AN ALGORITHMIC APPROACH TO ANALYZING SOCIAL PHENOMENA

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Sigal Oren

August 2013

AN ALGORITHMIC APPROACH TO ANALYZING SOCIAL PHENOMENA

Sigal Oren, Ph.D.

Cornell University 2013

Online information and interaction is becoming more and more prominent in our lives. This development is made possible by the growth of large-scale user-based applications on the Web (including sites such as Wikipedia and Facebook). As the number of people using these applications increases, the number of social interactions increases and we start to witness social phenomena which originally appeared in the offline world, as well as new ones.

Our main goal in this thesis is to obtain a better understanding of some of these phenomena both in the online and the offline world. We will concentrate on phenomena from two main domains. First, motivated by the increasing interest in polarization and the implications that social interactions in the online world has on it, we study how people form their opinion. We present and analyze two models of opinion formation and a more general model of culture dynamics describing the process by which people form opinions on a set of issues simultaneously.

Second, we consider how to allocate credit to incentivize effort. We explore this question in the realm of scientific communities by studying a simple game theoretic model illustrating the process by which researchers choose a research project. Our results are not restricted to the academic domain alone, as crowd sourcing sites like Wikipedia are already implementing number of credit-allocation conventions familiar from the scientific community. We also take a special interest in studying the effects long range reasoning has on individuals' choices in other academic domains.

We will take the algorithmic approach in which we first try to construct a model of the phenomena in question. For the most part of this thesis we choose to model individuals as strategic agents maximizing some utility function. Then we analyze the model using tools from various fields such as game theory, computer science and statistical physics. Finally, we use our analysis to derive lessons for designing new systems.

BIOGRAPHICAL SKETCH

Sigal grew up in Haifa, Israel, where she got a BSc in software engineering from the Technion in 2008. After spending some time working as a software engineer, she joined in 2009 the PhD program of Cornell's computer science department. She expects to graduate in the Summer of 2013.

To my parents, Ethy and Aharon.

ACKNOWLEDGEMENTS

I cannot find words to express my gratitude to Jon Kleinberg for being the best advisor I could ever wished for. Jon was a pleasure to work with in every aspect: incredible mentor, teacher and collaborator. The most exciting part of my PhD was undoubtedly the meetings with Jon. During our meetings not only have I learned about new problems and new technical tools, but more importantly I learned how to think. I would miss them dearly.

As an undergrad at the Technion I was working as a part-time software engineer and was sure that this is my future career path. This all changed when Ron Lavi approached me and suggested I will do research with him. With Ron's guidance I discovered how exciting, creative and joyful research is. I will ever be indebted to Ron.

During my PhD I had the pleasure of attending Michael Macy's social dynamics lab group meetings. In these meeting I got a chance to learn how sociologists think on problems similar to my own research problems. I deeply thank Michael and all the lab members for sharing with me their work and giving me feedback on my own work. I would also like to thank Michael for his valuable input as one of my committee members.

I am grateful to the faculty at Cornell CS department. In particular, I thank David Bindel for being a great collaborator and willing to patiently explain linear algebra concepts. Johannes Gehrke was an excellent resource for doing research in databases. I also thank Joe Halpern for serving on my committee and for interesting conversations. Bobby Kleinberg's help, starting from the time I applied to Cornell, was invaluable. I am grateful to him for this and also for wonderful free dinners and a rescue from a snow storm.

I was lucky to spend an awesome summer at MSR Silicon Valley where I was mentored by Alex Slivkins and Moshe Babaioff. I had many fascinating conversations with Alex both on research and on life in general. This was also where I started working with David Kempe. Among many other things, I hope I learned from David how to write clearer and better proofs. I had great fun working with Moshe Babaioff, Shahar Dobzinski and Aviv Zohar on Bitcoin and learned on the importance of hummus in facilitating the research process. My only regret is not buying Bitcoins back then while they were cheap.

I was fortunate to spend 3 months in Microsoft Research Herzelia under the mentorship of Moshe Tennenholtz. I am thankful for Moshe for sharing with me his vast knowledge and original opinions on research and professional issues alike. I also greatly benefited from the time I spent working with Michael Schapira, discussing with the same intensity how to fix internet protocols and which ice cream flavor we are going to eat today.

I would also like to pay special thanks to all my collaborators and friends from Cornell whom I did not mention earlier: Ashwinkumar Badanidiyuru, Flavio Chierichetti, Lucja Kot, Konstantinos Mamouras and Lior Seeman. I am grateful to Vasilis Syrgkanis who was an amazing officemate and friend; Hu Fu for teaching me about chinese culture and more importantly how to cook Chinese food; Renato Paes Leme for being an invaluable source of advice and to Lior Seeman for being a close friend and a terrific study partner.

I also thank my parents Ethy and Aharon and my sister Inbal for all their support. Lastly, I wish to thank Shahar for being a great teacher, collaborator and partner (not necessarily in this order). I also thank him for his immense support and in particular for doing the dishes (almost) every day.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 The Online World

Every day billions surf the web, hundreds of millions login to Facebook, tens of millions buy online, hundreds of thousands Wikipedia articles are edited – and the numbers are just going up. The world today is not what it used to be 20 years ago. It is faster, global, collaborative, selective – it is an online world. The world is changing but the people are in part the same people, with the same aspirations, objectives and biases. As a result, some social phenomena reappear in the online world and occasionally take a more extreme form.

This new world is governed by algorithms – from determining which content you will see in your Facebook News Feed, and which ads Google shows when you search "hotel" to movie recommendations on Netflix. As the amount of data stored on each of us grows, the power of these algorithms grows as well. This entails great social and economic potential. The key for fulfilling this potential is obtaining a deeper understanding of the relevant social phenomena.

Among the opportunities the online world encompasses is the possibility of alleviating (or even eliminating) social phenomena that are perceived as negative. It enables us to keep a complete history on users' actions which together with a strict set of rules can automate decision processes. For example, Wikipedia keeps detailed statistics for every editor (number of edits, participance in discussions, new articles created). The detailed statistics together with the online nature of the process allow to allocate credit according to a strict set of rules, almost eliminating the human biases usually involved in this process. As ideal as this might sound, a lack of thorough understanding of the underlying reasons for these phenomena might actually transform this opportunity into a double edge sword.

Example: Polarization

The online world gives us greater freedom in choosing our news sources and friends. It should not come as a surprise that many of us exercise this power to follow news sources that are closer to our views and interact with those that are similar to us. In the offline world we have a limited number of news sources we can choose from; our friends are usually our classmates, coworkers or neighbors. In contrast, the online world offers numerous different news sources to choose from, and a variety of people to interact with. It is therefore much easier to find others who are more similar to us.

The result can be a polarized society (see, e.g., [101]). This is not the worst of it, since exposure to similar opinions alone can lead to even more extreme opinions and isolation from the rest of society. As the psychologist David G. Myers writes [78]:

> As the Internet connects the like-minded and pools their ideas, White supremacists may become more racist, Obama-despisers more hostile, and militia members more terror prone ... In the echo chambers of virtual worlds, as in real worlds, separation + conversation = polarization.

A related by-product of the online world is a possible undermining of traditional establishments that are considered as drivers of diversity and tolerance. One example of this was discussed in [22]: future freshmen begin to cherry pick their future college friends by Facebook even before the school year starts.

Usually users cannot navigate by themselves the vast amounts of information offered by the online world. They need someone to *filter* the information for them. For example, search results provided by Google are based on personal information to provide better filtering. While this personalization approach has many merits, it may enhance polarization [81]. On the other hand, the necessity of filtering also gives us the key for reducing polarization. Instead of completely personalizing search results or feeds we can try to mix in some content that feature slightly different views. By doing this carefully enough, to not alienate users, we can actually use the online world to increase the exposure to different opinions and reduce polarization. While such applications are still speculative, there are clearly mechanisms in place at present – such as the Facebook News Feed – where algorithmic ideas can affect the mixture of information that a user receives. [77] takes the first steps in this direction by suggesting algorithms for creating blog aggregators exposing readers to a greater variety of opinions. Later in the thesis we return to some of these issues at a more technical level, studying opinion formation in Chapters 2 and 3, and culture dynamics in Chapter 4.

1.2 The Algorithmic Approach as Applied in this Thesis

In this thesis we apply the algorithmic approach to analyzing social phenomena. Our approach consists of three different steps: modeling, analysis and design. We now elaborate on each of them separately:

1. **Modeling:** We apply methods from two different fields: (i) Game theory – where we assume that users are rational and strategic agents maximizing

some utility function. (ii) Statistical physics – where users are modeled as particles obeying some set of rules and the properties of the resulting dynamical system are analyzed.

When trying to model a social phenomena either as a game or as a dynamical system, we consider the simplest model which can still teach us something new about the world. In many of the chapters, we start from a very elementary – almost toy like model – and gradually develop it together with our understanding to a more complex model. The challenge here is to decide which parts of the problem are to be included in the model and which should be left aside.

2. **Analysis:** For most of this thesis we will choose to model individuals as rational agents. Such game theoretic modeling raises the question of how will the selfish players play the game? In particular, even when a set of social conventions is not presented as a game, it often creates incentives; combined with individual self-interest, this channels behavior toward certain outcomes and away from others. Game theory suggests that players often play strategies that are in some form of equilibrium (such as Nash equilibrium). This raises a set of computer science questions: Can they find which strategies to play efficiently? How efficient is the worst Nash equilibrium comparing to the optimal solution induced by a central planner (this ratio is termed "the price of anarchy" - see, e.g., [80])? What about the best Nash equilibrium (price of stability [7])? We draw on tools and techniques from the field of Algorithmic Game Theory to answer some of these questions.

3. **Design:** After performing the first two steps and gaining a better understanding of some social phenomena that has negative implications, we can ask a natural question: what can we do about that? Concretely, can we

change the players' utility functions in a way that both makes sense in real life and manages to close the gap between the worst/best Nash equilibrium and the optimal solution? We will address this question for several of the social phenomena we discuss in this thesis.

In this thesis we apply the algorithmic approach to social phenomena in two domains. First, we consider how people form their opinion on either a single issue or a set of issues. Second, we address questions such as how to allocate credit to incentivize contribution and how do competition and long range reasoning affect this allocation. We will now elaborate more on these two domains and describe our progress in studying these questions.

1.3 Opinion and Culture Dynamics in Social Networks

In general, when modeling a social phenomenon on a social network there are many choices to make. Should we consider a fixed network and concentrate only on modeling the phenomenon in question (ignoring in a sense how this network was formed)? Or should we also take into account the process by which the network is evolving? Should the nodes of the network represent individuals? Or maybe they should represent types (where similar individuals are clustered to one type)? At large, different social phenomena call for applying different techniques.

Here we consider opinion formation procedures that take place on a network. Such procedures can be described by a repeated process in which at each step each individual inquires his friends (neighbors in the network) for their opinions and then aggregates this information (possibly together with some private information) to form his opinion for the next step. As the modest task of understanding how opinions form in a fixed social network is not well understood yet, we will start our inquiries there.

A natural staring point is the prominent DeGroot model operating on a fixed network and modeling opinions as real numbers. We will see how casting this model as a game can earn us new insights. Then, we will consider the opinion formation process for issues in which opinions cannot be meaningfully modeled as real numbers. Lastly, we will switch gears and – instead of analyzing the process by which an opinion on a single issue is formed – we will consider how opinions on a set of issues are formed simultaneously. This is usually referred to as culture dynamics. Here, we will present a model generalizing (at least to some extent) the previous works on this subject. As in this work we are interested in the interactions and influences between different "cultures" we will take the statistical physics approach of modeling individuals as particles and consider networks defined on types instead of individuals.

Opinion Formation (Chapter 2)

The starting point for this work is the DeGroot model [35] suggesting the following concrete implementation of the opinion formation process previously described: Each individual holds a numerical opinion and arrives at a shared opinion (consensus) by repeated averaging of his opinion with his neighbors' opinions.

Motivated by the observation that in real life consensus is rarely reached, we study a related sociological model suggested by Friedkin and Johnsen [42]. In this model, individuals have some internal opinions which do not change over the averaging process. This small modification of the model yields a diversity of opinions instead of consensus.

Informally speaking, the question we want to answer is "how much does society lose from disagreement?". To answer this we interpret the repeated averaging as best-response dynamics in an underlying game with natural payoffs. Let us define these payoffs more formally: the cost player i incurs for expressing opinion z_i when the rest of the players express opinions described by the vector z is:

$$c_i(z) = (s_i - z_i)^2 + \sum_{j \in N(i)} w_{i,j}(z_i - z_j)^2,$$

where s_i denotes the internal opinion of player i, $N(i)$ denotes the set of neighbors i has in the network and $w_{i,j}$ is the weight of the edge connecting between i and j.

The repeated averaging converges to the unique Nash equilibrium and enables us to answer the price of anarchy question: what is the cost of disagreement in this model relative to a social optimum? By drawing a connection between these agreement models and extremal problems for generalized eigenvalues we can show that for undirected graphs the price of anarchy is at most 9/8.

We show that for directed graphs the price of anarchy is unbounded. For this class of graphs, a design question becomes relevant: can we improve the cost of the Nash equilibrium by getting individuals to interact with some other individuals? In the language of our model: can we reduce the cost of disagreement at equilibrium by adding edges to the graph? We show that this can be done. However, we also show that finding the optimal set of edges to add according to various natural restrictions is NP-hard. Therefore approximation algorithms are of interest. Potentially these ideas might be used to reduce polarization in society, as we discussed earlier.

Discrete Preferences (Chapter 3)

The model described in Chapter 2 allows players to adopt arbitrarily fine-grained "average" opinions from among any set of options. Most of the dynamics and equilibrium properties of it are driven by this type of averaging. However, sometimes an "average" opinion cannot be meaningfully defined. Consider for example choosing a favorite movie genre; what is the average of drama and action?

This can be classified as an instance of a larger family of games featuring a tension between coordination and individual preferences while the strategies available to the players come from a fixed, discrete set, and where players may have different intrinsic preferences among the possible strategies. It is natural to model the tension among these different preferences by positing a distance function ($d(\cdot, \cdot)$) on the strategy set that determines a notion of similarity among strategies; a players payoff is determined by the distance from her chosen strategy to her preferred strategy and to the strategies chosen by her network neighbors:

$$c_i(z) = \alpha \cdot d(s_i - z_i) + \sum_{j \in N(i)} (1 - \alpha) \cdot d(z_i - z_j).$$

This cost function is defined similarly to the cost function in Chapter 2. As before s_i is i's preferred strategy (or opinion) and z_i is the strategy he plays. α is a parameter that essentially controls the extent to which players are more concerned with their preferred strategies or their network neighbors. The behavior of the game undergoes qualitative changes as we vary α.

Even when there are only two strategies available, this framework already leads to natural questions about a version of the classical Battle of the Sexes problem played on a graph; such questions generalize issues in the study of network coordination games. We are able to shed some light on some of them by studying the game through the price of stability lens (the ratio between the best Nash equilibrium and the optimal solution). We show that in this case the price of stability is non-monotonic in α and exhibits "saw-tooth" behavior with infinitely many local minima in the interval $[0, 1]$.

More generally we show that for $\alpha < 1/2$ the price of stability is equal to 1 for any discrete preference game in which the distance function on the strategies is a tree metric; as a special case, this includes the Battle of the Sexes on a graph. We also show that trees essentially form the maximal family of metrics for which the price of stability is 1, and produce a collection of metrics on which the price of stability converges to an asymptotically tight bound of 2.

Culture Dynamics (Chapter 4)

Up till this point we were discussing the dynamics of a single opinion. However, our opinions on different issues might exhibit correlations, both simple and complex. For example, someone that opposes gun control is more likely to be pro-life. The process by which we shape our opinions is also a contributing factor to the generation of complex correlations. We usually do not form an opinion on each issue separately; we read news and have conversations with our friends on many issues. Naturally, people that share our opinions on some issues have easier time (and more opportunities) influencing us on others. This idea was formalized in the beautiful work of Axelrod [10].

The process that looks at how individuals form opinions on a set of issues (or features) is referred to as cultural dynamics. Axelrod's cultural dynamics model relies on two main forces: *influence* (the tendency of people to become similar to others they interact with) and *selection* (the tendency to be affected most by the behavior of others who are already similar). Influence tends to promote homogeneity within a society[1], while selection frequently causes fragmentation. When both forces are in effect simultaneously, it becomes an interesting question to analyze which societal outcomes should be expected.

One drawback of the Axelrod's model and other relevant models is that the influence and interaction is limited to a few specific patterns. We try to provide a model that keeps true to the influence and selection forces while allowing arbitrary interaction and influence patterns. To allow us to get a more global picture, instead of considering a social network, our model cluster people of the same beliefs to types. The network and the analysis is done on this higher level.

Our model posits an arbitrary graph structure describing which types of people can influence one another: this captures effects based on the fact that people are only influenced by sufficiently similar interaction partners. In a generalization of the model, we introduce another graph structure describing which types of people even so much as come in contact with each other. These restrictions on interaction patterns can significantly alter the dynamics of the process at the population level.

For the basic version of the model, in which all individuals come in contact with all others (the global model), we achieve an essentially complete characterization of (stable) equilibrium outcomes and prove convergence from all starting states. For the other extreme case (the local model), in which individuals only come in contact with others who have the potential to influence them, the underlying process is significantly more complicated; nevertheless we present an analysis for certain graph structures.

[1]Under the assumption that the choice of interaction partners is independent of similarity.

At a higher level, formalizing the distinction between interaction and influence ties back to the issue of polarization discussed earlier. In particular, our work suggests that the choices of who to interact with and who to be influenced by can have a great impact on the final outcome. We later show an example suggesting that increasing the interaction with different types might drive polarization as it dilutes the influence more similar types have on one another.

1.4 Competition and Credit

Incentivizing people to contribute effort is essential in many domains. Many times, workers should be incentivized to perform their work the best they can, scientists to stretch the limits of knowledges and kids to do their schoolwork. The online world readily provides additional concrete examples such as Wikipedia, Stack Overflow, Amazon Mechanical Turk and more. The rise of these crowd sourcing sites brings with it many interesting questions: How do users choose which Wikipedia page to edit or which Amazon Mechanical Turk tasks to take? How do credit allocation rules affect those choices?

Credit here can be interpreted in many different ways: it can be some reputation measure, possibly symbolized by badges as in Stack Overflow or by a promotion as in Wikipedia; alternatively it can simply be a monetary payment. The exact ways credit is perceived and allocated can have profound implications on the users' choices and in turn on the system's efficiency.

We study these questions from two perspectives, both are motivated by analogous questions from the academic world. We first consider how should we best allocate credit in a setting in which each user chooses one project to work on and only cares about maximizing his utility for this choice. After observing that allocating credit fairly can sometime be suboptimal, we tap into our experiences from the research community and show that the observed phenomena of credit misallocation might actually be a solution to this suboptimality.

Next, we add another dimension to the problem – time. We consider how the choices a user makes change once long range effects are taken into account. The analogous question we study here is how academic departments choose which faculty candidates to hire – when the candidate choice is affected by the departments' reputations and the departments care about maximizing their reputation in the long run.

Incentives to Contribute Effort (Chapter 5)

We start our exploration of this topic in a domain where this problem has a long history: Science. Psychologists, philosophers, sociologists and economists all studied the process by which scientists get credit for their work. There are many documented cases in which scientific credit was allocated unfairly. To name just a few: in mathematics, Newton is credited for the discovery of Calculus which was independently discovered by Leibniz; in Biology, Watson and Crick are credited for the discovery of the DNA double helix structure and the contribution of Rosalind Franklin is somewhat neglected. Some more examples are described in [87].

We suggest a novel explanation for this phenomenon. We address two types of unfair credit allocations:

- Researchers who solve more technically difficult problems tend to get more credit than those solving technically easier problems which are equally or even more important. This type of unfairness is a great concern for at least part of the theoretical computer science community [48].

- Famous researchers tend to get more credit than less famous ones for solving the exact same problem independently. This is a well-known phenomenon both in sociology and philosophy of science. Merton termed it the "Matthew effect" [70].

Academic credit is allocated in many ways. Examples include: prizes, papers accepted to prestigious venues, and grants. We suggest a simple game theoretic model that attempts to capture how researchers choose which project to work on. Our model is based on a model of Kitcher from the philosophy of science [62]. Our game-theoretic model assumes that players are selfish and care only about maximizing their own utility. We show that if credit is allocated fairly then the selfish behavior of the agents would lead to a suboptimal result. However, by introducing some unfairness it is possible to direct the players to the optimal assignment. We conclude that the unfairness in allocating credit might not be just the artifact of human biases but can actually be of help to a research community's collective productivity.

Competition and Temporal Dynamics Effects (Chapter 6)

Previously, we discussed how credit should be allocated when users are myopic and only care about maximizing their short term utility. Here, we take another step and try to understand what happens when the choices made are not myopic but also take into account the effects of current actions on the future. Consider for example a firm, currently hiring employees, that has to decide which caliber of candidates to pursue. Should the firm try to increase its reputation by making offers to higher-quality candidates, despite the risk that the candidates might reject the offers and leave the firm empty-handed? Or is it better to play it safe and go for weaker candidates who are more likely to accept the offer? The question acquires an added level of complexity once we take into account the effect one hiring cycle has on the next: hiring better employees in the current cycle increases the firm's reputation, which in turn increases its attractiveness for higher-quality candidates in the next hiring cycle.

Even though we framed the previous question in a setting considerably different than the settings previously discussed, we can still identify in it ingredients common to other settings featuring competition as well. To take just one example from the crowd sourcing world, consider a website such as DesignCrowd.com in which consumers publish various design requests for logos, brochures, etc. Similarly to a firm choosing which candidate to pursue a designer chooses which project to work on. A designer can work on a high profile project which will probably attract more competition but winning it will help him win more high profile projects later on; or he could go for less popular projects, thus increasing the probability that his work is selected and he gets paid.

We develop a model that captures these long-range planning and evolving reputational effects in a setting where two firms repeatedly compete for job candidates over multiple periods. Within this model, we attempt to estimate the effect that reasoning about future hiring cycles has on the efficiency of the job market: do people end up unnecessarily unemployed while the firms compete over the top candidates, or does the evolution of reputation over time eventually converge to a two-tiered system in which the firms each target different parts of the market?

Our model sets up this trade-off in a stylized setting, governed by a parameter q that captures the difference in strength between the top candidate in each hiring cycle and the next best. Using a standard economic model of competition between parties of unequal strength, we show that when q is relatively low, the efficiency of the job market is improved by long-range reputational effects, but when q is relatively high, taking future reputations into account can sometimes reduce the efficiency. While this trade-off arises naturally in the model, the multi-period nature of the strategic reasoning it induces adds new sources of complexity to the analysis. We obtain a tight bound of $\frac{2}{1+\sqrt{1.5}} \approx 0.898$ on the ratio of the welfare at the canonical equilibrium of the model to the socially optimal welfare.

1.5 Bibliographic Notes

Chapter 2 is based on joint work with David Bindel and Jon Kleinberg [17]. Chapter 3 is based on joint work with Flavio Chierichetti and Jon Kleinberg [29]. Chapter 4 is based on joint work with David Kempe, Jon Kleinberg and Alex Slivkins [61]. Chapter 5 is based on joint work with Jon Kleinberg [64]. Chapter 6 is based on joint work with Jon Kleinberg currently under submission. Related to the second theme of this thesis is a joint work with Moshe Babaioff, Shahar Dobzinski and Aviv Zohar [11] on incentives in Bitcoin which is omitted to keep this thesis at a reasonable length.

CHAPTER 2

OPINION FORMATION: OPINIONS AS REAL NUMBERS

2.1 Introduction

An active line of recent work in economic theory has considered processes by which a group of people in a social network can arrive at a shared opinion through a form of repeated averaging [3, 32, 36, 49, 59]. This work builds on a basic model of DeGroot [35], in which we imagine that each person i holds an *opinion* equal to a real number z_i, which might for example represent a position on a political spectrum, or a probability that i assigns to a certain belief. There is a weighted graph $G = (V, E)$ representing a social network, and node i is influenced by the opinions of her neighbors in G, with the edge weights reflecting the extent of this influence. Now, in each time step node i updates her opinion to be a weighted average of her current opinion and the current opinions of her neighbors.

This body of work has developed a set of general conditions under which such processes will converge to a state of *consensus*, in which all nodes hold the same opinion. This emphasis on consensus, however, can only model a specific type of opinion dynamics, where the opinions of the group all come together. As the sociologist David Krackhardt has observed,

> We should not ignore the fact that in the real world consensus is usually not reached. Recognizing this, most traditional social network scientists do not focus on an equilibrium of consensus. They are instead more likely to be concerned with explaining the lack of consensus (the variance) in beliefs and attitudes that appears in actual social influence contexts [66].

In this chapter we study a model of opinion dynamics in which consensus is not reached in general, with the goal of quantifying the inherent social cost of this lack of consensus. To do this, we first need a framework that captures some of the underlying reasons why consensus is not reached, as well as a way of measuring the cost of disagreement.

Lack of Agreement and its Cost

We begin from a variation on the DeGroot model due to Friedkin and Johnsen [42], which posits that each node i maintains a persistent *internal opinion* s_i. This internal opinion remains constant even as node i updates her overall opinion z_i through averaging. More precisely, if $w_{i,j} \geq 0$ denotes the weight on the edge (i,j) in G, then in one time step node i updates her opinion to be the average

$$z_i = \frac{s_i + \sum_{j \in N(i)} w_{i,j} z_j}{1 + \sum_{j \in N(i)} w_{i,j}}, \tag{2.1}$$

where $N(i)$ denotes the set of neighbors of i in G. Note that, in general, the presence of s_i as a constant in each iteration prevents repeated averaging from bringing all nodes to the same opinion. In this way, the model distinguishes between an individual's intrinsic belief s_i and her overall opinion z_i; the latter represents a compromise between the persistent value of s_i and the expressed opinions of others to whom i is connected. This distinction between s_i and z_i also has parallels in empirical work that seeks to trace deeply held opinions such as political orientations back to differences in education and background, and even to explore genetic bases for such patterns of variation [5].

Now, if consensus is not reached, how should we quantify the cost of this lack of consensus? Here we observe that since the standard models use averaging as their basic mechanism, we can equivalently view nodes' actions in each time step as myopically optimizing a quadratic cost function: Updating z_i as in Equation (2.1) is the same as choosing z_i to minimize

$$(z_i - s_i)^2 + \sum_{j \in N(i)} w_{i,j}(z_i - z_j)^2. \tag{2.2}$$

We therefore take this as the *cost* that i incurs by choosing a given value of z_i, so that averaging becomes a form of cost minimization.

Given this view, we can think of repeated averaging as the trajectory of best-response dynamics in a one-shot, complete information game played by the nodes in V, where i's strategy is a choice of opinion z_i, and her payoff is the negative of the cost in Equation (2.2).

Nash Equilibrium and Social Optimality in a Game of Opinion Formation

It was already observed in [42] that repeated averaging always converges. In Section 6.2 we repeat the convergence proof and show it actually converges to the unique Nash equilibrium of the game defined by the individual cost functions in (2.2): each node i has an opinion x_i that is the weighted average of i's internal opinion and the (equilibrium) opinions of i's neighbors. This equilibrium will not in general correspond to the *social optimum*, the vector of node opinions y that minimizes the *social cost*, defined to be sum of all players' costs: $c(y) = \sum_i \left((y_i - s_i)^2 + \sum_{j \in N(i)} w_{i,j}(y_i - y_j)^2 \right)$.

