buyer-seller networks. *American Economic Review*, 91(3):485–508, 2001.
- [72] Bibb Latané. The psychology of social impact. *American Psychologist*, 36(4):343–356, 1981.
- [73] Bibb Latané, Andrej Nowak, and James H. Liu. Measuring emergent social phenomena: dynamism, polarization, and clustering as order parameters of social systems. *Behavioral Science*, 39, 1994.
- [74] Fragkiskis Malliaros and Michalis Vazirgiannis. Clustering and community detection in directed networks: A survey. *Physics Reports*, 2013.
- [75] Ana Mauleon and Vincent Vannetelbosch. *Network formation games*, chapter 8. The oxford handbook of the Economics of Networks, 2016.
- [76] D. McKenzie and H. Rapoport. Self-selection patterns in mexico-u.s. migration: The role of migration networks. *Review of economics and statistics*, 92(4):811–821, 2010.
- [77] Isabel Melguizo. Persistence of disagreement. *miemo*, 2018.
- [78] Carl D. Meyer. *Matrix Analysis and Applied linear algebra*. SIAM, 2000.
- [79] F. H. Moraga. New evidence on emigrant selection. *Review of economics and statistics*, 93(1):72–96, 2011.
- [80] K. Munshi. Social networks and migration. *Annual Review of Economics*, 12:503–524, 2020.
- [81] K. Munshi and N. Wilson. Identity, occupational choice, and mobility: Historical conditions and current decisions in the american midwest. *mimeo*, 2011.
- [82] M.E.J. Newman. Assortative mixing in networks. *Phys. Rev. Lett*, 89, 2002.
- [83] E. Oster. Psocal: stata module to calculate treatment effects and relative degree of selection under proportional selection of observables and unobservables. *Statistical Software Components S457677*, Boston College department of economics(revised 18), 2013.
- [84] Emily Oster. Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business and Economic Statistics*, 37(2):187–204, 2019.

- [85] H. Rapoport, S. Sardoschau, and A. Silve. International migration and cultural convergence. *mimeo*, 2017.
- [86] Y. Rubinstein and D. Brenner. Pride and prejudice: using ethnic-sounding names and inter-ethnic marriages to identify labour market discrimination. *Review of Economic Studies*, 81(1):389–425, 2013.
- [87] Evan Sadler. Influence campaigns. *mimeo*, 2019.
- [88] E. Sanderson and F. Windmeijer. A weak instrument f-test in linear iv models ith multiple endogenous variables. *Journal of Econometrics*, 180:212–221, 2016.
- [89] E. Seneta. *Non-negative matrices and markov chains*. Springer Science and Business Media, 2006 edition, 1981.
- [90] Fanqi Shi and Yiqing Xing. Screening with network externalities. *Available at SSRN 3286427*, 2018.
- [91] Teresa M. Silva, Luís Silva, and Sara Fernandes. Convergence time to equilibrium distributions of autonomous and periodic non-autonomous graphs. *Linear Algebra and its Applications*, 488:199–215, 2016.
- [92] Daniel Smith, Philip Schlaepfer, Katie Major, Mark Dyble, Abigail E. Page, James Thompson, Nikhil Chaudhary, Gul Deniz Salali, Ruth Mace, Leonora Astete, Marilyn Ngales, Lucio Vinicius, and Andrea Bamberg Migliano. Cooperation and the evolution of hunter-gatherer storytelling. *Nature Communications*, 8(1853), 2017.
- [93] G. Stasser, L. Taylor, and C. Hanna. Information sampling in structured and unstructured discussions of three and six person groups. *Journal of Personality and social psychology*, 57:67–78, 1989.
- [94] Garold Stasser and William Titus. Pooling of unshared information in group decision making: Biased information sampling during discussion. *Journal of Personality and social psychology*, 48(6):1467–1478, 1985.
- [95] Garold Stasser and William Titus. Hidden profiles: a brief history. *Psychological Inquiry*, 14(3):304–313, 2003.
- [96] Fernando Vega-Redondo. *Links and Actions in Interplay*, chapter 9. The Oxford Handbook of the Economics of Networks, 2016.
- [97] S. Wasserman and K. Faust. *Social Network Analysis : methods and applications*. Cambridge university press, 1994.

-
- [98] Robert Wilson et al. Information, efficiency, and the core of an economy. *Econometrica*, 46(4):807–816, 1978.
- [99] Lin Xiao and Stephen Boyd. Fast linear iterations for distributed averaging. *Systems and Control Letters*, 53, 2004.
- [100] E. Yildiz, A. Ozdaglar, D. Acemoglu, A. Saberi, and A. Scaglione. Binary opinion dynamics with stubborn agents. *ACM Trans. Econ. Comp.*, 2013.
- [101] Yves Zenou. *Key Players*, chapter 11. The oxford handbook of the Economics of Networks, 2016.
- [102] Shoshana Zuboff. *The Age of Surveillance Capitalism*. Profile Books Ltd, 2019.