The sub-optimality of the Nash equilibrium can be viewed in terms of the *externality* created by a player's personal optimization: by refusing to move further toward their neighbors' opinions, players can cause additional cost to be incurred by these neighbors. In fact we can view the problem of minimizing social cost for this game as a type of *metric labeling problem* [23, 65], albeit a polynomial-time solvable case of the problem with a non-metric quadratic distance function on the real numbers [1]: we seek node labels that balance the value of a cost function at each node (capturing disagreement with node-level information) and a cost function for label disagreement across edges. Viewed this way, the sub-optimality of Nash equilibrium becomes a kind of sub-optimality for local optimization.

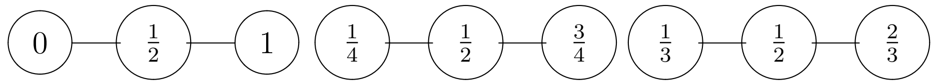
A natural question for this game is thus the *price of anarchy*, defined as the ratio between the cost of the Nash equilibrium and the cost of the optimal solution.

Our Results: Undirected Graphs

The model we have described can be used as stated in both undirected and directed graphs — the only difference is in whether i's neighbor set $N(i)$ represents the nodes with whom i is connected by undirected edges, or to whom i links with directed edges. However, the behavior of the price of anarchy is very different in undirected and directed graphs, and so we analyze them separately, beginning with the undirected case.

As an example of how a sub-optimal social cost can arise at equilibrium in an undirected graph, consider the graph depicted in Figure 2.1 — a three-node path with uniform edge weights in which the nodes have internal opinions 0, 1/2, and 1 respectively. As shown in the figure, the ratio between the social cost of the Nash equilibrium and the social optimum is 9/8. Intuitively, the reason for the higher cost of the Nash equilibrium is that the center node — by symmetry — cannot

[1] In the next chapter we will consider the version of the question in which the distance function is indeed a metric.



(a) *internal opinions.* (b) *Nash equilibrium.* (c) *optimal solution.*

Figure 2.1: An example in which the two players on the sides do not compromise by the optimal amount, given that the player in the middle should not shift her opinion. The social cost of the optimal set of opinions is 1/3, while the cost of the Nash equilibrium is 3/8.

usefully shift her opinion in either direction, and so to achieve optimality the two outer nodes need to compromise more than they want to at equilibrium. This is a reflection of the externality discussed above, and it is the qualitative source of sub-optimality in general for equilibrium opinions — nodes move in the direction of their neighbors, but not sufficiently to achieve the globally minimum social cost.

Our first result is that the very simple example in Figure 2.1 is in fact extremal for undirected graphs: we show that for any undirected graph G and any internal opinions vector s, the price of anarchy is at most 9/8. We prove this by casting the question as an extremal problem for quadratic forms, and analyzing the resulting structure using eigenvalues of the Laplacian matrix of G. From this, we obtain a characterization of the set of graphs G for which some internal opinions vector s yields a price of anarchy of 9/8.

We show that this bound of 9/8 continues to hold even for some generalizations of the model — when nodes i have different coefficients w_i on the cost terms for their internal opinions, and when certain nodes are "fixed" and simply do not modify their opinions.

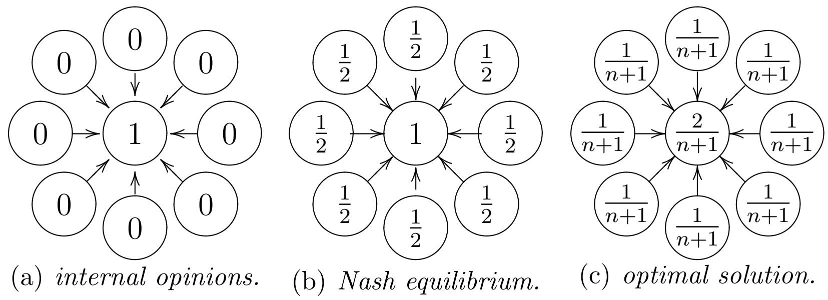


Figure 2.2: An example demonstrating that the price of anarchy of a directed graph can be unbounded.

Our Results: Directed Graphs

We next consider the case in which G is a directed graph; the form of the cost functions remains exactly the same, with directed edges playing the role of undirected ones, but the range of possible behaviors in the model becomes very different. This is due to the fact that nodes can now exert a large influence over the network without being influenced themselves. Indeed, as Matt Jackson has observed, directed versions of repeated averaging models can naturally incorporate "external" media sources; we simply include nodes with no outgoing links, so that they maintain their internal opinion [59].

We first show that the spectral machinery developed for analyzing undirected graphs can be extended to the directed case; through an approach based on generalized eigenvalue problems we can efficiently compute the maximum possible price of anarchy, over all choices of internal node opinions, on a given graph G. However, in contrast to the case of undirected graphs, the price of anarchy can be very large in some instances; the simple example in Figure 2.2 shows a case in which $n-1$ nodes with internal opinion 0 all link to a single node that has internal opinion 1 and no out-going edges, producing an in-directed star. As a result, the social cost of the Nash equilibrium is $\frac{1}{2}(n-1)$, whereas the minimum social cost is at most

1, since the player at the center of the star could simply shift her opinion to 0. Intuitively, this corresponds to a type of social network in which the whole group pays attention to a single influential "leader" or "celebrity"; this drags people's opinions far from their internal opinions s_i, creating a large social cost. Unfortunately, the leader is essentially unaware of the people paying attention to her, and hence has no incentive to modify her opinion in a direction that could greatly reduce the social cost.

In Section 2.4 we show that a price of anarchy lower-bounded by a polynomial in n can in fact be achieved in directed graphs of constant degree, so this behavior is not simply a consequence of large in-degree. It thus becomes a basic question whether there are natural classes of directed graphs, and even bounded-degree directed graphs, for which a constant price of anarchy is achievable.

Unweighted Eulerian directed graphs are a natural class to consider — first, because they generalize undirected graphs, and second, because they capture the idea that at least at a local level no node has an asymmetric effect on the system. We use our framework for directed graphs to derive two bounds on the price of anarchy of Eulerian graphs: For Eulerian graphs with maximum degree Δ we obtain a bound of $\Delta + 1$ on the price of anarchy. For the subclass of Eulerian *antisymmetric* directed graphs[2] with maximum degree Δ and edge expansion α, we show a bound of $O(\Delta^2\alpha^{-2})$ on the price of anarchy.

[2]An Eulerian *antisymmetric* directed graph is an Eulerian graph that does not contain any pair of oppositely oriented edges (i, j) and (j, i).

Our Results: Modifying the Network

Finally, we consider an algorithmic problem within this framework of opinion formation. The question is the following: if we have the ability to modify the edges in the network (subject to certain constraints), how should we do this to reduce the social cost of the Nash equilibrium by as much as possible? This is a natural question both as a self-contained issue within the mathematical framework of opinion formation, and also as discussed earlier in the context of applications: many social media sites overtly and algorithmically consider how to balance the mix of news content [4, 13, 76, 77] and also the mix of social content [12, 100] that they expose their users to, so as to optimize user engagement on the site.

Adding edges to reduce the social cost has an intuitive basis: it seems natural that exposing people to others with different opinions can reduce the extent of disagreement within the group. When one looks at the form of the social cost $c(y)$, however, there is something slightly counter-intuitive about the idea of adding edges to improve the situation: the social cost is a sum of quadratic terms, and by adding edges to G we are simply adding further quadratic terms to the cost. For this reason, in fact, adding edges to G can never improve the optimal social cost. But adding edges *can* improve the social cost of the Nash equilibrium, and sometimes by a significant amount — the point is that adding terms to the cost function shifts the equilibrium itself, which can sometimes more than offset the additional terms. For example, if we add a single edge from the center of the star in Figure 2.2 to one of the leaves, then the center will shift her opinion to $2/3$ in equilibrium, causing all the leaves to shift their opinions to $1/3$, and resulting in a $\Theta(n)$ improvement in the social cost. In this case, once the leader pays attention to even a single member of the group, the social cost improves dramatically.

We focus on three main variants on this question: when all edges must be added *to* a specific node (as in the case when a site can modify the amount of attention directed to a media source or celebrity); when all edges must be added *from* a specific node (as in the case when a particular media site tries to shift its location in the space of opinions by blending in content from others); and when edges can be added between any pair of nodes in the network (as in the case when a social networking site evaluates modifications to its feeds of content from one user to another [12, 100]).

In Section 2.5 we show that, in the previously discussed variants, the problem of where to add edges to optimally reduce the social cost is NP-hard. On the positive side, we obtain a $\frac{9}{4}$-approximation algorithm when edges can be added between arbitrary pairs of nodes.

2.2 Convergence and Nash Equilibrium Uniqueness

In this section, we show that the opinion game has a unique Nash equilibrium to which the repeated averaging process converges. Our proof makes use of some matrix notation which will also turn out to be useful for the rest of the chapter. Given an edge-weighted graph, we write W for the weighted adjacency matrix, and let D be the diagonal matrix of node degrees $d_i = \sum_{j \neq i} w_{i,j}$. The *weighted graph Laplacian* is $L = D - W$. A useful fact about the Laplacian is that it is a *positive semidefinite* matrix. This implies that all of its eigenvalues are non-negative.

We are now ready to show that the Nash equilibrium is unique and is the limit of the repeated averaging process.

Claim 2.2.1 *The opinion game admits a unique Nash equilibrium.*

Proof: In the Nash equilibrium x each player chooses an opinion which minimizes her cost; in terms of the derivatives of the cost functions, this implies that $c_i'(x) = 0$ for all i. Thus, to find the players' opinions in the Nash equilibrium we should solve the following system of equations: $\forall i \ (x_i - s_i) + \sum_{j \in N(i)} w_{i,j}(x_i - x_j) = 0$. Therefore in the Nash equilibrium each player holds an opinion which is a weighted average of her internal opinion and the Nash equilibrium opinions of all her neighbors. After some rearranging we get that $\forall i \ \sum_{j \in N(i)} w_{i,j}(x_i - x_j) + x_i = s_i$. This system of equations can be succinctly written as $(L + I)x = s$. Observe that, $L + I$ is a positive definite matrix as the Laplacian matrix is a positive semidefinite matrix and once we add the identity matrix to it all its eigenvalues are strictly greater than 0. Since a positive definite matrix is invertible, we have that the unique Nash equilibrium is $x = (L + I)^{-1}s$. □

Claim 2.2.2 *The repeated averaging process defined by the Friedkin and Johnsen update rule (Equation 2.1) converges to the Nash equilibrium $x = (L + I)^{-1}s$.*

Proof: Changed back to the previous version of the proof and added that the fixed point is a Nash equilibrium. I think this version is easier to understand even though it is longer. Let $z(t)$ be the opinions vector at time t. Let b be a "normalized" internal opinions vector defined as follows $b_i = \dfrac{s_i}{1 + \sum_{j \in N(i)} w_{i,j}}$. We also define a matrix R to be the "normalized" adjacency matrix of G: $R_{i,i} = 0$ and for $i \neq l$, $R_{i,l} = \dfrac{w_{i,l}}{1 + \sum_{j \in N(i)} w_{i,j}}$. With this notation in place the update rule in Equation 2.1 can now be written as $z(t) = F(z(t-1)) = Rz(t-1) + b$. If v^1 and v^2 are arbitrary vectors, then

$$\|F(v^1) - F(v^2)\|_\infty = \|R(v^1 - v^2)\|_\infty \leq \|R\|_\infty \|v^1 - v^2\|_\infty$$

where $\|v\|_\infty = \max_i |v_i|$ is the usual max norm and the associated matrix norm $\|R\|_\infty$ is given by

$$\|R\|_\infty = \max_i \sum_j |R_{i,j}| = \max_i \frac{\sum_{j \in N(i)} w_{i,j}}{1 + \sum_{j \in N(i)} w_{i,j}} < 1.$$

Thus, F is a contraction mapping in the max norm, and so the iteration converges to a unique fixed point. To compute this fixed point, we observe that $R = (D+I)^{-1}W$ and that $b = (D+I)^{-1}s$. Thus, we can alternatively write the repeated averaging process as $z(t+1) = (D+I)^{-1}(Wz(t) + s)$. After some rearranging we get that the fixed point of this process is indeed the Nash equilibrium: $x = (D+I-W)^{-1}s = (L+I)^{-1}s$

□

2.3 Undirected Graphs

We first consider the case of undirected graphs and later handle the more general case of directed graphs. The main result in this section is a tight bound on the price of anarchy for the opinion-formation game in undirected graphs. After this, we discuss two slight extensions to the model: in the first, each player can put a different amount of weight on her internal opinion; and in the second, each player has several fixed opinions she listens to instead of an internal opinion. We show that both models can be reduced to the basic form of the model which we study first.

For undirected graphs we can simplify the social cost to the following form:

$$c(z) = \sum_i (z_i - s_i)^2 + 2 \sum_{(i,j) \in E, i > j} w_{i,j}(z_i - z_j)^2.$$

We write this concisely in matrix form as $c(z) = z^T A z + ||z - s||^2$, where the matrix $A = 2L$ captures the tension on the edges. Recall that L is the weighted Laplacian of G and is defined by setting $L_{i,i} = \sum_{j \in N(i)} w_{i,j}$ and $L_{i,j} = -w_{i,j}$. The optimal solution is the y minimizing $c(\cdot)$. By taking derivatives, we see that the optimal solution satisfies $(A + I)y = s$. Since the Laplacian of a graph is a positive semidefinite matrix, it follows that $A + I$ is positive definite. Therefore, $(A+I)y = s$ has a unique solution: $y = (A+I)^{-1}s$. In comparison, as we showed in Claim 2.2.1, the Nash equilibrium is $x = (L+I)^{-1}s = (\frac{1}{2}A + I)^{-1}s$.

We now begin our discussion on the price of anarchy (PoA) of the opinion game — the ratio between the cost of the optimal solution and the cost of the Nash equilibrium.

Our main theorem is that the price of anarchy of the opinion game is at most 9/8. Before proceeding to prove the theorem we present a simple upper bound of 2 on the PoA for undirected graphs. To see why this holds, note that the Nash equilibrium actually minimizes the function $z^T(\frac{1}{2}A)z + ||z - s||^2$ (one can verify that this function's partial derivatives are the system of equations defining the Nash equilibrium). This allows us to write the following bound on the PoA:

$$\begin{aligned} PoA = \frac{c(x)}{c(y)} &\leq \frac{2(x^T(\frac{1}{2}A)x + ||x-s||^2)}{c(y)} \\ &\leq \frac{2(y^T(\frac{1}{2}A)y + ||y-s||^2)}{c(y)} \\ &\leq \frac{2c(y)}{c(y)} = 2. \end{aligned}$$

We note that this bound holds only for the undirected case, as in the directed case the Nash equilibrium does not minimize $z^T(\frac{1}{2}A)z + ||z-s||^2$ anymore.

We now state the main theorem of this section.

Theorem 2.3.1 *For any graph G and any internal opinions vector s, the price of anarchy of the opinion game is at most* $9/8$.

Proof: The crux of the proof is relating the price of anarchy of an instance to the eigenvalues of its Laplacian. Specifically, we characterize the graphs and internal opinion vectors with maximal PoA. In these worst-case instances at least one eigenvalue of the Laplacian is exactly 1, and the vector of internal opinions is a linear combination of the eigenvectors associated with the eigenvalues 1, plus a possible constant shift for each connected component. As a first step we consider two matrices B and C that arise by plugging the Nash equilibrium and optimal solution we previously computed into the cost function and applying simple algebraic manipulations:

$$
\begin{aligned}
c(z) &= \|z - s\|^2 + z^T A z \\
&= (z^T z - 2s^T z + s^T s) + z^T A z \\
&= z^T (A + I) z - 2s^T z + s^T s.
\end{aligned}
$$

$$
\begin{aligned}
c(y) &= s^T (A+I)^{-1}(A+I)(A+I)^{-1} s - 2s^T (A+I)^{-1} s + s^T s \\
&= s^T [\underbrace{I - (A+I)^{-1}}_{B}] s.
\end{aligned}
$$

$$
\begin{aligned}
c(x) &= s^T (L+I)^{-1}(A+I)(L+I)^{-1} s - 2s^T (L+I)^{-1} s + s^T s \\
&= s^T (L+I)^{-1} \left[A + I - 2(L+I) + (L+I)^2\right] (L+I)^{-1} s \\
&= s^T [\underbrace{(L+I)^{-1}(A+L^2)(L+I)^{-1}}_{C}] s.
\end{aligned}
$$

Next, we show that the matrices A, B, C are *simultaneously diagonalizable*: there exists an orthogonal matrix Q such that $A = Q\Lambda^A Q^T$, $B = Q\Lambda^B Q^T$ and

$C = Q\Lambda^C Q^T$, where for a matrix M the notation Λ^M represents a diagonal matrix with the eigenvalues $\lambda_1^M, \ldots, \lambda_n^M$ of M on the diagonal.

Lemma 2.3.2 *A,B and C are simultaneously diagonalizable by a matrix Q whose columns are eigenvectors of A.*

Proof: It is a standard fact that any real symmetric matrix M can be diagonalized by an orthogonal matrix Q such that $M = Q\Lambda^M Q^T$. Q's columns are eigenvectors of M which are orthogonal to each other and have a norm of one. Thus in order to show that A, B and C can be diagonalized with the same matrix Q it is enough to show that all three are symmetric and have the same eigenvectors. For this we use the following basic fact:

> *If λ^N is an eigenvalue of N, λ^M is an eigenvalue of M and w is an eigenvector of both then:*
>
> 1. *$\frac{1}{\lambda^M}$ is an eigenvalue of M^{-1} and w is an eigenvector of M^{-1}.*
> 2. *$\lambda^N + \lambda^M$ is an eigenvalue of $N + M$ and w is an eigenvector of $N + M$.*
> 3. *$\lambda^N \cdot \lambda^M$ is an eigenvalue of NM and w is an eigenvector of NM.*

From this we can show that any eigenvector of A is also an eigenvector of B and C. Recall that A is a symmetric matrix, thus, it has n orthogonal eigenvectors which implies that A,B and C are all symmetric and share the same basis of eigenvectors. Therefore A,B and C are simultaneously diagonalizable. $\square$

We can now express the PoA as a function of the eigenvalues of B and C. By defining $\hat{s} = Q^T s$ we have:

$$\begin{aligned} PoA &= \frac{c(x)}{c(y)} = \frac{s^T C s}{s^T B s} = \frac{s^T Q \Lambda^C Q^T s}{s^T Q \Lambda^B Q^T s} \\ &= \frac{\hat{s}^T \Lambda^C \hat{s}}{\hat{s}^T \Lambda^B \hat{s}} = \frac{\sum_{i=1}^n \lambda_i^C \hat{s}_i^2}{\sum_{i=1}^n \lambda_i^B \hat{s}_i^2} \leq \max_i \frac{\lambda_i^C}{\lambda_i^B} \end{aligned}$$

The final step of the proof consists of expressing λ_i^C and λ_i^B as functions of the eigenvalues of A (denoted by λ_i) and finding the value for λ_i maximizing the ratio between λ_i^C and λ_i^B.

Lemma 2.3.3 $\max_i \frac{\lambda_i^C}{\lambda_i^B} \leq 9/8$. *The bound is tight if and only if there exists an i such that $\lambda_i = 2$.*

Proof: Using the basic facts about eigenvalues which were mentioned in the proof of Lemma 2.3.2, we get:

$$\begin{aligned} \lambda_i^B &= 1 - \frac{1}{\lambda_i + 1} = \frac{\lambda_i}{\lambda_i + 1}. \\ \lambda_i^C &= \frac{\lambda_i + \lambda_i^2/4}{(\lambda_i/2 + 1)^2} = \frac{\lambda_i^2 + 4\lambda_i}{(\lambda_i + 2)^2}. \end{aligned}$$

We can now write $\lambda_i^C / \lambda_i^B = \phi(\lambda_i)$, where ϕ is a simple rational function:

$$\phi(\lambda) = \frac{(\lambda^2 + 4\lambda)/(\lambda + 2)^2}{\lambda/(\lambda + 1)} = \frac{(\lambda^2 + 4\lambda)(\lambda + 1)}{(\lambda + 2)^2 \lambda} = \frac{(\lambda + 4)(\lambda + 1)}{(\lambda + 2)^2} = \frac{\lambda^2 + 5\lambda + 4}{\lambda^2 + 4\lambda + 4}.$$

By taking the derivative of ϕ, we find that ϕ is maximized over all $\lambda \geq 0$ at $\lambda = 2$ and $\phi(2) = 9/8$.

The eigenvalues λ_i are all non-negative, so it is always true that $\max_i \phi(\lambda_i) \leq 9/8$. If 2 is an eigenvalue of A (and hence 1 is an eigenvalue of the Laplacian) then there exists an internal opinions vector s for which the PoA is 9/8. What is the internal opinions vector maximizing the PoA? Rewriting our expression from above, we have

$$PoA = \frac{\sum_{i=1}^{n} \hat{s}_i^2 \lambda_i^B \phi(\lambda_i)}{\sum_{i=1}^{n} \hat{s}_i^2 \lambda_i^B},$$

i.e. the price of anarchy is a weighted average of the values $\phi(\lambda_i)$, where the weights are given by $\hat{s}_i^2 \lambda_i^B$. The only way to achieve the maximum value is if the only nonzero weights are on eigenvalues maximizing $\phi(\lambda)$. Because λ_i^B is positive whenever λ_i is positive, this means that to achieve a PoA of 9/8, $\hat{s}_i^2$ can only be nonzero if $\lambda_i = 2$ or $\lambda_i = 0$. Recall that $s = Q\hat{s}$, where the columns of Q are the eigenvectors of A. Thus, any internal opinion vector that is an eigenvector of A with eigenvalue 2 plus some null vector of A will achieve the maximal price of anarchy. In particular, since the all-ones vector is always an eigenvector of the weighted Laplacian matrix (and thus A) we have that there exists an internal opinions vector where all the opinions are positive for which the maximal PoA is achieved. □

With Lemma 2.3.3, we have completed the proof of Theorem 2.3.1. □

Corollary 2.3.4 *We can scale the weights of any graph to make its PoA be* 9/8*. If α is the scaling factor for the weights, then the eigenvalues of the scaled A matrix are $\alpha\lambda_i$. Therefore by choosing $\alpha = \frac{2}{\lambda_i}$ for any eigenvalue other than* 0 *we get that there exists an internal opinions vector for which the PoA is* 9/8*.*

2.3.1 Arbitrary Node Weights and Players with Fixed Opinions

Our first extension is a model in which different people put different weights on their internal opinion. In this extension, each node in the graph has a strictly positive weight γ_i and the cost function is

$$c^1(z) = \sum_i \left(\gamma_i (z_i - s_i)^2 + \sum_{j \in N(i)} w_{i,j}(z_i - z_j)^2 \right) = (z-s)^T \Gamma (z-s) + z^T A z,$$

where Γ is the diagonal matrix of node weights γ_i and $A = 2L$ is defined as in the previous section. In the next claim we show that the bound of 9/8 on the PoA holds even in this model:

Claim 2.3.5 *The PoA of the game with arbitrary strictly positive node weights is bounded by* 9/8.

Proof: We define the scaled variables $\hat{z} = \Gamma^{1/2} z$, $\hat{s} = \Gamma^{1/2} s$ and the scaled matrices $\hat{A} = \Gamma^{-1/2} A \Gamma^{-1/2}$ and $\hat{L} = \Gamma^{-1/2} L \Gamma^{-1/2} = \hat{A}/2$. Now, in terms of the scaled variables, we have

$$\begin{aligned} c^1(z) &= (z-s)^T \Gamma (z-s) + z^T A z \\ &= (\hat{z} - \hat{s})^T (\hat{z} - \hat{s}) + \hat{z}^T \hat{A} \hat{z} \\ &= \|\hat{z} - \hat{s}\|^2 + \hat{z}^T \hat{A} \hat{z}, \end{aligned}$$

and the Nash equation $(L + \Gamma)x = \Gamma s$ can similarly be multiplied by $\Gamma^{-1/2}$ and rewritten in terms of the scaled variables as $(\hat{L} + I)\hat{x} = \hat{s}$. Thus, in terms of the scaled variables the problem takes exactly the same form as in the previous section, and the argument of Theorem 2.3.1 applies.

$\square$

Next we show how to handle the case in which a subset of the players may have node weights of 0, which can equivalently be viewed as a set of players who have no internal opinion at all.

Lemma 2.3.6 *The PoA of the game with non-negative node weights is bounded by* 9/8.

Proof: We assume without loss of generality that every connected component of G includes at least one player i with weight $\gamma_i > 0$. Observe that otherwise, the cost associated with this connected component both in the Nash equilibrium and in the optimal solution is 0 and hence we can ignore this component. We begin by showing the matrix $L + \Gamma$ is positive definite, and thus nonsingular. Both L and Γ are semidefinite, so we only need to show that their sum is definite. If $z^T Lz = 0$, then z is constant over each connected component of G; and if $z^T \Gamma z = 0$, then z is zero for at least one node in each component. Therefore, $z^T Lz = z^T \Gamma z = 0$ if and only if $z = 0$. Because A=2L , the same argument shows that $A + \Gamma$ is nonsingular. Thus, the Nash equilibrium equations $(L + \Gamma)x = \Gamma s$ and the social optimality equation $(A + \Gamma)y = \Gamma s$ each have a unique solution.

To show the PoA is bounded by 9/8, consider a modified problem with weights $\gamma_i + \epsilon$ for $\epsilon \geq 0$. Because $L + \Gamma + \epsilon I$ and $A + \Gamma + \epsilon I$ are nonsingular for all $\epsilon \geq 0$, the Nash equilibrium $x^\epsilon = (L + \Gamma + \epsilon I)^{-1}(\Gamma + \epsilon I)s$ and the social optimum $y^\epsilon = (A + \Gamma + \epsilon I)^{-1}(\Gamma + \epsilon I)s$ are both continuous functions of ϵ. The modified social cost $c^{1,\epsilon}(z)$ is a continuous function of both ϵ and z. Therefore,

$$PoA^\epsilon = \frac{c^{1,\epsilon}(x^\epsilon)}{c^{1,\epsilon}(y^\epsilon)}$$

is a continuous function in ϵ, provided that $c^{1,\epsilon}(y^\epsilon) \neq 0$. Thus, there are two cases:

1. At least one connected component of G includes nodes i and j with γ_i and γ_j both positive and $s_i \neq s_j$. In this case $c^{1,\epsilon}(y^\epsilon) \neq 0$ for any $\epsilon > 0$, and PoA^ϵ is a continuous function of ϵ for any $\epsilon \geq 0$. Because $PoA^\epsilon \leq 9/8$ for $\epsilon > 0$, continuity implies $PoA^0 \leq 9/8$.

2. No connected component of G includes nodes i and j with γ_i and γ_j both positive and $s_i \neq s_j$. In this case, $x^\epsilon = y^\epsilon$ is constant on each connected compoment, and $c^{1,\epsilon}(x^\epsilon) = c^{1,\epsilon}(y^\epsilon) = 0$ for any $\epsilon \geq 0$. In this case, we define the price of anarchy to be 1, which is bounded by 9/8.

$\square$

In the second model we present, some nodes have *fixed opinions* and others do not have an internal opinion at all. We partition the nodes into two sets A and B. Nodes in B are completely fixed in their opinion and are non-strategic, while nodes in A have no internal opinion – they simply want to choose an opinion that minimizes their disagreement with their neighbors (which may include a mix of nodes in A and B). We can think of nodes in A as people forming their opinion and of nodes in B as news sources with a specific *fixed* orientation. We denote the fixed opinion of a node $j \in B$ by s_j. The cost for player $i \in A$ in this model is

$$c_i^2(z) = \sum_{j:(i,j)\in E_{AB}} w_{i,j}(z_i - s_j)^2 + \sum_{j:(i,j)\in E_{AA}} w_{i,j}(z_i - z_j)^2,$$

where E_{AB} and E_{AA} denote the edges between A and B and between A and A, respectively. Note that this clearly generalizes the original model, since we can construct a distinct node in B to represent each internal opinion. Next, we perform the reduction in the opposite direction, reducing this model to the basic model. To do this, we assign each node an internal opinion equal to the weighted average of the opinions of her fixed neighbors, and a weight equal to the sum of her fixed neighbors' weights. We then show that the PoA of the fixed opinion model is bounded by the PoA of the basic model and thus get:

Proposition 2.3.7 *The PoA of the fixed opinion model is at most* 9/8 .

Proof: We assume without loss of generality that any component of the subgraph G' on nodes in A has an edge to some node in B. This is a valid assumption as the cost associated with any component that do not have such an edge is 0 both in the Nash equilibrium and in the optimal solution. Hence, we can ignore such components. For each player $i \in A$ with edges into B, define the total edge weight γ_i and the weighted mean s_i and variance ν_i of the fixed opinions of neighbors as

$$\begin{aligned}\gamma_i &= \sum_{j:(i,j)\in E_{AB}} w_{i,j} \\ s_i &= \frac{1}{\gamma_i} \sum_{j:(i,j)\in E_{AB}} w_{i,j} s_j \\ \nu_i &= \frac{1}{\gamma_i} \sum_{j:(i,j)\in E_{AB}} w_{i,j}(s_j^2 - s_i^2).\end{aligned}$$

For convenience, let γ_i, s_i, and ν_i be zero for nodes with no edges into B. Note that $\sum_{j:(i,j)\in E_{AB}} w_{i,j}(z_i - s_j)^2$

$$\begin{aligned}&= \left(\sum_{j:(i,j)\in E_{AB}} w_{i,j}\right) z_i^2 - 2\left(\sum_{j:(i,j)\in E_{AB}} w_{i,j} s_j\right) z_i + \left(\sum_{j:(i,j)\in E_{AB}} w_{i,j} s_j^2\right) \\ &= \gamma_i z_i^2 - 2\gamma_i s_i z_i + \gamma_i(\nu_i + s_i^2) \\ &= \gamma_i (z_i - s_i)^2 + \gamma_i \nu_i.\end{aligned}$$

Thus, the cost for a player $i \in A$ in the fixed opinion game is

$$c_i^2(z) = \gamma_i \nu_i + \gamma_i (z_i - s_i)^2 + \sum_{j:(i,j)\in E_{AA}} w_{i,j}(z_i - z_j)^2 = c_i^1(z) + \gamma_i \nu_i,$$

where $c^1(z)$ is the cost for the game played on the subgraph G' over the nodes in A with node weights γ_i and intrinsic opinions s_i. Similarly, the social cost in the fixed opinion game is

$$c^2(z) = c^1(z) + \nu,$$

where $\nu = \sum_i \gamma_i \nu_i$. Because the variance costs are independent of z, we have that $\partial c_i^2/\partial z_j = \partial c_i^1/\partial z_j$; thus, the equations for the Nash opinion vector and the optimal opinion vector are the same as in the previous model with arbitrary weights. From Lemma 2.3.6, we know $c^1(x) \le \frac{9}{8}c^2(y)$; therefore,

$$c^2(x) = c^1(x) + \nu \le \frac{9}{8}c^1(y) + \nu \le \frac{9}{8}c^1(y) + \frac{9}{8}\nu = \frac{9}{8}c^2(y).$$

Hence, the price of anarchy is again bounded by 9/8.

$\square$

2.4 Directed Graphs

We begin our discussion of directed graphs with an example showing that the price of anarchy can be unbounded even for graphs with bounded degrees. Our main result in this section is that we can nevertheless develop spectral methods extending those in Section 2.3 to find internal opinions that maximize the PoA for a given graph. Using this approach, we identify classes of directed graphs with good PoA bounds.

In the introduction we have seen that the PoA of an in-directed star can be unbounded. As a first question, we ask whether this is solely a consequence of the unbounded maximum in-degree of this graph, or whether it is possible to have an unbounded PoA for a graph with bounded degrees. Our next example shows that one can obtain a large PoA even when all degrees are bounded: we show that the PoA of a bounded degree tree can be $\Theta(n^c)$, where $c \le 1$ is a constant depending on the in-degrees of the nodes in the tree.

Example 2.4.1 *Let G be a 2^k-ary tree of depth $\log_{2^k} n$ in which the internal opinion of the root is 1 and the internal opinion of every other node is 0. All edges are directed toward the root. In the Nash equilibrium all nodes at layer i hold the same opinion, which is 2^{-i}. (The root is defined to be at layer 0.) The cost of a node at layer i is $2 \cdot 2^{-2i}$. Since there are 2^{ik} nodes at layer i, the total social cost at Nash equilibrium is $\sum_{i=1}^{\log_{2^k} n} 2^{ik}2^{1-2i} = 2\sum_{i=1}^{\log_{2^k} n} 2^{(k-2)i}$. For $k > 2$ this cost is $2^{k-1}\frac{(2^{k-2})^{\log_{2^k} n} - 1}{2^{k-2} - 1} = 2^{k-1}\frac{n^{\frac{k-2}{k}} - 1}{2^{k-2} - 1}$. The cost of the optimal solution is at most 1; in fact it is very close to 1, since in order to reduce the cost the root should hold an opinion of ϵ very close to 0, which makes the root's cost approximately 1. Therefore the PoA is $\Theta(n^{\frac{k-2}{k}})$. It is instructive to consider the PoA for extreme values of k. For $k = 2$, the PoA is $\Theta(\log n)$, while for $k = \log n$ we recover the in-directed star from the introduction where the PoA is $\Theta(n)$. For intermediate values of k, the PoA is $\Theta(n^c)$. For example, for $k = 3$ we get that the PoA is $\Theta(n^{\frac{1}{3}})$.*

For directed graphs we do not consider the generalization to arbitrary node weights (along the lines of Section 2.3.1), noting instead that introducing node weights to directed graphs can have a severe effect on the PoA. That is, even in graphs containing only two nodes, introducing arbitrary node weights can make the PoA unbounded. For example, consider a graph with two nodes i and j. Node i has an internal opinion of 0 and a node weight of 1, while node j has an internal opinion of 1 and a node weight of ϵ. There is a directed edge (i, j) with weight 1. The cost of the Nash equilibrium is $1/2$, but the social cost of the optimal solution is smaller than ϵ. Thus, from now on we restrict our attention to uniform node weights.

2.4.1 The Price of Anarchy in a General Graph

For directed graphs we can define matrices B and C, similarly to their definition for undirected graphs, such that the cost of the optimal solution and the cost of the Nash equilibrium are respectively $c(y) = s^T Bs$ and $c(x) = s^T Cs$. Recall that the matrix A is used in the social cost function to capture the cost associated with the edges of the graph (disagreement between neighbors). We define it for directed graphs by setting $A_{i,j} = -w_{i,j} - w_{j,i}$ for $i \neq j$ and $A_{i,i} = \sum_{j \in N(i)} w_{i,j} + \sum_{\{j | i \in N(j)\}} w_{j,i}$. The matrix A can be also interpreted as the weighted Laplacian for an undirected graph where the weight on the undirected edge (i, j) is the sum of the weights in the directed graph for edges (i, j) and (j, i). Note that A is no longer a linear function of L, which is what makes analyzing the PoA of directed graphs more challenging. Recall that the matrices B and C are defined as follows:

$$B = I - (A + I)^{-1}$$
$$C = (L + I)^{-T}(A + L^T L)(L + I)^{-1}.$$

The price of anarchy, therefore, is $\dfrac{s^T Cs}{s^T Bs}$ as before. The primary distinction between the price of anarchy in the directed and undirected cases is that in the undirected case, B and C are both rational functions of A. In the directed case, no such simple relation exists between B and C, so that we cannot easily bound the generalized eigenvalues for the pair (and hence the price of anarchy) for arbitrary graphs. However, given a directed graph our main theorem shows that we can always find the vector of internal opinions s yielding the maximum PoA:

Theorem 2.4.2 *Given a graph G it is possible to find the internal opinions vector s yielding the maximum PoA up to a precision of ϵ in polynomial time.*

Proof: The total social cost is invariant under constant shifts in opinion. Therefore, without loss of generality, we restrict our attention to the space of opinion vectors with mean zero. Let us define a matrix $P \in \mathbb{R}^{n\times(n-1)}$ to have $P_{j,j} = 1$, $P_{j+1,j} = -1$, and $P_{i,j} = 0$ otherwise. The columns of P are a basis for the space of vectors with mean zero; that is, we can write any such vector as $s = P\hat{s}$ for some $\hat{s}$. We also define matrices $\bar{B} = P^T BP$ and $\bar{C} = P^T CP$, which are positive definite if the symmetrized graph is connected. The price of anarchy for internal opinion vector s is $\dfrac{\hat{s}^T\bar{C}\hat{s}}{\hat{s}^T\bar{B}\hat{s}}$ which is also known as the generalized Rayleigh quotient $\rho_{\bar{C},\bar{B}}(\hat{s})$. To compute the maximum value that the PoA can take we observe that the stationary points of $\rho_{\bar{C},\bar{B}}(\cdot)$ satisfy the generalized eigenvalue equation $(\bar{C} - \rho_{\bar{C},\bar{B}}(\hat{s})\bar{B})\hat{s} = 0$. In particular, the maximal price of anarchy is the largest generalized eigenvalue, and the associated eigenvector $\hat{s}_*$ corresponds to the maximizing choice of internal opinions.

The solution of generalized eigenvalue problems is a standard technique in numerical linear algebra, and there are good algorithms that run in polynomial time; see [50, §8.7]. In particular, because $\bar{B}$ is symmetric and positive definite, we can use the Cholesky factorization $\bar{B} = R^T R$ to reduce the problem to the standard eigenvalue problem $(R^{-T}\bar{C}R^{-1} - \lambda I)(R\hat{s}) = 0$. $\square$

2.4.2 Upper Bounds for Classes of Graphs

Our goal in this section is rather simple: we would like to find families of graphs for which we can bound the price of anarchy. The main tool we use is bounding the cost of the Nash equilibrium by a function of a simple structure. By using a function that has a similar structure to the social cost function we are able to frame the bound as a generalized eigenvalue problem that can be solved using techniques similar to the ones that were used in proving Theorem 2.3.1.

Proposition 2.4.3 *Let $\mathcal{G}$ be a graph family for which there exists a β such that for any $G \in \mathcal{G}$ and any internal opinion vector s, we have $c(x) \leq \min_z \tilde{c}(z)$, where $\tilde{c}(z) = \beta(z^T Az) + ||z - s||^2$. Then, $\forall G \in \mathcal{G}$ and internal opinion vector s,* $\mathrm{PoA}(G) \leq \frac{\beta+\beta\lambda_2}{1+\beta\lambda_2}$, *where λ_2 is the second smallest eigenvalue of A.*

Proof: Let $\tilde{y} = (\beta A + I)^{-1} s$ be the vector minimizing $\tilde{c}(\cdot)$. We can derive the following bound on the price of anarchy:

$$PoA(G) = \frac{c(x)}{c(y)} \leq \frac{\tilde{c}(\tilde{y})}{c(y)} = \frac{s^T \tilde{C} s}{s^T B s},$$

where $\tilde{C}$ and B are defined similarly to the matrices in Theorem 2.3.1:

$$B = I - (A + I)^{-1}$$
$$\tilde{C} = I - (\beta A + I)^{-1}$$

and are simultaneously diagonalizable. If λ_i is an eigenvalue of A then $\lambda_i^B = \frac{\lambda_i}{1+\lambda_i}$ and $\lambda_i^{\tilde{C}} = \frac{\beta\lambda_i}{1+\beta\lambda_i}$. As before, the maximum PoA is achieved when $\lambda_i^{\tilde{C}}/\lambda_i^B = \frac{\beta\lambda_i}{1+\beta\lambda_i}/\frac{\lambda_i}{1+\lambda_i} = \frac{\beta\lambda_i+\beta}{\beta\lambda_i+1}$ is maximized. The maximum here is taken over all eigenvalues different than 0 as we know that the PoA for the internal opinions vector associated with eigenvalue 0 (which is a constant vector) is 1. Therefore, for a connected graph the maximizing eigenvalue is λ_2. □

An immediate corollary is that if there exists a β as in Proposition 2.4.3 then the PoA is bounded by this β.

We say that an unweighted bounded degree *antisymmetric* expander is an unweighted directed graph that does not contain any pair of oppositely oriented edges (i, j) and (j, i), and whose symmetrized graph has maximum degree Δ and edge expansion α. We show:

Claim 2.4.4 *Let $\mathcal{G}_e \subseteq \mathcal{G}$ be a graph family consisting of unweighted bounded degree antisymmetric expanders for which the β defined in Proposition 2.4.3 exists. The PoA of $G \in \mathcal{G}_e$ is bounded by $O(\Delta^2/\alpha^2)$.*

Proof: For an antisymmetric graph, the matrix A is simply the Laplacian of the underlying graph; this is why we require in this claim that the graph is antisymmetric.

If Δ is the maximum degree, then we have $\lambda_2 \leq \lambda_n \leq \Delta$. We also have that $\lambda_2 \geq \alpha^2/2\Delta$ [30]. We can now use this to bound the expression we got in Proposition 2.4.3 for the PoA in terms of the graph's expansion as follows:

$$\frac{\beta + \beta\lambda_2}{1 + \beta\lambda_2} \leq \frac{\beta + \beta\lambda_2}{\beta\lambda_2} \leq \frac{1 + \lambda_2}{\lambda_2} \leq \frac{2\Delta(1 + \Delta)}{\alpha^2} = O(\Delta^2/\alpha^2).$$

□

The next natural question is for which graph families such a β exists. Intuitively, such a β exists whenever the cost of the Nash equilibrium is smaller than the cost of the best consensus — that is, the optimal solution restricted to opinion vectors in which all players hold the same opinion (constant vectors). This is true since the function $\beta(z^T Az) + ||z - s||^2$ is the social cost function of a network in which the weights of all edges have been multiplied by β. However using this intuition for finding graph families for which β exists is difficult and furthermore does not help in computing the value of β (or a bound on it). Hence, we take a different approach. In Lemma 2.4.5, we introduce an intermediate function $g(\cdot)$ with the special property that its minimum value is the same as the cost of the Nash equilibrium. By showing that there exists a β such that $g(z) \leq \beta z^T A z + ||z - s||^2$ we are able to present bounds for Eulerian bounded-degree graphs and additional bounds for Eulerian bounded-degree antisymmetric expanders. As a first step, we prove the following:

Lemma 2.4.5 *For Eulerian graphs, the social cost at Nash equilibrium is* $c(x) = \min_z g(z)$*, where* $g(z) = z^T M z + \|z - s\|^2$ *and* $M = A + LL^T$.

Proof: In the Eulerian case we have $A = L + L^T$, and we use this to simplify the expression

$$C = (L+I)^{-T}(A + L^T L)(L+I)^{-1}.$$

We first note that

$$A + L^T L \quad = \quad L + L^T + L^T L \quad = \quad (L+I)^T (L+I) - I,$$

then substitute to find

$$\begin{aligned} C &= (L+I)^{-T}\left[(L+I)^T(L+I) - I\right](L+I)^{-1} \\ &= I - (L+I)^{-T}(L+I)^{-1} \\ &= I - \left[(L+I)(L+I)^T\right]^{-1} \\ &= I - (M+I)^{-1}. \end{aligned}$$

Where the last transition was based in the fact that

$$(L+I)(L+I)^T - I \quad = \quad L + L^T + LL^T \quad = \quad A + LL^T = M$$

Because M is positive semidefinite, $g(z)$ has a unique minimizer $\hat{y} = (M+I)^{-1}s$; proceeding as in the derivation for the cost at optimality, we find the minimum value achieved is

$$g(\hat{y}) = \hat{y}^T(M+I)\hat{y} - 2s^T\hat{y} + s^Ts = s^T\left[I - (M+I)^{-1}\right]s = s^TCs = c(x).$$

$\square$

Recall that Δ is the maximum degree of an Eulerian graph. We are now ready to prove the following proposition:

Proposition 2.4.6 *For unweighted bounded degree Eulerian graphs* $c(x) \leq \min_z(\Delta+1)(z^TAz) + ||z-s||^2$.

Proof: By Lemma 2.4.5 we have that for Eulerian graphs $c(x) = \min_z g(z) = \min_z z^T(A+LL^T)z + ||z-s||^2$. What remains to show is that for $\beta = \Delta+1$ it holds that $g(z) \leq \beta z^TAz + ||z-s||^2$. After some rearranging this boils down to showing that the following holds: $z^TLL^Tz \leq (\beta-1)z^TAz$.

Note that A is the Laplacian for a symmetrized version of the graph; assuming this graph is connected (since otherwise we can work separately in each component), this means A has one zero eigenvalue corresponding to the constant vectors, and is positive definite on the space orthogonal to the constant vector. Similarly, LL^T has a zero eigenvalue corresponding to the constant vectors, and is at least positive semi-definite on the space orthogonal to the constant vectors. Since A is positive definite on the space of non-constant vectors, the smallest possible β can be computed via the solution of a generalized eigenvalue problem

$$\beta = 1 + \max_{z \neq \alpha e} \frac{z^TLL^Tz}{z^TAz},$$

where e denotes the all-ones vector. In the case of an unweighted graph, one get a bound via norm inequalities. Using the fact that the graph is Eulerian, L^T is also a graph Laplacian, and we can write

$$\left(L^T z\right)_i = \sum_{j=1}^{n} w_{j,i}(z_i - z_j),$$

so

$$z^T L L^T z = \sum_{i=1}^{n} \left(\sum_{j=1}^{n} w_{j,i}(z_i - z_j) \right)^2 .$$

Similarly, we expand the quadratic form $z^T A z$ into

$$z^T A z = \sum_{i<j} (w_{i,j} + w_{j,i})(z_i - z_j)^2 = \sum_{i=1}^{n} \left(\sum_{j=1}^{n} w_{j,i}(z_i - z_j)^2 \right) .$$

Now, recall that in general $\left(\sum_{j=1}^{d} x_j\right)^2 \leq d \sum_{j=1}^{d} x_j^2$, which means that in the unweighted case $\left(\sum_{j=1}^{n} w_{j,i}(z_i - z_j)\right)^2 \leq d_i \left(\sum_{j=1}^{n} w_{j,i}(z_i - z_j)^2\right)$. where $d_i = \sum_j w_{j,i}$ is the in-degree of node i (which is the same as the out-degree). Therefore,

$$\frac{z^T L L^T z}{z^T A z} \leq \frac{\sum_{i=1}^{n} d_i \sum_{j=1}^{n} w_{j,i}(z_i - z_j)^2}{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{j,i}(z_i - z_j)^2} \leq \max_i d_i = \Delta.$$

So for a general Eulerian graph, $\beta \leq 1 + \Delta$.

□

We observe that for a cycle the bound of 2 on the price of anarchy is actually tight:

Observation 2.4.7 *The PoA of a directed cycle is bounded by* 2 *and approaches* 2 *as the size of the cycle grows.*

Proof: For a cycle it is the case that $A = LL^T$; therefore $g(z) = 2(z^T A z) + ||z - s||^2$, and hence the bound assumed in Proposition 2.4.3 is actually a tight bound. In order to show that the PoA indeed approaches 2 we need to show that λ_2 approaches 0 as the size of the cycle grows. The fact that A is the Laplacian of an undirected cycle comes to our aid and provide us an exact formula for λ_2: $\lambda_2 = 2(1 - \cos(\frac{2\pi}{n}))$, where n is the size of the cycle ([30]), and this concludes the proof. $\square$

For general Eulerian graphs we leave open the question of whether the bound of $\Delta + 1$ is a tight bound or not. Indeed, it is an intriguing open question whether there exists a Eulerian graph with PoA greater than 2.

2.5 Adding Edges to the Graph

The next thing we consider is the following class of problems: Given an unweighted graph G and a vector of internal opinions s, find edges E' to add to G so as to minimize the social cost of the Nash equilibrium. We begin with a general bound linking the possible improvement from adding edges to the price of anarchy. Let G be a graph (either undirected or directed). Denote by $c_G(\cdot)$ the cost function and by x and y the Nash equilibrium and optimal solution respectively. Let G' be the graph constructed by adding edges to G. Then: $\frac{c_G(x)}{c_{G'}(x')} \leq \frac{c_G(x)}{c_{G'}(y')} \leq \frac{c_G(x)}{c_G(y)} = \text{PoA}(G)$. To see why this is the case, we first note that $c_{G'}(y') \leq c_{G'}(x')$ since the cost of the Nash equilibrium cannot be smaller than the optimal solution. Second, $c_G(y) \leq c_{G'}(y')$ simply because $c_{G'}(\cdot)$ contains more terms than $c_G(\cdot)$. Therefore we have proved the following proposition:

Proposition 2.5.1 *Adding edges to a graph G can improve the cost of the Nash equilibrium by a multiplicative factor of at most the PoA of G.*

We study three variants on the problem, discussed in the introduction. In all variants, we seek the "best" edges to add in order to minimize the social cost of the Nash equilibrium. The variants differ mainly in the types of edges we may add.

Adding edges from a specific node

First, we consider the case in which we can only add edges from a specific node w. Here we imagine that node w is a media source that therefore does not have any cost for holding an opinion, and so we will use a cost function that ignores the cost associated with it when computing the social cost. Hence, our goal is to find a set of nodes F such that adding edges from node w to all the nodes in F minimizes the cost of the Nash equilibrium while ignoring the cost exhibited by w. By reducing the subset sum problem to this problem we show that:

Proposition 2.5.2 *Finding the best set of edges to add from a specific node w is NP-hard.*

Proof: Denote by $G + F$ the graph constructed by adding to G edges from w to all nodes in F. Our goal is to find a set F minimizing $\tilde{c}_{G+F}(x)$, where x is a Nash equilibrium in the graph $G + F$ and $\tilde{c}$ denotes the total cost of all nodes in x except for node w. We show that finding this set is NP-hard by reducing the subset sum problem to this problem. Recall that in the subset sum problem we are given a set of positive integers $a_1, \ldots, a_n$ and a number t. We would like to know if there exists any subset S such that $\sum_{j \in S} a_j = t$. Given an instance of the subset problem, we reduce it to the following instance of the opinion game. The instance contains an in-directed star with n peripheral nodes that have an internal opinion of 0 and a center node w which has an internal opinion of 1 and n isolated nodes that have internal opinions of $-\frac{a_i}{t}$. This construction is illustrated in Figure 2.3.

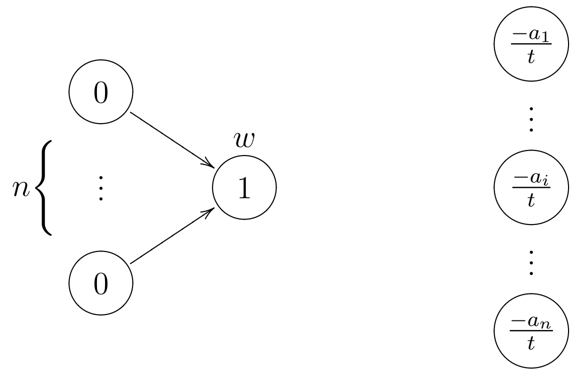


Figure 2.3: An illustration of the instance structured for Proposition 2.5.2.

Lemma 2.5.3 *For the graph G and the vector of internal opinions s defined above, there exists a set F such that $\tilde{c}_{G+F}(x) = 0$ if and only if the answer to the subset problem is yes.*

Proof: As seen in the introduction, in the Nash equilibrium each one of the peripheral nodes holds an opinion of $\frac{1}{2}x_w$. Node w hold an opinion of $x_w = \dfrac{1 + \sum_{j \in F} s_j}{1 + |F|}$. As we assume that w does not incur any cost, the cost of the Nash equilibrium in $G + F$ is just the cost of the n peripheral nodes:

$$\tilde{c}_{G+F}(x) = n\left((\frac{1}{2}x_w - 0)^2 + (x_w - \frac{1}{2}x_w)^2\right) = 2n\left(\frac{1 + \sum_{j \in F} s_j}{2(1 + |F|)}\right)^2.$$

Clearly the cost is nonnegative as it is a sum of quadratic terms; moreover it equals 0 if and only if $\sum_{j \in F} s_j = -1$. Defining $F' = \{j \in F | s_j < 0\}$, we have $\sum_{j \in F'} s_j = -1$. By the reduction we have that $\sum_{j \in F'} -\frac{a_j}{t} = -1$; if we multiply by $-t$ we get that $\sum_{j \in F'} a_j = t$ implying that there exists a solution to the subset sum problem. □ □

Adding edges to a specific node

Next, we consider the case in which we can only add edges to a specific node. We can imagine again that node w is a media source; in this case, however, our goal is to find the best set of people to expose to this media source. By reducing the minimum vertex cover problem to this problem we show that:

Proposition 2.5.4 *Finding the best set of edges to add to a specific node w is NP-hard.*

Proof: Given an instance of the minimum vertex cover problem, consisting of an undirected graph $G' = (V', E')$, we construct an instance of the opinions game as follows:

- For each edge $(i, j) \in E'$ we create a vertex $v_{i,j}$ with internal opinion 1.
- For every $v_{i,j}$ we create an in-directed star with 24 peripheral nodes that have an internal opinion of 0. We later refer to node $v_{i,j}$ and all the nodes directed to it as *$v_{i,j}$'s star*.
- For each vertex $i \in V'$ we create a vertex u_i with internal opinion 1.
- For each edge $(i, j) \in E'$ we create directed edges $(v_{i,j}, u_i)$ and $(v_{i,j}, u_j)$.
- We create an isolated node w with internal opinion -3.

We illustrate some of this construction in Figure 2.4.

Let T be the set of vertices such that adding edges from all the nodes in T to node w minimizes the cost of the Nash equilibrium. Denote by $G + T$ the graph constructed by adding to G edges from all nodes in T to w. Consider some node

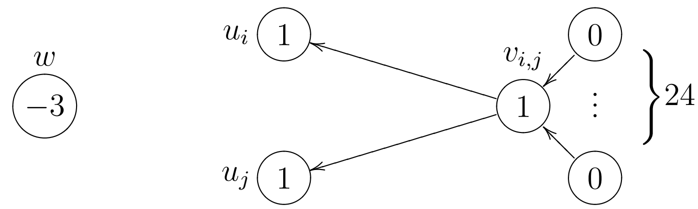


Figure 2.4: A partial illustration of the construction in Proposition 2.5.4.

b with internal opinion of 0. Observe that since nodes with internal opinion of 0 do not have any incoming edges, any edge that would be added from b to w would only effect b's cost. Since adding an edge from b to w amounts to adding a positive term to b's cost function, it cannot be the case that this improves b's cost and thus the total social cost. Thus, T contains only vertices of type $v_{i,j}$ and u_i. In the table in Figure 2.5 we compute $v_{i,j}$'s opinion in the Nash equilibrium and the cost of its star as a function of which vertices that influence $v_{i,j}$ are in T. For example in the first row we consider the case in which $v_{i,j}, u_i, u_j \notin T$. In this case, $v_{i,j}$'s opinion is $(1+1+1)/3 = 1$ and the cost of its star is $\frac{1}{2} \cdot 24 = 12$. We use the costs in this table to reason about the structure of T and the cost of the Nash equilibrium in $G+T$.

	Configuration	$v_{i,j}$'s opinion	$v_{i,j}$'s star cost
1	$v_{i,j}, u_i, u_j \notin T$	1	12
2	$v_{i,j} \in T, u_i, u_j \notin T$	0	12
3	$v_{i,j}, u_i \in T, u_j \notin T$	$-1/2$	14
4	$v_{i,j}, u_i, u_j \in T$	-1	20
5	$v_{i,j}, u_j \notin T, u_i \in T$	$1/3$	4
6	$v_{i,j} \notin T, u_i, u_j \in T$	$-1/3$	4

Figure 2.5: The total cost of $v_{i,j}$'s star for different configurations

In Lemma 2.5.5 we show how to construct from T a set T' such that $c_{G+T'}(x') = c_{G+T}(x)$ and T' is a *pseudo vertex cover*. We say that a set T' is a pseudo vertex cover if it obeys two properties: (i) it contains only vertices of the type u_i. (ii) the vertices in V' corresponding to the u_i's in V constitute a vertex cover in G'.

Next, we consider the cost of the Nash equilibrium in the graph $G + S$ where S is a pseudo vertex cover: By the table in Figure 2.5 we have the cost associated with every $v_{i,j}$'s star is 4. This is by the fact that S is a pseudo vertex cover and hence the only applicable cases are 5 and 6, in both cases the total cost of $v_{i,j}$'s star is 4. Also, note that the cost for each $u_i \in S$ is 8. Hence, the total cost of the Nash equilibrium for network $G+S$ is $f(S) = 4|E|+8|S|$. By construction, T' is a pseudo vertex cover and it also minimizes $f(\cdot)$, since $c_{G+T}(x) = c_{G+T'}(x') = 4|E| + 8|T'|$ and T is optimal. Therefore T' corresponds to a minimum vertex cover in G'. A key element in this reduction is the property that the cost of $v_{i,j}$'s star is the same, whether $u_i \in T'$ or both u_i and u_j belong to T'.

Lemma 2.5.5 *There exists a pseudo vertex cover T' such that $c_{G+T'}(x') = c_{G+T}(x)$*

Proof: First, we obtain T'' by removing from T all vertices of type $v_{i,j}$. We have that $c_{G+T''}(x'') \leq c_{G+T}(x)$ since by examining the table in Figure 2.5 we observe that including vertices of type $v_{i,j}$ in T'' can only increase the cost of the Nash equilibrium. Since T is optimal, it has to be the case that $c_{G+T''}(x'') = c_{G+T}(x)$. Next, to get T' we take T'' and for each vertex $v_{i,j}$ such that $u_i, u_j \notin T''$ we add u_i to T'. By adding these vertices we have not increased the cost since in the worst case $v_{i,j}$'s star and u_i have a total cost of 12 which is the same as their previous total cost. As before by the optimality of T we could not have reduced the cost by adding the vertices, therefore it still holds that $c_{G+T'}(x') = c_{G+T}(x)$. To complete the proof observe that by construction T' is a pseudo vertex cover. □

□

Adding an arbitrary set of edges

In the last case we consider, which is the most general one, we can add any set of edges. For this case we leave open the question of the hardness of adding an unrestricted set of edges. We do show that finding the best set of k arbitrary edges is NP-hard. This is done by a reduction from k-dense subgraph [39] :

Proposition 2.5.6 *Finding a best set of arbitrary k edges is NP-hard.*

Proof: We show a reduction from the "Dense k-Subgraph Problem" defined in [39]: given an undirected graph $G' = (V', E')$ and a parameter k, find a set of k vertices with maximum average degree in the subgraph induced by this set. Given an instance of the "Dense k-Subgraph Problem" we create an instance of the opinion game as follows: (illustrated in Figure 2.6)

- For every edge $(i, j) \in E'$ we create a node $v_{i,j}$ with internal opinion 0.
- For every vertex $i \in V'$ we create a node u_i with internal opinion 1.
- For every $v_{i,j}$ we add directed edges $(v_{i,j}, u_i)$ and $(v_{i,j}, u_j)$.
- For every u_i we create an in-directed star with $2n^3$ peripheral nodes that have an internal opinion of 0. We later refer to node u_i and all the nodes directed to it as *u_i's star*.
- Finally, we create a single isolated vertex w with internal opinion -1.

The proof is composed of two lemmas. In Lemma 2.5.7 we show that all edges in the minimizing set are of type (u_i, w). Then we denote by T the set of nodes

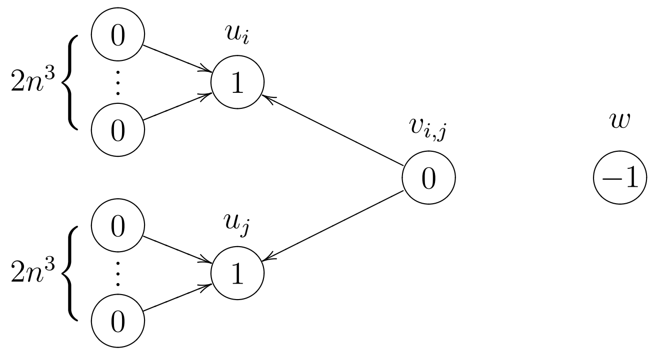


Figure 2.6: A partial illustration of the construction in Proposition 2.5.6.

of type u_i such that adding an edge from each one of these nodes to w minimizes the cost, and in Lemma 2.5.8 we show that T is a k densest subgraph.

Lemma 2.5.7 *The best set of edges to add contains only edges from nodes of type u_i to w.*

Proof: Observe that by construction connecting a node of type u_i to w reduces the cost of u_i's star from n^3 to less than 2. Thus, the reduction in the cost for choosing k nodes of type u_i and connecting them to w is at least $k(n^3 - 2)$. Now, assume towards a contradiction that the set T of best edges includes some edges of type different than (u_i, w). Consider the set of nodes that their cost was effected as the result of the addition of these edges. Let $t < k$ be the number of stars that include at least a single effected node. The total cost of all nodes in this effected set is at most $n^3t + n^2$. Since, the cost of a complete star is n^3 and there are at most n^2 possible effected nodes of type $v_{i,j}$, each has a cost of $\frac{2}{3} < 1$. This is an upper bound on the reduction in cost that adding the edges in T can achieve. Observe that $k(n^3 - 2) > n^3t + n^2$ since $(k - t)n^3 > n^2 + 2k$ as $n^3 > n^2 + 2k$ for any interesting value of n (recall that $k < n$). Thus we have that the origin of each edge in T has to be a different node in u_i's star for some i.

For every star that does not have any incoming edges it is easy to observe that the best edge to add from this star is an edge between u_i to w, as this edge reduces the cost by $n^3 - 2$. Therefore, either from each star there exists an edge of type (u_i, w), or there exists a cycle, connecting several stars to one another. Also, observe that this cycle can only include nodes of type u_i since they are the only ones that influence other nodes. By including a peripheral node the only change in cost would be of this specific node's cost and not of the star it belongs to. It remains to rule out the existence of a cycle consisting of some of the u_i's. To see why this cannot be the case, note that in equilibrium all the nodes of type u_i express the exact same opinion. Thus, connecting them in a cycle would not change their opinion and in turn would not reduce the cost associated with their stars at all.

□

Lemma 2.5.8 *The previously defined set T is a solution to the dense k-subgraph problem.*

Proof: The key point is the fact that the cost associated with a node of type $v_{i,j}$ is 0 if and only if both u_i and u_j are in T; otherwise this cost is exactly $\frac{2}{3}$. When $u_i \in T$, the opinion of u_i in the Nash equilibrium is 0 since it is averaging between 1 and -1. Therefore node $v_{i,j}$'s associated cost in the Nash equilibrium is:

- 0 - if both u_i and u_j are in T - since $v_{i,j}$ holds opinion 0.
- $\frac{2}{3}$ - if both u_i and u_j are not in T - since $v_{i,j}$'s opinion is $\frac{2}{3}$ and therefore the cost is $(0 - \frac{2}{3})^2 + 2(1 - \frac{2}{3})^2 = \frac{2}{3}$.

- $\frac{2}{3}$ - if only one of u_i, u_j is in T - then $v_{i,j}$'s opinion is $\frac{1}{3}$ and therefore the cost is $(0-\frac{1}{3})^2+(0-\frac{1}{3})^2+(1-\frac{1}{3})^2=\frac{2}{3}$.

Hence to minimize the cost of the Nash equilibrium we should choose a set T maximizing the number of nodes of type $v_{i,j}$ for which both u_i and u_j are in T. In the graph G' from the k-dense subgraph problem that set T is a set of vertices and what we are looking for is the set T with an induced graph that has the maximum number of edges. By definition this set is exactly a k-densest subgraph. □ □

Finding approximation algorithms for all of the problems discussed in propositions 2.5.2, 2.5.4, and 2.5.6 is an interesting question. As a first step we offer a $\frac{9}{4}$-approximation for the problem of optimally adding edges to a directed graph G — a problem whose hardness for exact optimization we do not know. The approximation algorithm works simply by including the reverse copy of every edge in G that is not already in G; this produces a bi-directed graph G'.

Claim 2.5.9 $c_{G'}(x') \le \frac{9}{4} c_G(y)$.

Proof: By Theorem 2.3.1 we have that $c_{G'}(x') \le \frac{9}{8} c_{G'}(y')$. Also notice that in the worst case, in order to get from G to G', we must double all the edges in G. Therefore $c_{G'}(y') \le 2c_G(y)$. By combining the two we have that $c_{G'}(x') \le \frac{9}{4} c_G(y)$. □

Observe that for undirected graph, we are restricted to include only the reverse copy of every edge in G that *is not* already in G. For some instances this prevents us from adding edges to construct the graph which A is the Laplacian of and therefore the Nash equilibrium of the new graph is the optimal solution of the original one. Once, we deal with weighted graph the restriction no longer holds and therefore by include reverse copies of all edges that do appear in G we can achieve an approximation ratio of 2.

CHAPTER 3

DISCRETE PREFERENCES AND COORDINATION

3.1 Introduction

In the previous chapter we analyzed a model of opinion formation in which each person has an internal opinion and has to choose a potentially different opinion to express with the goal of minimizing her cost. This is just one example for settings in which the outcome does not depend only on personal choices, but also on the choices of the people they interact with. A natural model for such situations is to consider a game played on a graph that represents an underlying social network, where the nodes are the players. Each node's personal decision corresponds to selecting a strategy, and the node's payoff depends on the strategies chosen by itself and its neighbors in the graph [18, 38, 73].

Coordination and Internal Preferences

A fundamental class of such games involves payoffs based on the interplay between *coordination* — each player has an incentive to match the strategies of his or her neighbors — and *internal preferences* — each player also has an intrinsic preference for certain strategies over others, independent of the desire to match what others are doing. Trade-offs of this type come up in a very broad collection of situations, and it is worth mentioning several that motivate our work here.

- In the context of opinion formation, as was discussed in the previous chapter, a group of people or organizations might each possess different internal views, but they are willing to express or endorse a "compromise" opinion so as to be in closer alignment with their network neighbors.

- Questions involving technological compatibility among firms tend to have this trade-off as a fundamental component: firms seek to coordinate on shared standards despite having internal cost structures that favor different solutions.
- Related to the previous example, a similar issue comes up in cooperative facility location problems, where firms have preferences for where to locate, but each firm also wants to locate near the firms with which it interacts.

In the previous chapter we discussed the line of work beginning in the mathematical social sciences that is concerned with opinion formation — where the possible strategies correspond to a continuous space such as $\Re^d$ [42, 59]. This makes it possible for players to adopt arbitrarily fine-grained "average" strategies from among any set of options, and most of the dynamics and equilibrium properties of such models are driven by this type of averaging. In particular, dynamics based on repeated averaging have been shown in early work to exhibit nice convergence properties [42]. In the previous chapter we contributed to this line of work by developing bounds on the relationship between equilibria and social optima.

Discrete Preferences

In many settings that exhibit a tension between coordination and individual preferences, however, there is no natural way to average among the available options. Instead, the alternatives are drawn from a fixed discrete set — for example, there is only a given set of available technologies for firms to choose among, or a fixed set of political candidates to endorse or vote for. On a much longer time scale, there is always the possibility that additional options could be created to interpolate between what's available, but on the time scale over which the strategic interaction takes place, the players must choose from among the discrete set of alternatives that is available.

Among a small fixed set of players, coordination with discrete preferences is at the heart of a long line of games in the economic theory literature — perhaps the most primitive example is the classic *Battle of the Sexes* game, based on a pedagogical story in which one member of a couple wants to see movie A while the other wants to see movie B, but both want to go to a movie together. This provides a very concrete illustration of a set of payoffs in which the (two) players have (i) conflicting internal preferences (A and B respectively), (ii) an incentive to arrive at a compromise, and (iii) no way to "average" between the available options.

But essentially nothing is known about the properties of the games that arise when we consider such a payoff structure in a network context. Even the direct generalization of Battle of the Sexes (BoS) to a graph is more or less unexplored in this sense — each node plays a copy of BoS on each of its incident edges, choosing a single strategy A or B for use in all copies, incurring a cost from miscoordination with neighbors and an additional fixed cost when the node's choice differs from its inherent preference. Indeed, as some evidence of the complexity of even this formulation, note that the version in which each node has an intrinsic preference for A is equivalent to the standard network coordination game, which already exhibits rich graph-theoretic structure [73]. And beyond this, of course, lies the prospect of such games with larger and more involved strategy sets.

Formalizing Discrete Preference Games

In this chapter, we develop a set of techniques for analyzing this type of discrete preference games on a network, and we establish tight bounds on the price of stability for several important families of such games.

To formulate a general model for this type of game, we start with an undirected graph $G = (V, E)$ representing the network on the players, and an underlying finite set L of strategies. Each player $i \in V$ has a *preferred strategy* $s_i \in L$, which is what i would choose in the absence of any other players. Finally, there is a metric $d(\cdot, \cdot)$ on the strategy set L — that is, a distance function satisfying (i) $d(i, i) = 0$ for all i, (ii) $d(i, j) = d(j, i)$ for all i, j, and (iii) $d(i, j) \leq d(i, k) + d(k, j)$ for all i, j and k. For $i, j \in L$, the distance $d(i, j)$ intuitively measures how "different" i and j are as choices; players want to avoid choosing strategies that are at large distance from either their own internal preference or from the strategies chosen by their neighbors.

Each player's objective is to minimize her cost (think of this as the negative of her payoff): for a fixed parameter $\alpha \in [0, 1]$, the cost to player i when players choose the strategy vector $z = \langle z_j : j \in V \rangle$ is

$$c_i(z) = \alpha \cdot d(s_i, z_i) + \sum_{j \in N(i)} (1 - \alpha) \cdot d(z_i, z_j),$$

where $N(i)$ is the set of neighbors of i in G. The parameter α essentially controls the extent to which players are more concerned with their preferred strategies or their network neighbors; we will see that the behavior of the game can undergo qualitative changes as we vary α.

We say that the above formulation defines a *discrete preference game*. Note that the network version of Battle of the Sexes described earlier is essentially the special case in which $|L| = 2$, and network coordination games are the special case in which $|L| = 2$ and $\alpha = 0$, since then players are only concerned with matching their neighbors. The case in which $d(\cdot, \cdot)$ is the distance metric among nodes on a path is also interesting to focus on, since it is the discrete analogue of the one-dimensional space of real-valued opinions from continuous averaging models [17, 42] — consider for example the natural scenario in which a finite number of discrete alternatives in an election are arranged along a one-dimensional political spectrum.

We also note that discrete preference games belong to the well-known framework of graphical games, which essentially consist of games in which the utility of every player depends only on the actions of its neighbors in a network. The interested reader is referred to the relevant chapter in [79] and the references within. In this context, Gottlob et al. proposed a generalization of Battle of the Sexes (BoS) to a graphical setting [51], but their formulation was much more complex than our starting point, with their questions correspondingly focused on existence and computational complexity, rather than on the types of performance guarantees we will be seeking.

For any discrete preference game, we will see that it is possible to define an exact potential function, and hence these games possess pure Nash equilibria.

Price of Stability in Discrete Preference Games

We can also ask about the *social cost* of a strategy vector $z = \langle z_j : j \in V \rangle$, defined as the sum of all players' costs:

$$c(z) = \sum_{i \in V} \alpha \cdot d(s_i, z_i) + 2 \sum_{(i,j) \in E} (1 - \alpha) \cdot d(z_i, z_j).$$

We note that the problem of minimizing the social cost is an instance of the *metric labeling problem*, in which we want to assign labels to nodes in order to minimize a sum of per-node costs and edge separation costs [23, 65].

Since an underlying motivation for studying this class of games is the tension between preferred strategies and agreement on edges, it is natural to study its consequences on the social cost via the price of anarchy and/or the price of stability. The price of anarchy is in fact too severe a measure for this class of games; indeed, as we discuss in the next section, it is already unbounded for the well-studied class of network coordination games that our model contains as a special case.

We therefore consider the price of stability, which turns out to impart a rich structure to the problem. The price of stability is also natural in terms of the underlying examples discussed earlier as motivation; in most of these settings, it makes sense to propose a solution — for example, a compromise option in a political setting or a proposed set of technology choices for a set of interacting firms — and then to see if it is stable with respect to equilibrium.

Overview of Results

As a starting point for reference, observe that network coordination games (where players are not concerned with their preferred strategies) clearly have a price of stability of 1: the players can all choose the same strategy and achieve a cost of 0. But even for a general discrete preference game with two strategies — i.e. Battle of the Sexes on a network — the price of stability is already more subtle, since the social optimum may have a more complex structure (as a two-label metric labeling problem, and hence a minimum cut problem).

We begin by giving tight bounds on the maximum possible price of stability in the two-strategy case as a function of the parameter α. The dependence on α has a complex non-monontonic character; in particular, the price of stability is equal to 1 for all instances if and only if $\alpha \leq 1/2$ or $\alpha = 2/3$, and more generally the price of stability as a function of α displays a type of "saw-tooth" behavior with infinitely many local minima in the interval $[0, 1]$. Our analysis uses a careful scheduling of the best-response dynamics so as to track the updates of players toward a solution with low social cost.

Above we also mentioned the distance metric of a path as a case of interest in opinion formation. We show that when $\alpha \leq 1/2$, the price of stability for instances based on such metrics is always 1, by proving the stronger statement that in fact the price of stability is always 1 for any discrete preference game based on a tree metric. Our analysis for tree metrics involves considering how players' best responses lie at the medians of their neighbors' strategies in the metric, and then developing combinatorial techniques for reasoning about the arrangement of these collections of medians on the underlying tree.

Like path metrics, tree metrics are also relevant to motivating scenarios in terms of opinion formation, when individuals classify the space of possible opinions according to a hierarchical structure rather than a linear one. To take one example of this, consider students choosing a major in college, where each student has an internal preference and an interest in picking a major that is similar to the choices of her friends. The different subjects roughly follow a hierarchy — on top we might have science, engineering, and humanities; under science we can have for example biology, physics, and other areas; and under biology we can have subjects including genetics and plant breeding. This setting fits our model since each person has some internal inclination for a major, but still it is arguably the case that a math major has more in common in her educational experience with her computer science friends than with her friends in comparative literature.

The two families of instances described above (two strategies and tree metrics) both have price of stability equal to 1 when $\alpha \leq 1/2$. But the price of stability can be greater than 1 for more general metrics when $\alpha \leq 1/2$. It is not hard to show (as we do in the next section) that the price of stability is always at most 2 for all α, and we match this bound by constructing and analyzing examples, based on perturbations of uniform metrics, showing that the price of stability can be arbitrarily close to 2 when $\alpha = 1/2$.

3.2 Preliminaries

Recall that in a discrete preference game played on a graph $G = (V, E)$ with strategy set L, each player $i \in V$ has a preferred strategy $s_i \in L$. The cost incurred by player i when all players choose strategies $z = \langle z_j : j \in V \rangle$ is

$$c_i(z) = \alpha \cdot d(s_i, z_i) + \sum_{j \in N(i)} (1 - \alpha) \cdot d(z_i, z_j).$$

The social cost of the game is the sum of all the players' costs:

$$c(z) = \sum_{i \in V} \alpha \cdot d(s_i, z_i) + 2 \sum_{(i,j) \in E} (1 - \alpha) \cdot d(z_i, z_j).$$

Another quantity that is useful to define is the contribution of player i to the social cost – by this we quantify not only the cost player i is exhibiting but also the cost it is inflicting on its neighbors:

$$sc_i(z) = \alpha \cdot d(s_i, z_i) + 2 \sum_{j \in N(i)} (1 - \alpha) \cdot d(z_i, z_j).$$

As is standard, we denote by z_{-i} the strategy vector z without the i^{th} coordinate.

We first show that this class of games includes instances for which the price of anarchy (PoA) is unbounded. A simple instance for which the PoA is unbounded is one in which the preferred strategy of all the players is the same – thus the cost of the optimal solution is 0, and it is also an equilibrium for all the players to play some other strategy and incur a positive cost.[1] In the next claim, we use this idea to construct, for every value of $\alpha < 1$, an instance for which the PoA is unbounded.

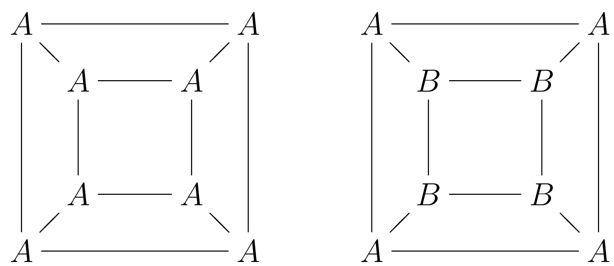
Claim 3.2.1 *For any $\alpha < 1$ there exists an instance for which the price of anarchy is unbounded.*

Proof: Assume the strategy space contains two strategies A and B, such that $d(A, B) = 1$. For any $0 < \alpha < 1$ we consider a clique of size $\lceil \frac{\alpha}{1-\alpha} \rceil + 1$ in which all players' preferred strategy is A and show it is an equilibrium for all the players to play strategy B. To see why, observe that if the rest of the players play strategy B, then player i's cost for playing strategy A is $(1 - \alpha) \cdot \lceil \frac{\alpha}{1-\alpha} \rceil$ which is at least α. Since the cost of player i for playing strategy B is α we have that it is an equilibrium for all players to play strategy B. The PoA of such an instance is unbounded as the cost of the equilibrium in which all players play strategy B is strictly positive but the cost of the optimal solution is 0.

To show that the PoA can be unbounded for $\alpha = 0$, a slightly different instance is required, which will be familiar from the literature of network coordination

[1] This type of equilibrium, in addition to simply producing an unbounded PoA, has a natural interpretation in our motivating contexts. In technology adoption, it corresponds to convergence on a standard that no firm individually wants, but which is hard to move away from once it has become the consensus. In opinion formation, it corresponds essentially to a kind of "superstitious" belief that is universally expressed, and hence is hard for people to outwardly disavow even though they prefer an alternate opinion.

games. When players do not have a preference the optimal solution is clearly for all players to play the same strategy, as such a solution has a cost of 0. However, Figure 3.1 depicts an instance for which there exists a Nash equilibrium in which not all the players play the same strategy and hence the cost of this equilibrium is strictly positive. □



(a) *An optimal solution* (b) *A Nash equilibrium*

Figure 3.1: An instance illustrating that the PoA can be unbounded even when the players do not have a preferred strategy (i.e., $\alpha = 0$).

We note that the worst equilibrium in the previous instances is not a strong equilibrium. Thus, if all the players could coordinate a joint deviation to strategy A they can all benefit. A natural question is what happens if we restrict ourselves to the worst strong Nash equilibrium (strong PoA), in which a simultaneous deviation by a set of players is allowed. Unfortunately, the strong PoA can still be quite high (linear in the number of players). Take for example $\alpha < \frac{1}{2}$ and consider a clique of size n in which all but one of the players prefer strategy A. In this case it is not hard to verify that the equilibrium in which all players play strategy B is strong. The reason is that the player that prefers strategy B cannot gain from any deviation, and the cost of any other player would increase by at least $1-\alpha-\alpha > 0$ for any deviation. The cost of such an equilibrium is $\alpha(n-1)$ in comparison to the optimal solution which has a cost of α.

As we just showed both the PoA and the strong PoA can be very high, and hence for the remainder of the chapter we focus on the qualities of the best Nash equilibrium, trying to bound the price of stability (PoS). We begin by showing that the price of stability is bounded by 2. This is done by a potential function argument which also proves that a Nash equilibrium always exists.

Claim 3.2.2 *The price of stability is bounded by* 2.

Proof: We first prove that the following function is an exact potential function:

$$\phi(z) = \alpha \sum_{i \in V} d(z_i, s_i) + (1-\alpha) \sum_{(i,j) \in E} d(z_i, z_j).$$

To see why, note that: $\phi(z_i, z_{-i}) - \phi(z'_i, z_{-i}) =$

$$\alpha \cdot d(z_i, s_i) + (1-\alpha) \sum_{j \in N(i)} d(z_i, z_j) - \Big(\alpha \cdot d(z'_i, s_i) + (1-\alpha) \sum_{j \in N(i)} d(z'_i, z_j)\Big)$$

$$= c_i(z_i, z_{-i}) - c_i(z'_i, z_{-i}).$$

Denote by x the global minimizer of the potential function and by y the optimal solution. By definition x is an equilibrium and it provides a 2-approximation to the optimal social cost since $c(x) \leq 2\phi(x) \leq 2\phi(y) \leq 2c(y)$. $\square$

3.3 The Case of Two Strategies: Battle of the Sexes on a Network

We begin by considering the subclass of instances in which the players only have two different strategies A and B. Without loss of generality we assume that $d(A, B) = 1$. We denote by $N_j(i)$ the set of i's neighbors using strategy j and by $\bar{s}_i$ the strategy opposite to s_i. When the strategy space contains only two strategies, a player's best response is to pick a strategy which is the weighted majority of its own preferred strategy and the strategies played by its neighbors. The next two observations formalize this statement and a similar statement regarding a player's strategy minimizing the social cost:

Observation 3.3.1 *The strategy s_i minimizes player i's cost ($c_i(z)$) if:*

$$(1-\alpha)N_{\bar{s}_i}(i) \leq \alpha + (1-\alpha)N_{s_i}(i) \implies N_{\bar{s}_i}(i) \leq \frac{\alpha}{1-\alpha} + N_{s_i}(i).$$

Observation 3.3.2 *The strategy s_i minimizes the social cost ($sc_i(z)$) if:*

$$2(1-\alpha)N_{\bar{s}_i}(i) \leq \alpha + 2(1-\alpha)N_{s_i}(i) \implies N_{\bar{s}_i}(i) \leq \frac{\alpha}{2(1-\alpha)} + N_{s_i}(i).$$

We present a simple best response order that results in a Nash equilibrium after a linear number of best responses. We will later see how this order can be used to bound the PoS.

Lemma 3.3.3 *Starting from some initial strategy vector, the following best response order results in a Nash equilibrium:*

1. *While there exists a player that can reduce its cost by changing its strategy to A, let it do a best response. If there is no such player continue to the second step.*
2. *While there exists a player that can reduce its cost by changing its strategy to B, let it do a best response.*

Proof: To see why the resulting strategy vector is a Nash equilibrium, observe that after the first step, all nodes are either satisfied with their current strategy choice, or can benefit from changing their strategy to B. This property remains true after some of the nodes change their strategy to B since the fact that a node has more neighbors using strategy B can only reduce the attractiveness of switching to strategy A. Thus, at the end of the second step all nodes are satisfied. $\square$

Next, we characterize the values of α for which the price of stability is 1.

Claim 3.3.4 *If $\alpha \leq \frac{1}{2}$ or $\alpha = \frac{2}{3}$, then in any instance there exists an optimal solution which is also a Nash equilibrium.*

Proof: Let y be an optimal solution minimizing the potential function $\phi(\cdot)$. Assume towards a contradiction that it is not a Nash equilibrium. Let player i be a player that prefers to switch to a different strategy than y_i. Denote player i's best response by x_i. By Observations 3.3.1 and 3.3.2, it is easy to see that if $y_i = s_i$, then the strategy minimizing player i's cost is also s_i. Thus, we have that $y_i \neq s_i$ and $x_i = s_i$. If s_i is a minimizer of the social cost function then (s_i, y_{-i}) is also an optimal solution. This contradicts the assumption that y is a minimizer of the potential function $\phi(\cdot)$ since $\phi(s_i, y_{-i}) < \phi(y)$. Therefore by Observations 3.3.1 and 3.3.2 we have that $N_{\bar{s}_i}(i) < \frac{\alpha}{1-\alpha} + N_{s_i}(i)$ and $\frac{\alpha}{2(1-\alpha)} + N_{s_i}(i) < N_{\bar{s}_i}(i)$. By combining the two inequalities we get that:

$$\frac{\alpha}{2(1-\alpha)} + N_{s_i}(i) < N_{\bar{s}_i}(i) < \frac{\alpha}{1-\alpha} + N_{s_i}(i).$$

Since $N_{\bar{s}_i}(i)$ and $N_{s_i}(i)$ are both integers, this implies that there exists an integer k such that $\frac{\alpha}{2(1-\alpha)} < k < \frac{\alpha}{1-\alpha}$. This holds for $k = 1$ if $\frac{\alpha}{2(1-\alpha)} < 1 < \frac{\alpha}{1-\alpha}$ (implying $\frac{1}{2} < \alpha < \frac{2}{3}$) or for some $k > 1$ if $\frac{\alpha}{1-\alpha} - \frac{\alpha}{2(1-\alpha)} > 1$ (implying $\alpha > \frac{2}{3}$). Thus, for $\alpha \leq \frac{1}{2}$ or $\alpha = \frac{2}{3}$ any instance admits an optimal solution which is also a Nash equilibrium. $\square$

It is natural to ask what is the PoS for the values of α for which we know the optimal solution is not a Nash equilibrium. The following theorem provides an answer to this question by computing the ratio between the optimal solution and a Nash equilibrium obtained by performing the sequence of best responses Lemma 3.3.3 prescribes.

Theorem 3.3.5 *For* $\frac{1}{2} < \alpha < 1$, *PoS* $\leq 2\left\lceil \dfrac{\alpha}{1-\alpha} - 1 \right\rceil \cdot \dfrac{1-\alpha}{\alpha}$.

Proof: Let x be the equilibrium achieved by the sequence described in Lemma 3.3.3 starting from an optimal solution y. Denote the strategy vector at the end of the first step by x_1 and at the end of the sequence by $x_2 = x$. We assume that a player performs a best response only when it can strictly decrease the cost by doing so, thus we only reason about the case where the player's best response is unique. In the following Lemma we bound the increase in the social cost inflicted by the players' unique best responses (the proof can be found below):

Lemma 3.3.6 *Let player i's* unique *best response when the rest of the players play* z_{-i} *be* x_i *then:*

1. *If* $x_i = \bar{s}_i$ *then* $c(\bar{s}_i, z_{-i}) - c(s_i, z_{-i}) \leq \alpha - 2(1-\alpha)\left\lfloor \dfrac{\alpha}{1-\alpha} + 1 \right\rfloor$.
2. *If* $x_i = s_i$ *then* $c(s_i, z_{-i}) - c(\bar{s}_i, z_{-i}) \leq -\alpha + 2(1-\alpha)\left\lceil \dfrac{\alpha}{1-\alpha} - 1 \right\rceil$.

Notice that by statement (1) of Lemma 3.3.6 a node that changes its strategy to a strategy different than its preferred strategy can only reduce the social cost. Also, note that if a node changes its strategy in the first step to A and in the second step back to B its total contribution to the social cost is non-positive. The reason is that in one of these changes the player changed its strategy from s_i to $\bar{s}_i$ and in the other from $\bar{s}_i$ to s_i. The effect of these two changes on the social cost sums up to $2(1-\alpha)\left(\lceil\frac{\alpha}{1-\alpha}\rceil - \lfloor\frac{\alpha}{1-\alpha}\rfloor - 2\right) \leq 0$. Thus we can ignore such changes as well.

The only nodes that are capable of increasing the social cost by performing a best response are ones that play in the optimal solution a different strategy than their preferred strategy ($y_i \neq s_i$). By definition their number equals exactly $\sum_i d(y_i, s_i)$ as $d(y_i, s_i) = 1$ if $y_i \neq s_i$ and 0 otherwise. Statement (2) of Lemma 3.3.6 guarantees us that each of these nodes can increase the social cost by at most $-\alpha + 2(1-\alpha)\left\lceil\frac{\alpha}{1-\alpha} - 1\right\rceil$. Thus, we get the following bound:

$$\begin{aligned} c(x) &\leq c(y) + \left(-\alpha + 2(1-\alpha)\left\lceil\frac{\alpha}{1-\alpha} - 1\right\rceil\right)\sum_{i\in V} d(y_i, s_i) \\ &= 2(1-\alpha)\sum_{(i,j)\in E} d(y_i, y_j) + 2(1-\alpha)\left\lceil\frac{\alpha}{1-\alpha} - 1\right\rceil\sum_{i\in V} d(y_i, s_i). \end{aligned}$$

We are now ready to compute the bound on the PoS:

$$\begin{aligned} PoS &\leq \frac{2\left\lceil\frac{\alpha}{1-\alpha} - 1\right\rceil \cdot (1-\alpha)\sum_{i\in V} d(y_i, s_i) + 2(1-\alpha)\sum_{(i,j)\in E} d(y_i, y_j)}{\alpha\sum_{i\in V} d(y_i, s_i) + 2(1-\alpha)\sum_{(i,j)\in E} d(y_i, y_j)} \\ &\leq \frac{2\left\lceil\frac{\alpha}{1-\alpha} - 1\right\rceil \cdot \frac{1-\alpha}{\alpha} \cdot \left(\alpha\sum_{i\in V} d(y_i, s_i) + 2(1-\alpha)\sum_{(i,j)\in E} d(y_i, y_j)\right)}{\alpha\sum_{i\in V} d(y_i, s_i) + 2(1-\alpha)\sum_{(i,j)\in E} d(y_i, y_j)} \\ &\leq 2\left\lceil\frac{\alpha}{1-\alpha} - 1\right\rceil \cdot \frac{1-\alpha}{\alpha}. \end{aligned}$$

□

Proof of Lemma 3.3.6: Notice we are only considering the effect player i's strategy has on the social cost, thus we have that: $c(\bar{s}_i, z_{-i}) - c(s_i, z_{-i}) = sc_i(\bar{s}_i, z_{-i}) - sc_i(s_i, z_{-i})$. Implying that $c(\bar{s}_i, z_{-i}) - c(s_i, z_{-i}) = \alpha + 2(1-\alpha)(N_{s_i}(i) - N_{\bar{s}_i}(i))$. For proving statement (1) we observe that since $\bar{s}_i$ is player i's unique best response then: $N_{\bar{s}_i}(i) > \frac{\alpha}{1-\alpha} + N_{s_i}(i)$ as player i has a strictly smaller cost for playing strategy $\bar{s}_i$ than for playing strategy s_i. Since, $N_{\bar{s}_i}(i)$ and $N_{s_i}(i)$ are integers this implies that $N_{\bar{s}_i}(i) - N_{s_i}(i) \geq \lfloor \frac{\alpha}{1-\alpha} + 1 \rfloor$ and the bound is achieved.

For proving statement (2), observe that $c(s_i, z_{-i}) - c(\bar{s}_i, z_{-i}) = -\alpha + 2(1 - \alpha)(N_{\bar{s}_i}(i) - N_{s_i}(i))$. Now since s_i is player i's best response we have that: $N_{\bar{s}_i}(i) < \frac{\alpha}{1-\alpha} + N_{s_i}(i)$ as player i's best response is to use strategy s_i. Which similarly to the previous bound implies that $N_{\bar{s}_i}(i) - N_{s_i}(i) \leq \lceil \frac{\alpha}{1-\alpha} - 1 \rceil$ as required. □

It is interesting to take a closer look at the upper bound on the PoS we computed (as we will see later this bound is tight). In Figure 3.2 we plot the upper bound on the PoS as a function of α. We can see that as α approaches 1 the PoS approaches 2 and also that for any $k \geq 2$, as ϵ approaches 0, the PoS of $\alpha = \frac{k-1}{k} + \epsilon$ also approaches to 2. This uncharacteristic saw-like behavior of the PoS originates from the fact that for every value of α the maximal PoS is achieved by a star graph. This is proved in the following claim.

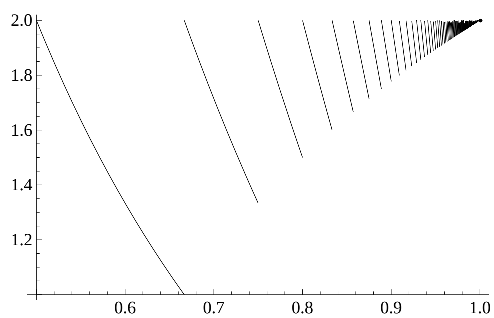


Figure 3.2: The tight upper bound on the PoS for two strategies as a function of α for the range $\frac{1}{2} < \alpha < 1$.

Claim 3.3.7 *For any $\alpha > 1/2$, $\alpha \neq 2/3$ there exists an instance achieving a price of stability of* $2\left\lceil \frac{\alpha}{1-\alpha} - 1 \right\rceil \cdot \frac{1-\alpha}{\alpha}$.

Proof: Consider a star consisting of $\left\lceil \frac{\alpha}{1-\alpha} - 1 \right\rceil$ peripheral nodes that prefer strategy A and a central node that prefers strategy B. In the optimal solution the central node plays strategy A for a cost of α. However, this is not a Nash equilibrium since for playing strategy B it exhibits a cost of $(1-\alpha)\cdot\left\lceil \frac{\alpha}{1-\alpha} - 1 \right\rceil < \alpha$. Thus, the central node prefers to play its preferred strategy. $\square$

Corollary 3.3.8 *As n goes to infinity the PoS for $\alpha = \frac{n-1}{n}$ approaches* 2.

3.4 Richer Strategy Spaces

In the previous section we have seen that even when there are only two strategies in the game (the Battle of the Sexes on a network), for at least some values of $\alpha > \frac{1}{2}$, the PoS can be quite close to 2.

These bounds carry over to larger strategy spaces since an instance can always use only two strategies from the strategy space. However, for $\alpha \leq \frac{1}{2}$ the PoS for the Battle of the Sexes on a network is 1, so a natural question is how bad the PoS can be once we have more strategies in the space. This is the question we deal with for the rest of the chapter.

3.4.1 Tree Metrics

We begin by considering the case in which the distance function on the strategy set is a tree metric, defined as the shortest-path metric among the nodes in a tree. (As such, tree metrics are a special case of graphic metrics, in which there is a graph on the elements of the space and the distance between every two elements is defined to be the length of the shortest path between them in the graph.) We show that if the distance function is a tree metric then the price of stability is 1 for any rational $\alpha \leq \frac{1}{2}$.

Denote by $C_i(z)$ and $SC_i(z)$ the strategies of player i that minimize $c_i(z) = \alpha \cdot d(z_i, s_i) + (1-\alpha)\sum_{j \in N(i)} d(z_i, z_j)$ and $sc_i(z) = \alpha \cdot d(z_i, s_i) + 2(1-\alpha)\sum_{j \in N(i)} d(z_i, z_j)$ respectively. We show that if for every player i the intersection of the two sets $C_i(z)$ and $SC_i(z)$ is always non-empty then the price of stability is 1:

Claim 3.4.1 *If for every player i and strategy vector z, $SC_i(z) \cap C_i(z) \neq \emptyset$, then $PoS = 1$.*

Proof: First, recall the potential function $\phi(\cdot)$ used in the proof of Claim 3.2.2, and consider an optimal solution y minimizing this potential function $\phi(\cdot)$. If y is also a Nash equilibrium then we are done. Else, there exists a node i that can strictly reduce its cost by performing a best response. By our assumption node i can do this by choosing a strategy $x_i \in SC_i(y) \cap C_i(y)$. The fact that $x_i \in SC_i(y)$ implies that the change in strategy of player i does not affect the social cost. Therefore, (x_i, y_{-i}) is also an optimal solution and $\phi(y) > \phi(x_i, y_{-i})$, in contradiction to the assumption that y is the optimal solution minimizing $\phi(\cdot)$. $\square$

Our goal now is to show that the conditions of Claim 3.4.1 hold for a tree metric. Our first step is relating the strategies that consist of a player's best response (or social cost minimizer) and the set of medians of a node-weighted tree.

Definition 3.4.2 (medians of a tree) *Given a tree T where the weight of node v is denoted $w(v)$, the set of T's medians is $M(T) = \arg\min_{u \in V}\{\sum_{v \in V} w(v) \cdot$*

$d(u,v)\}$.

Definition 3.4.3 *Given a network G, a tree metric T, a strategy vector z, a player i and non-negative integers q and r, we denote by $T_{i,z}(q,r)$ the tree T with the following node weights:*

$$w(v) = \begin{cases} q + r \cdot |\{j \in N(i) | z_j = v\}| & \text{for } v = s_i \\ r \cdot |\{j \in N(i) | z_j = v\}| & \text{for } v \neq s_i \end{cases}$$

Next we show that for $\alpha = \frac{a}{a+b}$, every player i and strategy vector z, it holds that $M(T_{i,z}(a,b)) = C_i(z)$. To see why, observe that by construction we have $M(T_{i,z}(a,b)) =$

$$\arg\min_{u \in V} \{(a + b \cdot |\{j \in N(i) | z_j = s_i\}|) \cdot d(u, s_i) + \sum_{v \neq s_i \in V} b \cdot |\{j \in N(i) | z_j = v\}| \cdot d(u,v)\}$$

$$= \arg\min_{u \in V} \{a \cdot d(u, s_i) + b \sum_{j \in N(i)} d(u, z_j)\} = C_i(z).$$

Similarly, it is easy to show that $M(T_{i,z}(a, 2b)) = SC_i(z)$. Thus, to show that $SC_i(z) \cap C_i(z) \neq \emptyset$ it is sufficient to show that $T_{i,z}(a,b)$ and $T_{i,z}(a,2b)$ share a median. This is done by using the following proposition:

Proposition 3.4.4 *Let T_1 and T_2 be two node-weighted trees with the same edges and nodes, then:*

- *If there exists a node v, such that for every $u \neq v \in V$, we have $w_1(u) = w_2(u)$ and for v we have $|w_1(v) - w_2(v)| = 1$, then T_1 and T_2 share a median.*
- *If T_1 and T_2 share a median, then it is also a median of their union $T_1 \cup T_2$. Where the union of T_1 and T_2 is a tree with the same nodes and edges where the weight of node v is $w_{1+2}(v) = w_1(v) + w_2(v)$.*

The proof of this proposition is based on combinatorial claims from Section 3.4.2 showing that a tree's medians and separators (defined below) coincide and establishing connections between the separators of different trees. Given Proposition 3.4.4 we can now show that $T_{i,z}(a,b)$ and $T_{i,z}(a,2b)$ share a median:

Lemma 3.4.5 *For $\alpha = \frac{a}{a+b} \leq \frac{1}{2}$, every player i and strategy vector z, $M(T_{i,z}(a,b)) \cap M(T_{i,z}(a,2b)) \neq \emptyset$.*

Proof: First, observe that by Proposition 3.4.4 we have that $T_{i,z}(0,1)$ and $T_{i,z}(1,1)$ share a median. As medians are invariant to scaling this implies that $T_{i,z}(0,b-a)$ and $T_{i,z}(a,a)$ also share a median. Next, by the second statement of Proposition 3.4.4 we have that any median they share is also a median of their union $T_{i,z}(a,b)$; let us denote this median by u. Since u is a median of $T_{i,z}(0,b)$ and $T_{i,z}(a,b)$, it is also a median of $T_i(a,2b)$ by applying Proposition 3.4.4 again. Thus we have that u is a median of both $T_{i,z}(a,b)$ and $T_{i,z}(a,2b)$ which concludes the proof. $\square$

Hence we have proven the following theorem:

Theorem 3.4.6 *If the distance metric is a tree metric then for rational $\alpha \leq \frac{1}{2}$, there exists an optimal solution which is also a Nash equilibrium (PoS=1).*

3.4.2 Combinatorial Properties of Medians and Separators in Trees

We now state and prove the combinatorial facts about medians and separators in trees that we used for the analysis above. This builds on the highly tractable structure of medians in trees developed in early work; see [44, 47] and the references therein.

Consider a tree where all nodes have integer weights, and denote the weight of the tree by $w(V) = \sum_{v \in V} w(v)$.

Definition 3.4.7 *A separator of a tree T is a node v such that the weight of each connected component of $T - v$ is at most $w(V)/2$.*

Claim 3.4.8 *A node u is a median of a tree T if and only if it is a separator of T.*

Proof: Let u be a median of a tree T, and assume towards a contradiction that it is not a separator; that is, there exists a component of $T - u$ of weight strictly greater than $w(V)/2$. Let v be the neighbor of u in this component. Consider locating the median at v. This reduces the distance to a total node weight of at least $w(V)/2 + 1/2$ by 1, and increases the distance to less than a total node weight of $w(V)/2$ by 1. Hence the sum of all distances decreases, and this contradicts the fact that u is a median. Thus, every median of a tree is also a separator.

To show that any separator is also a median, we show that for any two separators u_1 and u_2 it holds that $\sum_{v \in V} w(v) \cdot d(u_1, v) = \sum_{v \in V} w(v) \cdot d(u_2, v)$. Since, we know that there exists a median which is a separator this will imply that any separator is a median.

Denote by C the connected component of the graph $T - u_1 - u_2$ that includes the nodes on the path between u_1 and u_2. If u_1 and u_2 are adjacent let $C = \emptyset$. Denote the connected component of $T - u_1 - C$ that includes u_2 by C_2 and the connected component of $T - u_2 - C$ that includes u_1 by C_1. Note that by construction

C, C_1, C_2 are disjoint and the union of their nodes equals V. Since u_1 is a separator it holds that $w(C_2)+w(C) \le w(V)/2$. This in turn implies that $w(C_1) \ge w(V)/2$. Similarly, since u_2 is a separator it holds that $w(C_1) + w(C) \le w(V)/2$. This in turn implies that $w(C_2) \ge w(V)/2$. Therefore, it has to be the case that $w(C) = 0$, $w(C_1) = w(C_2) = w(V)/2$. We next show this implies that $\sum_{v \in V} w(v) \cdot d(u_1, v) = \sum_{v \in V} w(v) \cdot d(u_2, v)$. Observe that:

$$\begin{aligned}\sum_{v \in V} w(v) \cdot d(u_1, v) &= \sum_{v \in C_1} w(v) \cdot d(u_1, v) + \sum_{v \in C_2} w(v) \cdot \big(d(u_1, u_2) + d(u_2, v)\big) \\ &= \sum_{v \in C_1} w(v) \cdot d(u_1, v) + \sum_{v \in C_2} w(v) \cdot d(u_2, v) + w(C_2) \cdot d(u_1, u_2).\end{aligned}$$

and similarly that:

$$\sum_{v \in V} w(v) \cdot d(u_2, v) = \sum_{v \in C_2} w(v) \cdot d(u_2, v) + \sum_{v \in C_1} w(v) \cdot d(u_1, v) + w(C_1) \cdot d(u_1, u_2).$$

The claim follows as we have shown that $w(C_2) = w(C_1)$. $\square$

Next, we prove two claims relating the separators of different trees. We first show that if two trees differ only in the weight of a single node and the difference in weight of this node in the two trees is 1 – then they share a separator:

Claim 3.4.9 *Consider two trees T_1 and T_2 with the same set of edges and nodes that differ only in the weight of a single node v - such that $w_2(v) = w_1(v) + 1$. Then, T_1 and T_2 share a separator.*

Proof: We first handle the case where $w_1(V)$ is odd. Let u be a separator of T_1, then, in this case the size of each component in $T_1 - u$ is at most $w_1(V)/2 - 1/2$. Thus for the same separator u in T_2 the size of each component is at most $w_1(V) + 1/2 = w_2(V)/2$. Therefore u is still a separator. Assume that $w_1(V)$ is even. This implies that $w_1(V) + 1$ is odd. Consider a separator u' of T_2. Then the size of each connected component in $T_2 - u'$ is at most $w_2(V)/2 + 1/2$, since $w_1(V)$ is even, this implies that the weight of each connected component is bounded by $w_1(V)/2$ and therefore u' is also a separator of T_1. $\square$

We show that if u is a separator of both T_1 and T_2 it is also the separator of their union:

Claim 3.4.10 *Every separator that T_1 and T_2 share is also a separator of $T_1 \cup T_2$.*

Proof: Let u be a separator of both T_1 and T_2. This implies that in T_1 the weight of every connected component in $T_1 - u$ is at most $w_1(T_1)/2$ and in T_2 the weight of every connected component in $T_2 - u$ is at most $w_2(T_2)/2$. Hence, in $T_1 \cup T_2$ the weight of every connected component in $T_1 \cup T_2 - u$ is at most $w_1(T_1)/2 + w_2(T_2)/2 = w_{1+2}(T_1 \cup T_2)/2$. Thus, u is also a separator of $T_1 \cup T_2$. $\square$

The previous three claims establish the proof of Proposition 3.4.4. First Claim 3.4.8 shows that the set of medians and separators coincide. Then, Claim 3.4.8 Claim 3.4.10 prove the two statements of the proposition respectively.

3.4.3 Lower Bounds in Non-Tree Metrics

In some sense tree metrics are the largest class of metrics for which the optimal solution is always a Nash equilibrium for $\alpha \leq \frac{1}{2}$. The next example demonstrates that even when the distance metric is a simple cycle the PoS can be as high as $\frac{4}{3}$ for $\alpha = \frac{1}{2}$. In the following section, we give a family of more involved constructions that converge to the asymptotically tight lower bound of 2 on the price of stability.

Example 3.4.11 *Consider a metric which is a cycle of size $3k+1$ for some integer*

$k \geq 1$. Let A,B,C be three strategies in this strategy space such that $d(A, B) = k$, $d(A, C) = k$ and $d(B, C) = k + 1$. Consider an instance where a node with a preferred strategy A is connected to a node with preferred strategy B and to another node with a preferred strategy of C. Also, assume that the node with preferred strategy B is part of a clique of size $3k$ in which all nodes prefer strategy B. Similarly, the node with preferred strategy C is part of a clique of size $3k$ in which all nodes prefer strategy C.

Consider the following equilibrium in which the nodes in both cliques play their preferred strategies. Then, the central node should play its preferred strategy. To see why, note that for playing strategy A its cost is $(1/2)2k = k$. On the other hand, its cost for playing any other strategy x which is between A and B (including B) on the cycle is $\frac{1}{2}(d(x, A) + d(x, B) + d(x, C))$ which equals:

$$\frac{1}{2}\Big(d(A, B) + min\{d(x, A) + d(A, C), d(x, B) + d(B, C)\}\Big).$$

which is greater than k. Similarly one can show that the central player prefers strategy A over any strategy x. The cost of the Nash equilibrium is $2k$. Note that this is the best Nash equilibrium since the cost of any solution in which some of the nodes in a clique play a strategy different than their preferred strategy is at least $3k$. In the optimal solution the central node should play strategy B (or C) for a total cost of $(1/2)k + 2 \cdot (1/2)(k + 1) = (3/2)k + 1$. Thus we have that the price of stability approaches $4/3$ as k approaches infinity.

Note that this lower bound of $4/3$ is achieved on an instance in which the lowest-cost Nash equilibrium and the socially optimal solution differ only in the strategy choice of a single player. We now show that in such cases, where the difference between these two solutions consists of the decision of just a single player, $4/3$ is the maximum possible price of stability for $\alpha = \frac{1}{2}$. More generally we show that $\dfrac{2}{2-\alpha}$ is the maximum possible price of stability for $\alpha < \frac{1}{2}$.

By the definition of the model, a player's strategy only affects its cost and the cost of its neighbors. Recall that we denote this part of the social cost by $sc_i(z)$: $sc_i(z) = \alpha \cdot d(s_i, z_i) + 2(1-\alpha) \cdot \sum_{j \in N(i)} d(z_i, z_j)$. We now prove the following claim.

Claim 3.4.12 *Let $\alpha \leq \frac{1}{2}$. Fix an optimal solution y which is not a Nash equilibrium and let player i be a player that can reduce its cost by playing x_i. Then* $\dfrac{sc_i(x_i, y_{-i})}{sc_i(y)} < \dfrac{2}{2-\alpha}$.

Proof: Since x_i is player i's best response, then $\alpha \cdot d(s_i, x_i) + (1 - \alpha) \sum_{j \in N(i)} d(x_i, y_j) < c_i(y)$. By rearranging the terms we get that $(1 - \alpha) \sum_{j \in N(i)} d(x_i, y_j) < c_i(y) - \alpha \cdot d(s_i, x_i)$. This in turn implies that

$$\begin{aligned} sc_i(x_i, y_{-i}) &= \alpha \cdot d(s_i, x_i) + 2(1-\alpha) \sum_{j \in N(i)} d(x_i, y_j) \\ &< \alpha \cdot d(s_i, x_i) + 2\big(c_i(y) - \alpha \cdot d(s_i, x_i)\big) \\ &= 2c_i(y) - \alpha \cdot d(s_i, x_i). \end{aligned}$$

Thus, we have that

$$\frac{sc_i(x_i, y_{-i})}{sc_i(y)} < \frac{2c_i(y) - \alpha \cdot d(s_i, x_i)}{c_i(y) + (1-\alpha) \cdot \sum_{j \in N(i)} d(y_i, y_j)}.$$

If $\sum_{j \in N(i)} d(y_i, y_j) \geq c_i(y)$ then $sc_i(y) \geq c_i(y) + (1-\alpha)c_i(y) = (2-\alpha)c_i(y)$ and the claim follows. Else, we show that $d(s_i, x_i) > d(s_i, y_i) - \sum_{j \in N(i)} d(y_i, y_j)$ in Lemma 3.4.13 below; this in turn implies that

$$\begin{aligned} sc_i(x_i, y_{-i}) &< 2c_i(y) - \alpha \cdot d(s_i, x_i) \leq 2c_i(y) - \alpha\big(d(s_i, y_i) - \sum_{j \in N(i)} d(y_i, y_j)\big) \\ &= c_i(y) + \sum_{j \in N(i)} d(y_i, y_j). \end{aligned}$$

This brings us to the following bound:

$$\frac{sc_i(x_i, y_{-i})}{sc_i(y)} < \frac{c_i(y) + \sum_{j\in N(i)} d(y_i, y_j)}{c_i(y) + (1-\alpha)\cdot \sum_{j\in N(i)} d(y_i, y_j)}$$
$$= 1 + \frac{\alpha \sum_{j\in N(i)} d(y_i, y_j)}{c_i(y) + (1-\alpha)\cdot \sum_{j\in N(i)} d(y_i, y_j)}.$$

Recall that by our assumption $c_i(y) > \sum_{j\in N(i)} d(y_i, y_j)$, this implies that $c_i(y) + (1-\alpha)\cdot \sum_{j\in N(i)} d(y_i, y_j) > (2-\alpha)\sum_{j\in N(i)} d(y_i, y_j)$ and the claim follows.

$\square$

Lemma 3.4.13 *Let $\alpha \leq \frac{1}{2}$. Fix an optimal solution y which is not a Nash equilibrium and let player i be a player that can reduce its cost by playing x_i. Then: $d(s_i, x_i) > d(s_i, y_i) - \sum_{j\in N(i)} d(y_i, y_j)$.*

Proof: Note the following: by the triangle inequality for any player j it holds that: $d(x_i, y_j) \geq d(s_i, y_j) - d(s_i, x_i)$ and $d(s_i, y_j) \geq d(s_i, y_i) - d(y_i, y_j)$. By combining the two together we have that $d(x_i, y_j) \geq d(s_i, y_i) - d(y_i, y_j) - d(s_i, x_i)$. This gives us the following lower bound on $c_i(x_i, y_{-i})$:

$$\begin{aligned}
c_i(x_i, y_{-i}) &= \alpha \cdot d(s_i, x_i) + (1-\alpha) \sum_{j\in N(i)} d(x_i, y_j) \\
&\geq \alpha \cdot d(s_i, x_i) + (1-\alpha) \sum_{j\in N(i)} \Big(d(s_i, y_i) - d(y_i, y_j) - d(s_i, x_i) \Big) \\
&= \alpha \cdot d(s_i, x_i) + (1-\alpha)\cdot |N(i)| \cdot \big(d(s_i, y_i) - d(s_i, x_i) \big) - (1-\alpha) \sum_{j\in N(i)} d(y_i, y_j).
\end{aligned}$$

Since x_i minimizes player i's cost it has to be the case that: $c_i(x_i, y_{-i}) < c_i(y)$. Thus the following inequality holds:

$$\alpha \cdot d(s_i, x_i) + (1-\alpha)|N(i)| \cdot \big(d(s_i, y_i) - d(s_i, x_i) \big) - (1-\alpha) \sum_{j\in N(i)} d(y_i, y_j)$$
$$< \alpha \cdot d(s_i, y_i) + (1-\alpha) \sum_{j\in N(i)} d(y_i, y_j).$$

After some rearranging we get that:

$$d(s_i, x_i) > d(s_i, y_i) - \frac{2(1-\alpha)}{(1-\alpha)\cdot|N(i)| - \alpha} \sum_{j\in N(i)} d(y_i, y_j)$$

which implies that the claim holds whenever $\dfrac{2(1-\alpha)}{(1-\alpha)\cdot|N(i)| - \alpha} \leq 1$. For $\alpha \leq \frac{1}{2}$, this later bound occurs for $|N(i)| \geq 3$.

The case of $|N(i)| \leq 2$ is handled separately and requires we use the assumption that y is an optimal solution. Denote i's neighbors by j and k. Then:

$$\alpha\cdot d(s_i, y_j) + 2(1-\alpha)\cdot d(y_j, y_k) \geq \alpha\cdot d(s_i, y_i) + 2(1-\alpha)\big(d(y_i, y_j) + d(y_i, y_k)\big).$$

By the triangle inequality the previous inequality implies that $d(s_i, y_j) \geq d(s_i, y_i)$. When combining this with the fact that $d(x_i, y_j) \geq d(s_i, y_j) - d(s_i, x_i)$ we get that $d(x_i, y_j) \geq d(s_i, y_i) - d(s_i, x_i)$; similarly we get for k that $d(x_i, y_k) \geq d(s_i, y_i) - d(s_i, x_i)$. Therefore,

$$\begin{aligned} c_i(x_i, y_{-i}) &= \alpha\cdot d(s_i, x_i) + (1-\alpha)\big(d(x_i, y_j) + d(x_i, y_k)\big) \\ &\geq \alpha\cdot d(s_i, x_i) + 2(1-\alpha)\big(d(s_i, y_i) - d(s_i, x_i)\big). \end{aligned}$$

and since x_i is player i's best response it has to be the case that:

$$\alpha\cdot d(s_i, x_i) + 2(1-\alpha)\big(d(s_i, y_i) - d(s_i, x_i)\big) < \alpha\cdot d(s_i, y_i) + (1-\alpha)\big(d(y_i, y_j) + d(y_i, y_k)\big).$$

After some rearranging we get that:

$$(2-3\alpha)d(s_i, y_i) - (1-\alpha)\big(d(y_i, y_j) + d(y_i, y_k)\big) < (2-3\alpha)d(s_i, x_i).$$

By dividing both sides of the inequality by $(2-3\alpha)$ we get that $d(s_i, x_i) > d(s_i, y_i) - \sum_{j\in N(i)} d(y_i, y_j)$ holds whenever $\frac{1-\alpha}{2-3\alpha} \leq 1$. This completes the proof since for $\alpha \leq \frac{1}{2}$ it is always the case that $\frac{1-\alpha}{2-3\alpha} \leq 1$. $\square$

3.5 Lower Bounds on the Price of Stability

At the end of the previous section, we saw that even in very simple non-tree metrics, the price of stability can be greater than 1. We now give a set of stronger lower bounds, using a more involved family of constructions. First we give an asymptotically tight lower bound of 2 when $\alpha = \frac{1}{2}$, and then we adapt this construction to give non-trivial lower bounds for all $0 < \alpha < \frac{1}{2}$.

3.5.1 Price of stability for $\alpha = \frac{1}{2}$

The following example illustrates that the PoS for $\alpha = \frac{1}{2}$ can be arbitrarily close to 2. The network we consider is composed of a path of n nodes and two cliques of size n^2 connected to each of the endpoints of the path. We assume that the preferred strategy of node i on the path is s_i, the preferred strategy of all nodes in the leftmost clique is s_0, and the preferred strategy of all nodes in the rightmost clique is s_{n+1}. The following is a sketch of the network:

$$\bigcirc_0 \;\text{———}\; \bullet 1 - \cdots\cdots - \bullet n \;\text{———}\; \bigcirc_{n+1}$$

Since all the s_i's are distinct we use them also as names for the different possible strategies. We define the following distance metric on these strategies: for $i > j$, we have $d(s_i, s_j) = 1 + (i - j - 1)\epsilon$. (When $i < j$, we simply use $d(s_i, s_j) = d(s_j, s_i)$.) In Claim 3.5.1 below we show that the best Nash equilibrium is the one in which all players play their preferred strategies. The cost of this equilibrium is $c(s) = \frac{1}{2} \cdot 2 \sum_{i=0}^{n} d(s_i, s_{i+1}) = n + 1$. On the other hand, consider the assignment in which for some node i all the nodes up till node i choose strategy s_0 and all the nodes from node $i + 1$ choose strategy s_{n+1}. The cost of such assignment is

$\frac{1}{2}(n+2+O(\epsilon))$; therefore as n goes to infinity and ϵ to zero the PoS goes to 2.

Claim 3.5.1 *In the previously defined instance the best Nash equilibrium is for each player to play its preferred strategy.*

Proof: We show that the cost of any other equilibrium is at least $\frac{1}{2}n^2$. Consider an equilibrium in which there exists at least one node i that plays strategy s_j such that $j < i$. By the following lemma (which we prove below) this implies that node i's neighbors play strategies s_a and s_b such that $a, b < i$ or $a, b > i$.

Lemma 3.5.2 *Let s_a and s_b be the strategies played by player i's neighbors such that $a \leq b$. If $a \leq i \leq b$, then player i's best response is to play strategy s_i.*

Observe that for i's best response to be strategy s_j it clearly has to be the case that $a, b < i$. By applying Lemma 3.5.2 repeatedly, we get that in this equilibrium, all nodes $k > i$ play strategies k' such that $k' < k$. This includes the n^{th} node of the path, implying that its right neighbor which belongs to the right clique plays strategy $s_{k'}$ such that $k' < n+1$. The cost incurred by the nodes in the clique in any such equilibrium is at least n^2: indeed, if r nodes play a strategy different than their preferred one, they pay a cost of at least $\frac{1}{2}r$ and the remaining $n^2 - r$ pay a cost of at least $n^2 - r$ for the edges connecting them to one of the r nodes playing a strategy different than its preferred strategy. To complete the proof, one can use an analogous argument for the case in which there exists an equilibrium in which there is a node i playing strategy s_j such that $j > i$. $\square$

Proof of Lemma 3.5.2: We first note that by the definition of the metric it is never in i's best interest to play a strategy s_j such that $j \neq i, a, b$: playing such a strategy has cost $3 + O(\epsilon)$ whereas the cost of playing the preferred strategy is $2 + O(\epsilon)$.

Observe that player i prefers to play strategy s_i over strategy $s_a \neq s_i$ whenever $d(s_i, s_a) + d(s_i, s_b) < d(s_i, s_a) + d(s_a, s_b)$ implying that $d(s_i, s_b) < d(s_a, s_b)$. This conditions holds according to our assumptions since $1 + (b - i - 1)\epsilon = d(s_i, s_b) < d(s_a, s_b) = 1 + (b - a - 1)\epsilon$. For the same reason, player i prefers strategy s_i over $s_b \neq s_i$ since $1 + (i - a - 1) = d(s_i, s_a) < d(s_a, s_b) = 1 + (b - a - 1)$ under the lemma's assumptions. $\square$

3.5.2 Extension for $\alpha < \frac{1}{2}$

We extend the construction in Section 3.5.1 to $0 < \alpha < \frac{1}{2}$ by defining the following metric: for $i > j$, let $d(s_i, s_j) = 1 + (i - j - 1)\left(\frac{1-2\alpha}{1-\alpha}(1+\epsilon)\right)$. We consider the same family of instances defined in Section 3.5.1 except for the fact that we increase the size of the cliques to n^2/α.

Next, we show that Lemma 3.5.2 also holds for this newly defined family of instances. This fact together with the observation that the proof of Claim 3.5.1 carries over with only minor modifications, imply that in the best Nash equilibrium of the previously defined family of instances all players play their preferred strategies.

Lemma 3.5.3 *Let s_a and s_b be the strategies played by player i's neighbors such that $a \leq b$. If $a \leq i \leq b$, then player i's best response is to play strategy s_i.*

Proof: For proving this claim it will be easier to use an equivalent distance function which is: $d(s_i, s_j) = 1 + (|i-j|-1)\left(\frac{1-2\alpha}{1-\alpha}(1+\epsilon)\right)$. Also, we only present the proof for the case that $a < i < b$, the proof for the rest of the cases is very similar. We first show that player i prefers to play strategy s_i over playing any strategy s_j such that $j \neq a, b$. The cost of player i for playing s_i is:

$$\begin{aligned}
&(1-\alpha)\left(2 + (|i-a| - 1 + |b-i| - 1) \cdot \frac{1-2\alpha}{1-\alpha}(1+\epsilon)\right) \\
&= 2(1-\alpha) + (|b-a|-2) \cdot (1-2\alpha)(1+\epsilon) \\
&\leq 1 + (|b-a|-1) \cdot (1-2\alpha)(1+\epsilon).
\end{aligned}$$

The cost of playing strategy s_j such that $j \neq a, b$ is:

$$\begin{aligned}
&\alpha(1 + (|i-j|-1)\frac{1-2\alpha}{1-\alpha}(1+\epsilon)) \\
&\quad + (1-\alpha)\left(2 + (|j-a| - 1 + |j-b| - 1) \cdot \frac{1-2\alpha}{1-\alpha}(1+\epsilon)\right) \\
&\geq 2 - \alpha + (|b-a|-2) \cdot (1-2\alpha)(1+\epsilon).
\end{aligned}$$

The last transition is due to the fact that $|j-a| + |j-b| \geq |b-a|$. Thus we conclude that player i's best response can only be s_i, s_a or s_b. Next we consider strategies s_a and s_b. By writing the cost of playing each one of these strategies it is easy to see that these costs are greater than the costs for playing s_i.

The cost of playing strategy $s_a \neq s_i$ is:

$$\begin{aligned}
&\alpha(1 + (|i-a|-1)\frac{1-2\alpha}{1-\alpha}(1+\epsilon)) + (1-\alpha)\left(1 + (|b-a|-1) \cdot \frac{1-2\alpha}{1-\alpha}(1+\epsilon)\right) \\
&= 1 + (\frac{\alpha}{1-\alpha}(|i-a|-1) + |b-a| - 1) \cdot (1-2\alpha)(1+\epsilon).
\end{aligned}$$

The cost of playing strategy $s_b \neq s_i$ is:

$$\begin{aligned}
&\alpha(1 + (|b-i|-1)\frac{1-2\alpha}{1-\alpha}(1+\epsilon)) + (1-\alpha)\left(1 + (|b-a|-1) \cdot \frac{1-2\alpha}{1-\alpha}(1+\epsilon)\right) \\
&= 1 + (\frac{\alpha}{1-\alpha}(|b-i|-1) + |b-a| - 1) \cdot (1-2\alpha)(1+\epsilon).
\end{aligned}$$

This conclude the proof as we have shown that under the assumptions of the claim, player i's best response it to play its preferred strategy. $\square$

To get a lower bound on the PoS, we would like to simulate the technique we used in the proof for $\alpha = \frac{1}{2}$ to compare between the cost of two solutions: (i) a solution in which there is exactly one edge such that its two endpoints play different strategies, and (ii) the best Nash equilibrium. We refer to the first solution as a *bi-consensus solution*. The cost of the best bi-consensus solution is an upper bound on the optimal solution, and hence computing the ratio between the best Nash equilibrium and best bi-consensus solution gives a lower bound on the PoS achieved by instances defined above.

Observe that in the best bi-consensus solution nodes $i \in [1 \dots \lfloor n/2 \rfloor]$ play strategy s_0 and nodes $i \in [\lceil n/2 \rceil \dots n]$ play strategy s_{n+1}. The cost of this solution b is the following:

$$
\begin{aligned}
c(b) = \alpha\Big(& \sum_{i=1}^{\lfloor n/2 \rfloor} (1 + (i-1)(\frac{1-2\alpha}{1-\alpha}(1+\epsilon))) \; + \sum_{i=\lfloor n/2 \rfloor + 1}^{n} (1 + (n-i)(\frac{1-2\alpha}{1-\alpha}(1+\epsilon)))\Big) \\
& + 2(1-\alpha)(1 + (n+1-0-1)(\frac{1-2\alpha}{1-\alpha}(1+\epsilon))) \\
& \leq \alpha \cdot n + \alpha \frac{1}{4}(n-1)^2 \cdot \frac{1-2\alpha}{1-\alpha}(1+\epsilon) + 2(1-\alpha) + 2n(1-2\alpha)(1+\epsilon).
\end{aligned}
$$

Where the last transition is due to the fact that:

$$
\begin{aligned}
\sum_{i=1}^{\lfloor n/2 \rfloor} (i-1) + \sum_{i=\lfloor n/2 \rfloor + 1}^{n} (n-i) \leq \sum_{i=1}^{\lfloor n/2 \rfloor - 1} i + \sum_{i=1}^{\lfloor n/2 \rfloor} i &= (\lfloor n/2 \rfloor - 1) \cdot \lfloor n/2 \rfloor + \lfloor n/2 \rfloor \\
&= \lfloor n/2 \rfloor^2 = \frac{1}{4}(n-1)^2.
\end{aligned}
$$

The cost of the best Nash equilibrium x is simply $c(x) = 2(1-\alpha)(n+1)$. Interestingly, once we pick $\alpha < \frac{1}{2}$ the maximum PoS for this example is obtained for an intermediate value of n. By taking the first derivative of the function $\frac{c(x)}{c(b)}$ with respect to n and comparing it to 0, we get that the maximum PoS is achieved for $n = \lceil \frac{1-2\alpha-2\sqrt{2-7\alpha+6\alpha^2}}{-1+2\alpha} \rceil$ or $n = \lfloor \frac{1-2\alpha-2\sqrt{2-7\alpha+6\alpha^2}}{-1+2\alpha} \rfloor$.

In Figure 3.3 we plot the lower bound on the PoS that can be achieved by this example (solid line). As one might can expect, as α approaches $\frac{1}{2}$ the PoS approaches 2. For comparison, we also plot (via the dashed line) the lower bound of $\frac{2}{2-\alpha}$ on the PoS that we computed in Section 3.4.3 for instances in which the best Nash equilibrium differs from an optimal solution only in the strategy played by a single player. Interestingly, each of the two constructions offers a better lower bound on the PoS for a different interval of α. We cannot rule out that the maximum of these two constructions could match the best achievable upper bound on the PoS for all α, but there may also be other constructions that can achieve higher lower bounds on the PoS for some ranges of α.

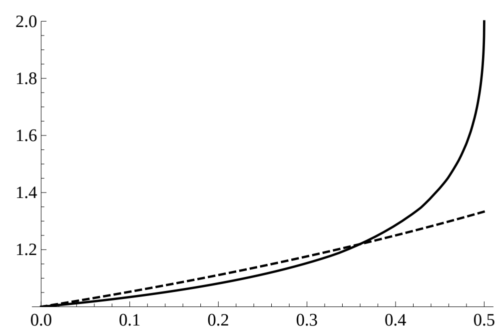


Figure 3.3: The PoS achievable by a path (solid) and by a single strategic node (dashed).

CHAPTER 4

CULTURAL DYNAMICS

4.1 Introduction

So far we have discussed models for opinion formation in which individuals form their opinion on a single topic. In this chapter we take a different approach and discuss how people form their opinions on a set of topics simultaneously. This is usually referred to as cultural dynamics, which builds on the interaction patterns between different cultures for answering questions such as how cultures evolve.

Human societies exhibit many forms of cultural diversity — in the languages that are spoken, in the opinions and values that are held, and in many other dimensions. An active body of research in the mathematical social sciences has developed models for reasoning about the origins of this diversity, and about how it evolves over time.

One of the fundamental principles driving cultural diversity is the tension between two forces: influence and selection. *Influence* refers to the tendency of people to become similar to those with whom they interact, whereas *Selection* is the tendency of people to interact with those who are more similar to them, and/or to be more receptive to influence from those who are similar.

Both of these forces lead toward outcomes in which people end up interacting with others like themselves, but in different ways: influence tends to promote homogeneity, as people shift their behaviors to become alike, while selection tends to promote fragmentation, in which a society can split into multiple groups that have less and less interaction with each other. Research that offers qualitative analyses for issues such as consensus-building, political polarization, or social stratification can often be interpreted through the lens of this influence-selection trade-off [31, 60, 69][1]. The trade-off between influence and selection, and the development of data analysis techniques to try separating the effects of the two, have been integral to understanding and promoting the adoption of products and behaviors in social networks [6, 8, 24, 67, 92], an active line of work at the interface of computing, economics, and statistics.

When both influence and selection are operating at the same time, how should we reason about their combined effects? Several lines of modeling work have approached this question, all starting from similar underlying motivations, but developing different mathematical formalisms.

1. Research on political opinions has studied populations in which each person holds an opinion. The opinion is represented by a number drawn from a bounded interval on the real line $\mathbb{R}^1$, or from a discrete set of points in an interval. (For example, the interval may represent the political spectrum from liberal to conservative.) Each person is influenced by the opinions of others who are sufficiently nearby on the interval, thus capturing the interplay between influence (people are shifting their opinions based on the opinions of others) and selection (people only pay attention to others whose opinions are sufficiently close) [15, 34, 56]. Other versions of models for opinion formation were discussed in the previous two chapters.

2. Axelrod proposed a model of cultural diversity in which there are several *dimensions* of culture, and each person has a value associated with each di-

[1]The term *homophily* is often used to refer to the mechanism of selection [69]. However, in other contexts, it is used to refer to the broader fact that people tend to be similar to their neighbors in a social network, regardless of the mechanism leading to this similarity. Hence, we use the more specific term *selection* here.

mension (e.g., a choice of language, religion, or political affiliation). Agents are more likely to interact when they agree on more dimensions; when two people interact, one person randomly chooses a dimension in which they differ, and changes his value so that they now match in this dimension [10]. For example, two people who have passions for similar sports and styles of food may end up having an easier time (and more opportunity for) associating, and hence an easier time influencing one another along another dimension such as religious beliefs. Again, the model represents an influence process in which the interactions are governed by selection based on (cultural) similarity. Axelrod's model has generated a large amount of subsequent work; see [27] for a survey.

3. Finally, Abrams and Strogatz exhibited some of the interesting effects that can occur even when there are only two types of people. They modeled a scenario in which people speak one of two languages. People mainly interact with speakers of their own language, but there is gradual "leakage" over time as speakers of one language may convert to become speakers of the other [2]. The Abrams-Strogatz model has also generated an active line of follow-up results, including explorations of its microfoundations through agent-based simulation [96] and analyses of the spatial effects and population density [82].

Commonalities among Models

Although the models described above differ in many details, they have the same underlying structure: the population is divided into a set of *types* (the opinions, the cultural choices, the language spoken), and a person of any given type may be influenced to switch types, but only by others whose types are sufficiently similar.

(In the case of the Abrams-Strogatz model, there is a preference for one's own type, but since there are only two types, all types can influence each other.) This process generates a "flow" as people migrate among different types, and we can ask questions about both dynamics (which outcomes the process will reach) and equilibria (which outcomes are self-sustaining, in the sense that the flows between types preserve the fraction of people who belong to each type). Following the language around Axelrod's work, we will refer to this type of process as representing the *cultural dynamics* of the population.

In addition to their similarities in structure, these cultural dynamics models also agree in their broad conclusions. In the first two models, the population gradually separates into distinct "islands" in the space of possible types; subsequently, no further interaction between the islands is possible. In the Abrams-Strogatz model, with just two types, the only outcomes that are stable under perturbations are the two extreme outcomes in which everyone ends up belonging to the same type. Typically, there is also an unstable equilibrium in which each language is spoken by a non-zero fraction of the population.

The most salient difference among the models is the structure that is imposed on the set of types. In each case, there is an undirected *influence graph* $\mathcal{T}$ on the set of types: when a person of type u interacts with a person of type v, the person of type u has the potential to switch to (or move towards) v *provided* that u and v are neighbors in $\mathcal{T}$ (i.e., provided that u and v are sufficiently similar according to the interpretation of the model). In the models of one-dimensional opinion dynamics on a discrete set, the graph $\mathcal{T}$ is the k^{th} power of a path for some $k \geq 1$ (types are similar enough when they are within k steps on the path); in Axelrod's model, the graph $\mathcal{T}$ is the k^{th} power of a (not necessarily binary) hypercube. The Abrams-Strogatz model shows that these kinds of processes can exhibit subtle behavior even on a two-node influence graph $\mathcal{T}$.

Cultural Dynamics on an Arbitrary Influence Graph

All of the prior results apply only to highly structured, symmetric graphs (essentially hypercubes and paths), whereas in some of the settings that the models seek to capture, the set of types can have a less orderly structure. A basic open problem is to analyze cultural dynamics on an arbitrary influence graph.

This is the problem we address in this chapter, where we develop techniques for resolving some of the main questions on arbitrary graphs. For a natural formulation of cultural dynamics on an arbitrary influence graph (which refer to as the *global model*, for reasons explained later), we prove convergence results and precisely characterize the set of all stable equilibria. We then consider generalizations of the global model, extending some of our convergence and stability results to these more general settings and posing several open questions.

The Global Model

We now describe the global model in more detail. Because the models from the earlier lines of work discussed above differ in many of their details, there is no meaningful way to simultaneously generalize all of them in a precise syntactic sense. Instead, our goal is to formulate a version of cultural dynamics that exhibits the same basic interplay of selection and influence — specifically, the idea that influence only happens among types that are "close together" — while allowing for an arbitrary graph on the set of types.

Let $\mathcal{T}$ be a graph on a set of types V; for each type $u \in V$, let $T_u \subseteq V$ denote the set of u's neighbors in $\mathcal{T}$. As is standard in many of the approaches to cultural dynamics, we model the population as a continuum: at the start of the process, each type $u \in V$ has a non-negative population *mass* associated with it, corresponding to the fraction of the population that initially has this type. (Consider, for example, the fraction of the world's population that belongs to a certain religion or speaks a certain language.) Time evolves in discrete steps $t = 0, 1, 2, \ldots$, and $x_u(t)$ denotes the mass on type u at time t. The full state of the population at time t is thus given by the *mass vector* $x(t)$, the vector of values $x_u(t)$ for all $u \in V$.

We define a discrete-time dynamical system in which $x_u(t+1)$ is determined in terms of the mass vector $x(t)$. The dynamical system is motivated by imagining that each person chooses a random other person to interact with. Selection effects are captured in two ways by the model: first, people are more likely to interact with their type; and second, they only have the potential to be influenced when they interact with an individual of their own or a neighboring type. Specifically, each person is α times more likely to choose an interaction partner of their own type than someone of a different type, for a parameter $\alpha \geq 1$. When a person of type u chooses to interact with a person of type v, such that $v \in T_u$, then with probability p, he will switch to type v, where $p \in (0, 1]$ is a fixed parameter. To express this dynamic numerically, we let $\mathcal{M}_u(t) = \alpha x_u(t) + \sum_{v \in V \setminus \{u\}} x_v(t)$. A person of type u chooses to interact with his own type with probability $\alpha x_u(t)/\mathcal{M}_u(t)$, and with any other type $v \neq u$ with probability $x_v(t)/\mathcal{M}_u(t)$. Since a person can change his type from u to v only when $v \in T_u$, the probability that a person of type u keeps his current type after one interaction is $(1-p) + p \cdot \dfrac{\alpha x_u(t) + \sum_{v \in V \setminus (T_u \cup \{u\})} x_v(t)}{\mathcal{M}_u(t)}$.

It is easy to translate this idea into a deterministic update rule on the continuum of people: following the motivation above, we set

$$x_u(t+1) = (1-p)\,x_u(t) + p{\cdot}x_u(t)\cdot\frac{\alpha x_u(t) + \sum_{v\in V\setminus(T_u\cup\{u\})} x_v(t)}{\mathcal{M}_u(t)} + p\sum_{v:u\in T_v} x_v(t)\cdot\frac{x_u(t)}{\mathcal{M}_v(t)}, \tag{4.1}$$

where the first two terms represent the mass that stays at u from time t to $t+1$ (because this fraction of u was not influenced or interacted either with other members of u, or with types $v \notin \{u\}\cup T_u$), and the third term incorporates all the mass that moves from other types v to u. This defines the update rule for the mass vectors x in the dynamical system.

It is natural to think of the parameter p as generally being very small, since most interactions between people do not lead to a change of type. However, as it turns out, the value of p does not have a major qualitative effect on our results. This is not surprising, since introducing $p<1$ (as opposed to $p=1$) just slows down the "flow" between any two types by a factor of $1/p$. We include p in the model in order to capture the range of possible speeds at which transitions can happen. For example, if the types in our model correspond to dialects of a language, we can choose a small p (since the probability that a person changes his dialect is very small). However, if instead the types represent opinions in the period before an election, people may switch much more rapidly, and a larger p is appropriate.

Convergence, Equilibria, and Stability in the Global Model

Our first result is that for any influence graph $\mathcal{T}$ and any initial mass vector x the system converges to a limit mass vector x^*. We prove this by establishing a system of invariants on the population masses over time, capturing a certain "rich-get-richer" property of the process — essentially, that the types of large mass will tend to grow at the expense of the types of small mass.

We next consider the equilibria of this model: we say that a mass vector x is an equilibrium if it remains unchanged after one application of the update rule. It is easy to construct examples of equilibria that are not stable, in the sense that an arbitrarily small perturbation of the masses x_u^* can — after further applications of the update rule — push the masses far away from the equilibrium. Such equilibria are less natural as predicted outcomes of the cultural dynamics being modeled, since the population would be unlikely to hold its position near this equilibrium.

To make this statement precise, we use the notion of *Lyapunov stability*. We say that an equilibrium x^* is *Lyapunov stable* (or, more briefly, *stable*) if given any $\varepsilon > 0$, there exists a $\delta > 0$ such that if $||x(t_0) - x^*|| < \delta$, then $||x(t) - x^*|| < \epsilon$, for all $t \geq t_0$.[2] For simplicity, we use the L_1 norm instead of the L_2 norm in what follows, although this is not an important distinction.

We prove that x^* is a stable equilibrium if and only if the set of active types $A(x^*) = \{u : x_u^* > 0\}$ is an independent set in the influence graph $\mathcal{T}$. The proof is based on the rich-get-richer properties of the process; these properties are used to show that after a sufficiently small perturbation to the population masses, the amount by which any type with positive mass can grow is bounded.

[2] One could ask about stronger notions of stability, in particular, *asymptotic stability*, which requires that there exists a $\delta_1 > 0$ such that if $||x(t_0) - x^*|| < \delta_1$, then $x(t) \to x^*$ as $t \to \infty$. Asymptotic stability is not a useful definition for our purposes; for example, if the underlying influence graph $\mathcal{T}$ has no edges, then any assignment of population masses is an equilibrium, but none are asymptotically stable, since there is no way for a small perturbation to converge back to the original state. On the other hand, all equilibria are Lyapunov stable in this simple example.

Interpretations of the Basic Results

These results establish, first of all, a precise sense in which the natural equilibria the model tend to break the population into non-interacting islands. In addition to offering a qualitative statement about fragmentation of opinions, the results also suggest a way of reasoning about the phenomenon by which opinions on different issues tend to become aligned, with an individual's views on one issue providing evidence for his or her views on another [85, 95]. To take a concrete example that already appears on the 2-dimensional hypercube (i.e. the 4-node cycle), consider a setting in which each individual has either a liberal or conservative view on fiscal issues and either a liberal or conservative view on social issues. If we assume that people only influence each other when they agree on at least one of these two categories of issues, then the graph on the set of types is a 4-node cycle. Since our results on stable equilibria indicate that independent sets are favored as outcomes, we can interpret the conclusion in this example as predicting that either the whole population will converge on a single node (representing a uniform choice of views), or on an independent set of two nodes, in which case an individual's opinion on fiscal issues has become correlated with his or her opinion on social issues.

It is also instructive to compare our results to the main result of [2] discussed above. Recall that they consider the influence graph $\mathcal{T} = K_2$ (two connected nodes), and they find that the two stable equilibria are the outcomes in which all the population mass is gathered at a single node. The family of dynamical systems they consider strictly subsumes ours in the special case of a two-node graph, but for the specific system we study, our results imply that their basic finding extends to arbitrary graphs: in any graph, the stable equilibria correspond to the non-empty independent sets, just as Abrams and Strogatz showed for the two-node graph K_2.

A Generalization: Limiting both Interaction and Influence

We now discuss a natural generalization of the model that is significantly more challenging to analyze. In the global model, the members of type u can interact with members of *all* other types, even though they are influenced only by the types in T_u. However, there are settings in which it is more natural to assume that the members of a type only ever *interact* with members of a subset of the other types. For example, this may be a reasonable assumption when types represent different languages.

To capture this idea, we now assume that there are two potentially distinct graphs on the set of types V: the influence graph $\mathcal{T}$ (as before), as well as an undirected *interaction graph* $\mathcal{S}$, where $\mathcal{T}$ is a subgraph of $\mathcal{S}$. Rather than interacting with a person chosen from the full population, a member of type u selects an interaction partner from the set S_u of u's neighbors in $\mathcal{S}$. It is straightforward to write the new update rule for this more general dynamical system, by summing over types in S_u instead of $V \setminus \{u\}$. Specifically, we can now define $\mathcal{M}_u(t) = \alpha x_u(t) + \sum_{v \in S_u} x_v(t)$, and get the following generalized update rule:

$$x_u(t+1) = (1-p)x_u(t) + p{\cdot}x_u(t) \cdot \frac{\alpha x_u(t) + \sum_{v \in S_u \setminus T_u} x_v(t)}{\mathcal{M}_u(t)} + p \sum_{v: u \in T_v} x_v(t) \cdot \frac{x_u(t)}{\mathcal{M}_v(t)}. \tag{4.2}$$

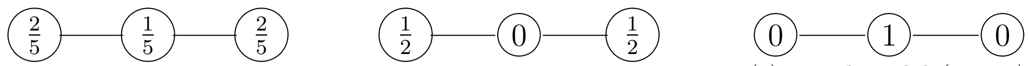
The global model is simply the special case in which the interaction graph $\mathcal{S}$ is the complete graph. The name *global model* emphasizes that each type interacts "globally," with all other types.[3]

[3]There are clearly many other potential generalizations which could incorporate notions of non-uniform interaction, including different interaction strengths between different pairs of types. Such extensions would lead to interesting questions as well. In this chapter, we focus on the generalization with two unweighted graphs $\mathcal{S}$ and $\mathcal{T}$ because it captures in a direct way some of the additional complexity that is introduced by simultaneously modeling limited interaction and influence.

The behavior of this general model is significantly more complex than the behavior of the global model; for instance, for arbitrary $\mathcal{S}$ and $\mathcal{T}$, it is not even clear whether the process will always converge. Intuitively, much of the difficulty comes from the fact that when we consider two neighboring types u and v, the sets of types that they are interacting with, S_u and S_v, can be quite different, whereas in the global model they are both the full set V. Among other things, this can lead to violations of the rich-get-richer property that was so useful for reasoning about the dynamics of the global model.

For the general model, we first establish a necessary condition for equilibria, as well as sufficient conditions for convergence and stability. We then focus further on the special case in which $\mathcal{S} = \mathcal{T}$. This is in a sense the opposite extreme from the global model; instead of making $\mathcal{S}$ as large as it can be, we make it as small as possible subject to the constraint that it contains $\mathcal{T}$ as a subgraph. Accordingly, we refer to the case $\mathcal{S} = \mathcal{T}$ as the *local model*. There are many interesting open questions surrounding the behavior of the local model; we make progress on these through initial convergence results and the identification of a large class of equilibria that are stable for all $\alpha > 1$: non-empty independent sets for which all nodes in the set are at a mutual distance of at least three. In fact, this is an "if and only if" characterization for an important class of influence graphs: those whose connected components are trees or, more generally, bipartite graphs.

An interesting observation is that the local and global models can have genuinely different behaviors starting from the same initial conditions: Figure 4.1 shows an example of an initial mass distribution on the three-node path for which the global model converges to an outcome in which the mass is divided evenly between the two endpoints, while the local model converges to the outcome in which all the mass is on the middle node.



(a) *Initial fractions.* (b) *Global model ($\alpha = 2$).* (c) *Local model ($\alpha = 2$).*

Figure 4.1: An instance in which different models predict convergence to different equilibria. The global model predicts an outcome in which two non-interacting types survive (polarization), whereas the local model predicts that only a single types survives (consensus).

At a higher level, formalizing the distinction between interaction ($\mathcal{S}$) and influence ($\mathcal{T}$) is a potentially promising activity more broadly, particularly in light of the considerable recent interest in the effects of information filtering on the political process. (See [81, 101] for popular media accounts, and [14] for recent experimental research.) The concern expressed in all these lines of work is that personalization on the Internet makes it possible to sharply restrict the diversity of information one sees, and thus risks accentuating the degree of polarization and fragmentation in political discourse — essentially, the risk is that people will only ever be exposed to those who already agree with them, making any kind of consensus almost impossible to achieve.

In this context, our general model also brings into the discussion the interesting contrast between interaction and influence. Personal filtering of information by Internet users can restrict the set of people they interact with (affecting the sets S_u), and it can also, separately, restrict the set of people who may be able to influence them (affecting the sets T_u). These two different effects are often bundled together in discussions of information filtering; it will be interesting to see whether treating them as genuinely distinct can shed additional light on this set of issues.

4.2 Observations on the General Model

In this section, we develop several observations that apply to the fully general model with an arbitrary interaction graph $\mathcal{S}$. In the subsequent sections, we utilize these observations to analyze the global model (where $\mathcal{S}$ is the complete graph) and the local model (where $\mathcal{S} = \mathcal{T}$).

We begin by restating the network dynamics in terms of flows on edges. Recall from the introduction that the fraction of people of type u who switch to type $v \in T_u$ is $p{\cdot}x_v(t)/\mathcal{M}_u(t)$, where $\mathcal{M}_u(t) = \alpha x_u(t) + \sum_{v \in S_u} x_v(t)$. Thus, the fraction of the *entire population* which is moving from u to v is $p{\cdot}x_u(t)x_v(t)/\mathcal{M}_u(t)$. At the same time as this mass of $p{\cdot}x_u(t)x_v(t)/\mathcal{M}_u(t)$ is moving from u to v, a mass of $p{\cdot}x_v(t)x_u(t)/\mathcal{M}_v(t)$ is moving from v to u. These movements partially cancel each other out, and motivate the following definition of the (directed) *flow* on the edge $(v, u) \in \mathcal{T}$:

$$f_{v \to u}(t) = p{\cdot}x_v(t)\, x_u(t) \left(\frac{1}{\mathcal{M}_v(t)} - \frac{1}{\mathcal{M}_u(t)} \right). \tag{4.3}$$

The updated mass can then be written as

$$x_u(t+1) = x_u(t) + \sum_{v \in T_u} f_{v \to u}(t). \tag{4.4}$$

It is easy to check that (4.4) is equivalent to the original dynamics (4.2). Interestingly, the $\mathcal{M}_u(t)$ terms emerge as crucial quantities that determine the direction of the flow; for that reason, we will call $\mathcal{M}_u(t)$ the *interaction mass* of node u at time t.

We say that a node u is *active* at time t if $x_u(t) > 0$, and *inactive* if $x_u(t) = 0$. We occasionally refer to a node u as x-*active* if it is active in x and x-*inactive* otherwise. The set of all active nodes under x is denoted by $A(x)$. Much of our analysis concerns the structure of the subgraph $\mathcal{T}_{\mathsf{act}}(x)$ of the influence graph $\mathcal{T}$ induced by the active nodes $A(x)$. We begin by characterizing when mass vectors are in equilibrium.

Proposition 4.2.1 *A vector x^* is an equilibrium if and only if each connected component C of $\mathcal{T}_{\mathsf{act}}(x^*)$ has the property that all nodes $u \in C$ have the same interaction mass $\mathcal{M}_u$.*

Proof: For the if direction, let $(u, v) \in \mathcal{T}$ be an arbitrary edge. If (u, v) lies inside a component C of $\mathcal{T}_{\mathsf{act}}(x^*)$, then by assumption, both u and v have the same interaction mass, so the flow between them will be 0. Otherwise, at least one of u, v is inactive, so that $x_u x_v = 0$, and again the flow is 0. Thus, x^* is an equilibrium.

For the converse direction, assume that for some connected component C of $\mathcal{T}_{\mathsf{act}}(x^*)$, the nodes do not all have the same interaction mass. Let $U \subseteq C$ be the set of all nodes $u \in U$ minimizing $\mathcal{M}_u(t)$. By assumption, $U \subsetneq C$, and because C is connected, there must be a node pair $u \in U, v \in C \setminus U$ which is adjacent in $\mathcal{T}$. Fix this node pair.

By definition of U, the flow along any edge $(u, v) \in \mathcal{T}$ is non-negative, and the flow along (u, v) is strictly positive. Thus, after one application of the update rule, the mass on node u strictly decreases, which means that x^* cannot be an equilibrium. $\square$

The following useful lemma relates convergence and the change in directions of flows:

Lemma 4.2.2 *If there exists a time t_0 such that the flows do not change direction after time t_0, then the system converges.*

Proof: Let G be the directed graph obtained by directing each edge (u, v) of $\mathcal{T}$ according to the direction of the corresponding flow $f_{u \to v}(t_0)$. By the assumption, these directions stay constant after time t_0. As flow always goes from types with smaller interaction mass to types with larger interaction mass, G must be acyclic. Let $v_1, v_2, \ldots, v_n$ be a topological sorting of the graph, so that all directed edges of G are of the form $(v_i, v_j), i < j$.

We define $X_k(t) = \sum_{i=1}^{k} x_{v_i}(t)$ to be the total mass at time $t \geq t_0$ on the k first nodes in the topological sorting. Because the total mass in the system is constant, and all flow goes from nodes with lower indices to nodes with higher indices, each of the $X_k(t)$ must be non-increasing as a function of t. Since they are also lower-bounded by 0, each $X_k(t)$ must converge to some value Z_k as $t \to \infty$. Therefore, each $x_{v_i}(t)$ converges to $Z_i - Z_{i-1}$ as $t \to \infty$. □

Recall that we are interested in characterizing stable equilibria. We next provide a sufficient condition.

Proposition 4.2.3 *An equilibrium x^* is stable if it satisfies the following two properties:*

1. *The active nodes form an independent set in the influence graph $\mathcal{T}$.*
2. *The interaction mass of every active node is strictly greater than the interaction mass of each of its inactive neighbors in the influence graph $\mathcal{T}$.*

Proof: Let x^* be an equilibrium for which both properties hold. A node u is called x^*-*active* if it is active in x^* and x^*-*inactive* otherwise. Let A be the set of all x^*-active nodes, and $\mathcal{M}_u^*$ denote the interaction mass of node u with respect to x^*. Define

$$\delta = \tfrac{1}{2\alpha+1} \min_{u \in A,\, v \notin A,\, (u,v) \in \mathcal{T}} (\mathcal{M}_u^* - \mathcal{M}_v^*) \;>\; 0 \tag{4.5}$$

by the second property of x^*. To show stability, we prove that whenever $||x^* - x(t_0)||_1 \leq \delta$, the system will satisfy $||x^* - x(t)||_1 \leq \delta$ for all time steps $t \geq t_0$.

The key step of the proof is to establish that for each node $u \in A$, the mass $x_u(t)$ is non-decreasing over time. For contradiction, assume that t is minimal, and $u \in A$ an x^*-active node, such that $x_u(t+1) < x_u(t)$. Because A is an independent set in $\mathcal{T}$, all of u's neighbors in $\mathcal{T}$ must be x^*-inactive. By definition of t, we have that $\sum_{v \notin A} x_v(t) \leq \delta$, and $\sum_{v \in A} x_v(t) \geq \sum_{v \in A} x_v^* - \delta$. In particular, this implies that $\mathcal{M}_u(t) \geq \mathcal{M}_u^* - \alpha\delta$ and for any neighbor v of u, $\mathcal{M}_v(t) \leq \mathcal{M}_v^* + \alpha\delta$, so we obtain that $\mathcal{M}_u(t) - \mathcal{M}_v(t) \geq (\mathcal{M}_u^* - \mathcal{M}_v^*) - 2\alpha\delta \overset{(4.5)}{\geq} (2\alpha+1)\delta - (2\alpha)\delta > 0$.

This implies that no flow could have been directed from u to v, for any $v \in T_u$, contradicting that u's mass decreased.

Finally, because each of the $x_u(t), u \in A$ is non-decreasing, mass can only move among x^*-inactive nodes, or from x^*-inactive nodes to x^*-active ones. Therefore, it holds that $\sum_{u \in A} |x_u(t) - x_u(t_0)| = \sum_{u \notin A} x_u(t_0) - \sum_{u \notin A} x_u(t)$. Thus,

$$\begin{aligned} ||x(t) - x^*||_1 &\leq \sum_{u \in A} |x_u(t) - x_u(t_0)| + \sum_{u \in A} |x_u(t_0) - x_u^*| + \sum_{u \notin A} x_u(t) \\ &= \sum_{u \in A} |x_u(t_0) - x_u^*| + \sum_{u \notin A} x_u(t_0) = ||x(t_0) - x^*||_1 \;\leq\; \delta, \end{aligned}$$

so the system is stable. $\square$

4.3 The Global Model

In this section, we analyze the global model. The definition of the general model states that flows are always directed from nodes with smaller interaction mass to nodes with larger interaction mass. For the global model, this property is simplified significantly: flow is always directed from types with smaller mass to types with larger mass. This property lets us achieve an almost complete understanding of the global model. We show that for this model, the system always converges, and we present a complete characterization of which equilibria are stable. First, we characterize equilibria by applying Proposition 4.2.1 to the global model.

Corollary 4.3.1 *Under the global model with $\alpha > 1$, the system is at equilibrium x^* if and only if the following holds: for every connected component C of $\mathcal{T}_{\mathsf{act}}(x^*)$, all nodes $u \in C$ have the same mass.*

Proof: Proposition 4.2.1 guarantees that x^* is at equilibrium if and only if for each edge $(u, u') \in \mathcal{T}_{\mathsf{act}}(x^*)$: $\alpha x_u^* + \sum_{v \in S_u} x_v^* = \alpha x_{u'}^* + \sum_{v \in S_{u'}} x_v^*$. In the global model, for any node u, the set S_u consists of of all types but u itself, implying that the sum cancels out, and we obtain $(\alpha - 1)x_u^* = (\alpha - 1)x_{u'}^*$. For $\alpha > 1$, this implies $x_u^* = x_{u'}^*$. $\square$

We next show that the system always converges; the proof relies on the key invariant that for any $1 \leq k \leq n$, the total mass of the k smallest types never increases over time. More formally, we define the following quantities:

Definition 4.3.2 *Let $y_1(t) \leq y_2(t) \leq \ldots \leq y_n(t)$ be the node masses sorted in* *non-decreasing order. Define*

$$Y_k(t) = \sum_{i \le k} y_i(t) \;=\; \min_{R:|R|=k} \sum_{v \in R} x_v(t) \tag{4.6}$$

to be the sum of the masses of the k smallest nodes at time t.

The following lemma formally captures the notion that the rich get richer in the global model.

Lemma 4.3.3 *For every k, the function $Y_k(t)$ is non-increasing in t.*

Proof: Let t, k be arbitrary, and assume for contradiction that $Y_k(t+1) > Y_k(t)$. Let $R(t)$ be the set of k nodes achieving the minimum in (4.6) at time t, and similarly for $R(t+1)$ at time $t+1$ (breaking ties arbitrarily). By assumption and the definition of $R(t+1)$, we have

$$\sum_{v \in R(t)} x_v(t) \;<\; \sum_{v \in R(t+1)} x_v(t+1) \;\le\; \sum_{v \in R(t)} x_v(t+1),$$

so the total mass of nodes in $R(t)$ must have strictly increased from time t to time $t+1$. In particular, because the total mass of all nodes is constant over time, this means that the flow from $\overline{R}(t)$ (the complement of $R(t)$) to $R(t)$ was strictly positive, so it must have been strictly positive along some edge (u, v) with $u \notin R(t), v \in R(t)$. But by definition of $R(t)$, we have that $x_u(t) \ge x_v(t)$, a contradiction to the flow dynamics of the global model. $\square$

Theorem 4.3.4 *Under the global model, the system converges for any influence graph and any starting mass vector $x(0)$.*

Proof: By Lemma 4.3.3, each function $Y_j(t)$ is non-increasing in t. As all masses are non-negative, the $Y_j(t)$ are also bounded below by 0. Hence, each function $Y_j(t)$ must converge to some value Z_j. Thus, each function $y_j(t)$ must converge to $Z_j - Z_{j-1} =: z_j$. It remains to show that this also implies convergence of $x(t)$.

Let $\delta > 0$ be at most the smallest difference between any two distinct z_j, i.e., $\delta \leq \min_{i,j:z_i \neq z_j} |z_i - z_j|$.

Let t_0 be large enough that $|y_i(t) - z_i| < \frac{\delta}{2n}$ for all $t \geq t_0$. Note that this also implies that $Y_j(t) - Z_j \leq \sum_{i=1}^{j} |y_i(t) - z_i| < j \cdot \frac{\delta}{2n}$.

We will show that if for some node v, there exists a time $t \geq t_0$ such that $x_v(t) = y_j(t)$ and $x_v(t+1) = y_{j'}(t+1)$, then $z_j = z'_j$. Assume towards a contradiction that this is not the case; hence, there exist v and $t \geq t_0$ such that $x_v(t) = y_j(t)$, $x_v(t+1) = y_{j'}(t+1)$ and $z_j \neq z'_j$. Furthermore assume that $j < j'$. (The proof for the case of $j' < j$ is similar.) Then, $x_v(t) \leq z_j + \frac{\delta}{2n}$ and $x_v(t+1) \geq z_{j'} - \frac{\delta}{2n} \geq z_j + \frac{2n-1}{2n}\delta$. Therefore, in time step t, v gained at least a mass of $\frac{2n-2}{2n}\delta$.

The source of this mass can only be nodes which at time t had smaller masses than v. Hence,

$$Y_{j-1}(t+1) - Z_{j-1} \;\leq\; Y_{j-1}(t) - Z_{j-1} - \frac{2n-1}{2n}\delta \;\leq\; (j-1) \cdot \frac{\delta}{2n} - \frac{2n-1}{2n}\delta \;<\; 0.$$

This is a contradiction to the fact that $Y_{j-1}(t)$ is non increasing and converges to Z_{j-1}. $\square$

4.3.1 Characterization of Lyapunov Stable Equilibria

For the global model, the properties required for Proposition 4.2.3 hold for any independent set, since the interaction mass of active types is always greater than the interaction mass of inactive types. Therefore, any equilibrium in which the set of active nodes is independent is stable. To complete the characterization, we show that the converse is also true.

Theorem 4.3.5 *In the global model, an equilibrium x^* is Lyapunov stable if and only if the active nodes form an independent set.*

Proof: It remains to prove the "only if" direction. Assume the active nodes in an equilibrium x^* do not form an independent set. We will prove that x^* is not Lyapunov-stable.

Let C be a connected component of size $|C| \geq 2$ in $\mathcal{T}_{\texttt{act}}(x^*)$. By the assumption that the active nodes in x^* do not form an independent set, such a connected component exists. Notice that each component of $\mathcal{T}_{\texttt{act}}(x^*)$ evolves in isolation, so we can focus on only C for the rest of the proof. Therefore, by Corollary 4.3.1, $x_v^* = \mu$ for all $v \in C$, for some value μ.

Let $u, v \in C$ be two arbitrary nodes, and $\delta > 0$ be arbitrarily small. Consider the following perturbation: $x_u = x_u^* + \delta, x_v = x_v^* - \delta$, and $x_w = x_w^*$ for all $w \neq u, v$. By Theorem 4.3.4, the system, starting from the perturbed vector x, will converge to some new equilibrium y. By Lemma 4.3.3, the smallest mass of any node in C will always be at most $\mu - \delta$ during the process. All y-active nodes must have the same mass; therefore, if all nodes were active in y, they would all have to have mass at most $\mu - \delta$, which would imply that mass has disappeared from C, a contradiction. Hence, at least one node of C must end up inactive in y. In particular, this means that $||x^* - y||$ is not bounded in terms of δ, and x^* is not Lyapunov stable. $\square$

4.4 The Local Model

In the previous section, we have given essentially complete characterizations of convergence and stability of equilibria under the global model, in which all types have the potential to interact, even though only certain pairs of types can influence each other (according to the graph $\mathcal{T}$).

We now consider the local model, which is at the other extreme of our general family: here, the interaction graph $\mathcal{S}$ is the same as the influence graph $\mathcal{T}$; hence, interactions occur only between individuals who also have the potential to influence each other. (We will generally denote this underlying graph by $\mathcal{T}$, with the understanding that $\mathcal{S} = \mathcal{T}$.) We find that the problems of convergence and stability are much more challenging in this case. For the global model, we were able to extract very useful organizing structures in the dynamical system that gave us a natural progress measure toward convergence. But as is well known, in general, a dynamical system on even a small number of variables may have convergence properties that are extremely difficult to analyze or express. Given the complex behavior of the update rules for the local model, we find that the convergence and stability questions are already difficult on graphs $\mathcal{T}$ with a small number of nodes, and we focus our results here on such cases. (Of course, based on the motivating premise of the model, even systems with a small number of variables are frequently natural, corresponding to selection and influence dynamics in societies with, for example, a small number of languages, a small number of political parties, or a small number of dominant religions or cultures.)

4.4.1 Convergence of a 3-path for $\alpha > 1$

We begin by considering the case $\alpha > 1$ and first prove the following theorem:

Theorem 4.4.1 *Under the local model, if the influence graph is a 3-path, then the system converges from any starting state.*

We first provide a brief outline. The subtle difficulty arises due to the fact that the flow between two types u and v does not necessarily go in the same direction over all time steps, but instead may change its direction. To keep track of the changes in direction, we define a *configuration* of the system to be a labeling of all edges (u, v) in $\mathcal{T}$ by the direction along which flow is traveling (i.e., whether from u to v or from v to u). In the case of a 3-node path, there are four possible configurations. We study transitions among the configurations as the system evolves over time; we show that each configuration is either a *sink*, which cannot transition to any other configuration, or it has the property that any change in the direction of an edge leads to a sink configuration. This ensures that there can be at most one change in the direction of flow as the system evolves; hence, there is a time t_0 such that for any $t > t_0$, no flow changes its direction. After this point, Lemma 4.2.2 guarantees that the system converges. For the case $\alpha \geq 2$, we show this fact only for the 3-path; for $\alpha < 2$, we establish a more general result, showing the same fact for arbitrary star graphs.

Lemma 4.4.2 *Consider the local model with $\alpha \geq 2$ such that the influence graph $\mathcal{T}$ is a 3-node path. Then, there is a time t_0 such that for any $t > t_0$, no flow changes its direction.*

Proof: Let the nodes of the path be $(1, 2, 3)$, in order. Consider an arbitrary time t. Consider the change in the interaction mass of node 1, for example, from step t to $t+1$. Recall that $\mathcal{M}_1(t) = \alpha x_1(t) + x_2(t)$. Node 1's interaction mass is decreased by $\alpha f_{1\to 2}(t)$ from flow leaving node 1 to node 2, and increased by $f_{1\to 2}(t) + f_{3\to 2}(t)$ from flow entering node 2. By applying the same reasoning to nodes 2 and 3, we get:

$$\begin{aligned}
\mathcal{M}_1(t+1) &= \mathcal{M}_1(t) + f_{3\to 2}(t) - (\alpha - 1) f_{1\to 2}(t), \\
\mathcal{M}_2(t+1) &= \mathcal{M}_2(t) + (\alpha - 1)(f_{1\to 2}(t) + f_{3\to 2}(t)), \\
\mathcal{M}_3(t+1) &= \mathcal{M}_3(t) + f_{1\to 2}(t) - (\alpha - 1) f_{3\to 2}(t).
\end{aligned} \tag{4.7}$$

Let $x_i = x_i(t), \mathcal{M}_i = \mathcal{M}_i(t), f_{i\to j} = f_{i\to j}(t)$ for $i, j = 1, 2, 3$. We will distinguish three cases based on the relative sizes of $\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3$.

1. If $\mathcal{M}_2 \geq \mathcal{M}_1$ and $\mathcal{M}_2 \geq \mathcal{M}_3$, then both $f_{1\to 2}$ and $f_{3\to 2}$ are non-negative. According to Equation (4.7), $\mathcal{M}_2$ increases by at least as much as both $\mathcal{M}_1$ and $\mathcal{M}_3$, so the same inequality will hold in the next step (and thus inductively forever). Thus, we have reached a sink configuration.

2. If $\mathcal{M}_2 < \mathcal{M}_1$ and $\mathcal{M}_2 < \mathcal{M}_3$, then both $f_{1\to 2}$ and $f_{3\to 2}$ are negative. By Equation (4.7), $\mathcal{M}_2$ decreases by at least as much as both $\mathcal{M}_1$ and $\mathcal{M}_3$, so again, the inequalities will hold forever, and we have reached a sink configuration.

3. The remaining case is that $\mathcal{M}_2 < \mathcal{M}_1$ and $\mathcal{M}_2 \geq \mathcal{M}_3$. (The case $\mathcal{M}_2 < \mathcal{M}_3, \mathcal{M}_2 \geq \mathcal{M}_1$ is symmetric.) Here, $\mathcal{M}_3$ decreases, $\mathcal{M}_1$ increases, and $\mathcal{M}_2$

may increase or decrease. If the relative order of $\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3$ stays the same for all steps after t, then we have reached a sink configuration. Otherwise, at some time $t' \geq t$, we must reach either a configuration with $\mathcal{M}_2(t') < \mathcal{M}_1(t'), \mathcal{M}_2(t') < \mathcal{M}_3(t')$ or with $\mathcal{M}_2(t') \geq \mathcal{M}_1(t'), \mathcal{M}_2(t') \geq \mathcal{M}_3(t')$. Either of those configurations is a sink configuration by the preceding two cases.

In summary, each configuration is either a sink configuration, or will reach a sink configuration at the next transition to a different order of interaction masses. $\square$

Next we prove that for $\alpha < 2$ the process converges on every star graph (and in particular on the 3-path).

Theorem 4.4.3 *Under the local model with $\alpha < 2$, if the influence graph is a star graph, then the system converges from any starting state.*

In the remainder of this section we prove Theorem 4.4.3. More specifically, we show that eventually the system enters a sink configuration.

Lemma 4.4.4 *Consider the local model such that the influence graph $\mathcal{T}$ is a star graph. Then, at any time, the number of edges with flow directed away from the center is at most $\lfloor \alpha \rfloor$.*

Proof: Denote the central node by u. Fix some time t, and let R be the set of all peripheral nodes v such that the flow on the edge (u, v) is directed from u to v. Because the flow is directed towards v, $\mathcal{M}_v > \mathcal{M}_u$ for all $v \in R$. Rearranging this inequality gives us that $(\alpha - 1)(x_v - x_u) > \sum_{w \neq u,v} x_w$, which implies in particular that $x_v > \frac{\sum_{w \neq u,v} x_w}{\alpha - 1}$. Summing over all $v \in R$ now implies that

$$\sum_{v \in R} x_v > \sum_{v \in R} \frac{\sum_{w \neq v,u} x_w}{\alpha - 1} \; > \; (|R| - 1)\frac{\sum_{v \in R} x_v}{\alpha - 1}.$$

Thus we have that $|R| \leq \lfloor \alpha \rfloor$. □

Lemma 4.4.5 *Consider the local model such that the influence graph $\mathcal{T}$ is a star graph, and with $\alpha < 2$. Then, a configuration in which a flow on exactly one edge is directed away from the center node is a sink configuration.*

Proof: Let u be the center node. Suppose that at time t, the system is in a configuration in which the flow on exactly one edge (u, v) is directed away from the center; so $\mathcal{M}_u(t) < \mathcal{M}_v(t)$.

We will prove that $\mathcal{M}_u(t+1) < \mathcal{M}_v(t+1)$. By Lemma 4.4.4 there can be at most one edge on which the flow is directed away from u, and (u, v) is such an edge; hence, the configuration at time $t+1$ is the same as at time t. To see that $\mathcal{M}_u(t+1) < \mathcal{M}_v(t+1)$, observe that the new interaction masses are

$$\mathcal{M}_u(t+1) = \mathcal{M}_u(t) + (\alpha - 1) \sum_{w \neq u,v} f_{w \to u}(t) - (\alpha - 1) f_{u \to v}(t)$$

$$\mathcal{M}_v(t+1) = \mathcal{M}_v(t) + (\alpha - 1) f_{u \to v}(t) + \sum_{w \neq u,v} f_{w \to u}(t).$$

Now, since $\mathcal{M}_u(t) < \mathcal{M}_v(t)$, their difference is

$$\mathcal{M}_u(t+1) - \mathcal{M}_v(t+1) < (\alpha - 2) \sum_{w \neq u,v} f_{w \to u}(t) - 2(\alpha - 1) f_{u \to v}(t).$$

Because $\alpha < 2$, the right-hand side is negative, and we have proved that $\mathcal{M}_u(t+1) < \mathcal{M}_v(t+1)$. □

Theorem 4.4.3 now follows from Lemmas 4.2.2, 4.4.4 and 4.4.5, as follows. If the system ever enters a configuration in which exactly one edge has flow directed away from the center, then by Lemma 4.4.5, it subsequently stays in this configuration forever, so by Lemma 4.2.2, the system converges. By Lemma 4.4.4, the only other alternative is that the system is always in the configuration with all edges directed inwards; then, again, it converges by Lemma 4.2.2.

4.4.2 Convergence on a Path for $\alpha = 1$

For $\alpha = 1$, we are able to prove convergence if the active subgraph is a path of $n \leq 5$ nodes. The proof requires different techniques than the ones we use for $\alpha > 1$: for paths of more than 3 nodes, flows on edges can change their direction infinitely often.

Assume that the active subgraph is an n-node path with nodes $(1, 2, \ldots, n)$. The endpoints of the path, nodes 1 and n, always have interaction masses no larger than their neighbors (nodes $2, n-1$), implying that their masses $x_1(t), x_n(t)$ are monotonically non-increasing. This implies convergence of $x_1(t)$, $x_n(t)$ as $t \to \infty$. In the following proposition, we will exploit the convergence at the endpoints to show that $x_2(t)$ and $x_{n-1}(t)$ must also converge. For a path of length at most 5, this implies convergence of the vector x to an equilibrium, as the total mass stays constant. Our technique does not apply beyond length 5; we do not know of a direct way to generalize the argument inductively to paths of arbitrary lengths.

Theorem 4.4.6 *Consider the local model with $\alpha = 1$. If the influence graph is a path of $n \leq 5$ nodes, then the system converges.*

Proof: We already argued above that $x_1(t)$ and $x_n(t)$ converge. If the path has 3 nodes, then $x_2(t) = 1 - x_1(t) - x_3(t)$ (by mass conservation), so $x_2(t)$ converges as well. So assume that $n \in \{4, 5\}$. Below, we show that $x_2(t)$ converges as well; a symmetric argument applies to $x_{n-1}(t)$. If the path has 4 nodes, we are done at this point. If the path has 5 nodes, then $x_3(t) = 1 - x_1(t) - x_2(t) - x_4(t) - x_5(t)$ must converge as well. Thus, $x(t)$ converges in all cases.

To prove that $x_2(t)$ converges, we distinguish two cases, based on $y_1 = \lim_{t\to\infty} x_1(t)$.

1. If $y_1 = 0$, there are two subcases. If $x_1(t) \geq x_2(t)$ for all t, then clearly, $x_2(t) \to 0$ as well. Otherwise, there exists a t_0 with $x_2(t_0) > x_1(t_0)$. By the definition of the model, specialized to the local model and $\alpha = 1$, we obtain that for any t,

$$x_1(t+1) = p{\cdot}x_1(t) \cdot \left(\frac{x_1(t)}{\mathcal{M}_1(t)} + \frac{x_2(t)}{\mathcal{M}_2(t)}\right) + (1-p) \cdot x_1(t),$$
$$x_2(t+1) = p{\cdot}x_2(t) \cdot \left(\frac{x_1(t)}{\mathcal{M}_1(t)} + \frac{x_2(t)}{\mathcal{M}_2(t)} + \frac{x_3(t)}{\mathcal{M}_3(t)}\right) + (1-p) \cdot x_2(t).$$

   Then, clearly, $x_1(t) < x_2(t)$ implies $x_1(t+1) < x_2(t+1)$. In particular, this means that $x_2(t) > x_1(t)$ for all $t \geq t_0$. In turn, this inequality is used in the last step of the following derivation:

$$\begin{aligned}
\max(f_{3\to 2}(t), 0) &\leq p{\cdot}\frac{x_2(t)\,x_3(t)}{\mathcal{M}_2(t)\,\mathcal{M}_3(t)} \cdot x_1(t) \\
&= p{\cdot}\frac{x_1(t)\,x_2(t)\,x_3(t)}{\mathcal{M}_1(t)\,\mathcal{M}_2(t)} \cdot \frac{\mathcal{M}_1(t)}{\mathcal{M}_3(t)} \\
&= f_{1\to 2}(t) \cdot \frac{x_1(t) + x_2(t)}{x_2(t) + x_3(t) + x_4(t)} \\
&\leq 2\, f_{1\to 2}(t).
\end{aligned}$$

   Thus, the total amount of flow entering node 2 after time t_0 is at most $3 \sum_{t=t_0}^{\infty} f_{1\to 2}(t) \leq 3\, x_1(t_0)$. The reason for the last inequality is that flow never enters node 1, so the total amount of flow that can leave node 1 for node 2 after t_0 is at most the amount that was at node 1 at time t_0.

Let $F^+(t)$ (resp., $F^-(t)$) be the total amount of flow that has entered (resp., left) node 2 up to time t. We have just proved that $F^+(t) - F^+(t_0) \leq 3x_1(t_0)$. Therefore, $F^+(t)$, being monotone and bounded, must converge. Because flow can only leave node 2 when it was already there, we get that $F^-(t) \leq x_2(0) + F^+(t)$ is also bounded, and must also converge. Hence, $x_2(t) = x_2(0) + F^+(t) - F^-(t)$, being the difference between two convergent sequences, must also converge.

2. If $y_1 > 0$, we will pursue a similar argument, but this time bounding the cumulative flow *out of* node 2 instead of into it. Because $x_2(t) \leq 1$ for all times t, this means that the cumulative flow into node 2 must also be bounded. Then, an identical argument to the previous paragraph shows that $x_2(t) = x_2(0) + F^+(t) - F^-(t)$ must converge.

   No flow can ever leave node 2 for node 1, so we just need to bound the flow from node 2 to node 3. Flow leaves node 2 for node 3 at time t if and only if $x_4(t) \geq x_1(t)$. Since $x_1(t) \geq y_1$, it follows that $\mathcal{M}_2(t), \mathcal{M}_3(t) \geq y_1$ as well. Therefore, we can bound the non-negative flow from node 2 to node 3 as follows:

$$\begin{aligned}\max(f_{2\to3}(t), 0) &\leq p \cdot \frac{x_2(t)\, x_3(t)}{\mathcal{M}_2(t)\, \mathcal{M}_3(t)} \cdot x_4(t) \;\leq\; p \cdot \frac{x_2(t)\, x_3(t)}{y_1^2} \\ &\leq p \cdot \frac{x_1(t)\, x_2(t)\, x_3(t)}{y_1^3} \;\leq\; \frac{1}{y_1^3} \cdot p \cdot \frac{x_1(t)\, x_2(t)\, x_3(t)}{\mathcal{M}_1(t)\mathcal{M}_2(t)} \;=\; \frac{1}{y_1^3} \cdot f_{1\to2}(t).\end{aligned}$$

   Thus, the total positive flow from node 2 to node 3 is bounded above by a constant times the total flow from node 1 to node 2, which in turn is at most $x_1(0)$.

$\square$

4.4.3 Characterization of Universally Stable Equilibria

Next, we turn our attention to Lyapunov-stable equilibria. We focus on a very strong notion of stability: stability for all $\alpha > 1$. Formally, we call a mass vector x a *universally stable equilibrium* if x is a Lyapunov-stable equilibrium for every $\alpha > 1$. Our goal here is to investigate which equilibria are universally stable. Such equilibria are robust to (a very idealized notion of) a changing the environment, as expressed by varying α.

Our main result for universally stable equilibria is a complete characterization for influence graphs that are forests, and more generally for influence graphs whose connected components are bipartite graphs.

Theorem 4.4.7 *Consider the local model with $\alpha > 1$. Assume that all connected components of the influence graph $\mathcal{T}$ are bipartite graphs. Then, a mass vector x^* is a universally stable equilibrium if and only if the distance between any two active nodes in x^* is at least* 3.

The proof of Theorem 4.4.7 consists of several sub-results, all of which hold for arbitrary influence graphs, and imply the desired characterization under the assumption in the theorem. It is worth noting that these sub-results constitute significant progress towards understanding the structure of universally stable equilibria for arbitrary influence graphs, as we discuss later. For brevity, if x is a mass vector such that the distance between any two active nodes is at least 3, we will say that x is *3-separated*.

The first proposition proves the "if" direction of Theorem 4.4.7. Its proof follows from our analysis in Section 4.2.

Proposition 4.4.8 *Under the local model with $\alpha > 1$, any 3-separated mass vector x^* is a Lyapunov-stable equilibrium.*

Proof: x^* is an equilibrium by Proposition 4.2.1 since its active nodes form an independent set. By Proposition 4.2.3, an equilibrium whose active nodes form an independent set is stable if the interaction mass of each active node is greater than the interaction mass of each of its neighbors. For $\alpha > 1$, this property holds when each inactive node has at most one active neighbor. In turn, this holds if and only if the distance between every two active nodes is at least 3; thus, all such equilibria are stable for every $\alpha > 1$. □

The "only if" direction of Theorem 4.4.7 is more complicated to prove. First, we show that if the active nodes of an equilibrium do form an independent set, then being 3-separated is necessary to ensure universal stability.

Proposition 4.4.9 *Let x^* be a universally stable equilibrium, and assume that the active nodes under x^* form an independent set. Then, x^* is 3-separated.*

Proof: For the sake of contradiction, suppose that x^* is not 3-separated. Then there exists an inactive node u with at least two active neighbors. Let A_u be the set of all active neighbors of node u. We use the following notation:

$$s = \sum_{v \in A_u} x^*_v; \quad \eta = \min_{v \in A_u} x^*_v; \quad \mu = \max_{v \in A_u} x^*_v.$$

Define $\alpha = 1 + \eta^2$. We will show that x^* is not Lyapunov-stable for this α.

To prove instability, consider a perturbation x which coincides with x^* on all nodes not in $\{u\} \cup A_u$, and satisfies

$$\begin{cases} x_v \leq x_v^* \quad \text{for all } v \in A_u \\ x_u = \delta \quad \text{for some } \delta \in (0, s - \mu - \eta^2) \\ x_u + \sum_{v \in A_u} x_v = \sum_{v \in A_u} x_v^* = s. \end{cases} \tag{4.8}$$

(Note that $s - \mu - \eta^2 > 0$ because A_u consists of at least two nodes.)

Under such an x, the interaction mass of node u is

$$\mathcal{M}_u \;=\; \alpha x_u + \sum_{v \in A_u} x_v \;=\; \eta^2 \cdot x_u + s,$$

while the interaction mass of any node $v \in A_u$ is

$$\mathcal{M}_v \;=\; \alpha x_v + x_u \;\leq\; \alpha x_v^* + x_u \;\leq\; (1 + \eta^2)\mu + x_u.$$

Since $x_u < s - \mu - \eta^2$, we have that $\mathcal{M}_v < (1 + \eta^2)\mu + s - \mu - \eta^2 < s$, and hence, $\mathcal{M}_u > \mathcal{M}_v$. Thus, under this perturbation, mass starts flowing from all nodes $v \in A_u$ to u, and this continues until $x_u \geq s - \mu - \eta^2$. Consequently, the system cannot reach any equilibrium with $x_u < s - \mu - \eta^2$; in particular, it cannot reach any equilibrium with $||x^* - x|| < s - \mu - \eta^2$. Since this holds for arbitrarily small δ, and $s - \mu - \eta^2 > 0$ is a constant independent of δ, we conclude that x^* is not Lyapunov-stable for this α. $\square$

With Proposition 4.4.9 in place, all that remains to complete the proof of the "only if" direction of Theorem 4.4.7 is to ensure that the active nodes in any universally stable equilibrium x^* of a bipartite graph form an independent set, i.e., that each connected component C of $\mathcal{T}_{\texttt{act}}(x^*)$ consists of a single node. This is implied by Lemma 4.4.10, which shows in general that if x^* is a universally stable equilibrium, then all the non-trivial connected components of the subgraph of its active nodes are *not* bipartite graphs. This completes the proof of Theorem 4.4.7, as any connected subgraph of a bipartite graph is a bipartite graph itself.

An additional benefit of Lemma 4.4.10 is that it applies to arbitrary influence graphs, and significantly limits the topologies a connected component of $\mathcal{T}_{\mathsf{act}}(x^*)$ can have for a universally stable equilibrium x^*. To state this lemma in the most general form, we define a class of regular graphs which in particular subsumes all bipartite graphs, all cliques, and all cycles whose length is a multiple of 3. We say that a d-regular graph is *locally balanced* if its vertices can be partitioned into k disjoint sets $V_1, V_2, \ldots, V_k$ such that each vertex $v \in V_i$ has exactly $d/(k-1)$ edges to each of the sets $V_j, j \neq i$.

Lemma 4.4.10 *Let x^* be a universally stable equilibrium and C a non-trivial connected component of its active subgraph $\mathcal{T}_{\mathsf{act}}(x^*)$. Then:*

(a) C is a regular graph, and x^* is uniform on C (i.e., $x^*_u = x^*_v$ for all $u, v \in C$).

(b) C is not *a bipartite graph, and, more generally, C is not locally balanced.*

Proof: We begin by proving part (a). Let $u, v \in C$ be a pair of *adjacent* nodes. The equilibrium conditions for $\alpha = 2$ imply that $2x^*_v + \sum_{w \in T_v} x^*_w = 2x^*_u + \sum_{w \in T_u} x^*_w$, and the ones for $\alpha = 3$ that $3x^*_v + \sum_{w \in T_v} x^*_w = 3x^*_u + \sum_{w \in T_u} x^*_w$. Subtracting the first equation from the second shows that $x^*_v = x^*_u$. Because C is a connected component, applying this argument along all edges in C proves that all nodes in C must have the same mass μ.

The interaction mass of node v with $\alpha = 2$ is therefore $\mathcal{M}_v = \mu \cdot (|T_v| + 1)$. Considering again a pair u, v of adjacent nodes, the equilibrium condition $\mathcal{M}_u =$

$\mathcal{M}_v$ implies that $|T_u| = |T_v|$. Again by connectivity of C, this implies that all nodes in C have the same degree, so C is regular.

Next, we prove part (b). Because x^* is universally stable, part (a) implies that C is d-regular for some $d \geq 1$, and $x^*_u = \mu$ (for some μ) for all $u \in C$. Assume for contradiction that C is locally balanced, and let $V_1, \ldots, V_k$ be the k partitions of C. Because $\mathcal{T}[V_i \cup V_j]$ is a $d/(k-1)$-regular bipartite graph for each pair $i \neq j$, all partitions V_i must have the same size $s = |C|/k$.

Set $\alpha = d + 1$, and let $\delta > 0$ be arbitrary. Consider perturbed vectors of the following form: $x_v = x^*_v + \frac{1}{s} \cdot \delta$ for every $v \in V_1$ and $x_u = x_u - \frac{1}{s(k-1)} \cdot \delta$ for every $u \notin V_1$. (That is, a total mass of δ is removed uniformly from nodes not in V_1, and added uniformly over the nodes in V_1.)

In moving from x^* to x, the interaction mass of each node $v \in V_1$ changes by $\alpha \cdot \frac{1}{s} \cdot \delta - d \cdot \frac{1}{s(k-1)} \cdot \delta > 0$, while the interaction mass of each node $u \notin V_1$ changes by

$$
\begin{aligned}
- \alpha \cdot \frac{1}{s(k-1)} \cdot \delta - \frac{d(k-2)}{k-1} \cdot \frac{1}{s(k-1)} \cdot \delta + \frac{d}{k-1} \cdot \frac{1}{s} \cdot \delta \\
= \Big(- (d+1) - \frac{d(k-2)}{k-1} + d \Big) \cdot \frac{1}{s(k-1)} \cdot \delta \;<\; 0.
\end{aligned}
$$

Thus, for any such vector $x(t) = x$, all flows are directed from nodes not in V_1 to nodes in V_1. Furthermore, by symmetry of the original vector x^* and the perturbation, the new mass vector $x(t+1)$ will be of the same form, for a different $\delta' > \delta$. Thus, the same argument will inductively apply in every time step. Hence, the direction of flows never changes, and Lemma 4.2.2 guarantees that the system converges. Since the interaction mass of all nodes in V_1 is only increasing, and the interaction mass of all nodes not in V_1 is only decreasing, the only equilibrium y the system can converge to is one in which all nodes outside of V_1 have zero mass.

In particular, this means that even starting from $||x^* - x|| = \delta'$ (which would correspond to using $\delta = \delta'/2$ in our analysis), $||x^* - y||$ is not bounded in terms of δ', so x^* is not stable. □

Lemma 4.4.10 considerably narrows down the set of equilibria for which the question of whether or not they are universally stable remains open.

More specifically, it only remains to consider mass vectors x in which there is a non-trivial connected C of $\mathcal{T}_{\texttt{act}}(x)$ such that C is a d-regular graph (for some $d \geq 1$), is not a locally balanced graph (in particular not a bipartite graph), and for every $u, v \in C$, $x_v = x_u$. We conjecture that such mass vectors are not universally stable; it would then follow that in any universally stable equilibrium, all components have size 1, and hence by Proposition 4.4.9 the active nodes would be at mutual distance 3. Accordingly, we formulate the following:

Conjecture 4.4.11 *A mass vector is a universally stable equilibrium if and only if its active nodes are at pairwise distance at least* 3 *in the influence graph.*

4.4.4 Characterization of Lyapunov Stable Equilibria for $\alpha = 1$

For $\alpha > 1$, we have shown that any equilibrium whose active nodes form an independent set of pairwise node distance at least 3 is stable. Perhaps surprisingly, this ceases to be true for $\alpha = 1$. Indeed, on the path of length 4, the equilibrium $x^* = (\frac{1}{2}, 0, 0, \frac{1}{2})$ is not stable.

We can see this instability as follows. Consider vectors of the form $x^{(\delta)} =$

$(\frac{1}{2}-\delta, \delta, \delta, \frac{1}{2}-\delta)$. Under $x^{(\delta)}$, for any $\delta \in (0, \frac{1}{2})$, the interaction mass of nodes 1 and 4 is strictly smaller than the interaction mass of nodes 2 and 3 (whose interaction masses are equal). This implies that no vector $x^{(\delta)}$ can be an equilibrium for $\delta \in (0, \frac{1}{2})$, and that flow will always be directed from nodes 1 and 4 to nodes 2 and 3. Furthermore, the flow from node 1 to node 2 is equal to the flow from node 4 to node 3, implying that at the next time step, the mass vector will be of the form $x^{(\delta')}$ with $\delta' > \delta$. As we know by Theorem 4.4.6 that the 4-path always converges, the system converges to some mass vector $y = x^{(\delta^*)}$ such that $\delta^* > 0$. [4] Since the update rule is continuous, this y must be an equilibrium. We have proved that the only such equilibrium is the one with $\delta^* = \frac{1}{2}$. Thus, starting from the perturbation $x^{(\delta)}$ of x^*, the system can only converge to a state y in which $y_1 = y_4 = 0$.

While a pairwise distance of 3 between active nodes is not enough to guarantee stability, a pairwise distance of 4 is sufficient.

Theorem 4.4.12 *Let x^* be a mass vector whose active nodes have pairwise distance at least 4. Then, x^* is a stable equilibrium for $\alpha = 1$.*

Proof: The proof is much more involved than the proof of Proposition 4.2.3, for the following reason: even for arbitrarily small perturbations to x^*, it is possible that inactive neighbors v of an active node u have higher interaction mass; thus, the conditions of Proposition 4.2.3 do not apply, and in fact, u could lose mass over time. However, we will be able to show that the total mass u loses, starting from a perturbation of magnitude at most δ, is bounded by a function $g(\delta) \to 0$ as $\delta \to 0$.

Let A be the set of all x^*-active nodes, and let $x(0)$ be a perturbation of x^*

[4] This also follows directly from our argument with $x^{(\delta)}$.

with $||x^* - x(0)||_1 \leq \delta \leq \frac{1}{8} \cdot \min_{u \in A} x_u^*$. We will show below that for each node $u \in A$, and all times t, we have that $|x_u(t) - x_u^*| \leq 2\delta$. Because

$$\sum_{v \notin A} |x_v(t) - x_v^*| \;=\; \sum_{v \notin A} x_v(t) \;=\; \sum_{u \in A} (x_u^* - x_u(t)) \;\leq\; \sum_{u \in A} |x_u(t) - x_u^*|,$$

we obtain that

$$||x_v(t) - x^*||_1 \;=\; \sum_{u \in V} |x_u(t) - x_u^*| \;\leq\; 2 \sum_{u \in A} |x_u(t) - x_u^*| \;\leq\; 4n\delta \;\rightarrow\; 0 \quad \text{as } \delta \rightarrow 0.$$

It remains to prove the inequality $|x_u(t) - x_u^*| \leq 2\delta$ for all nodes $u \in A$ and times t. Define $W = V \setminus (A \cup \bigcup_{u \in A} T_u)$ to be the set of all nodes at distance at least 2 from all active nodes. We will prove the inequality by showing that any flow from u to its neighbors v can be "charged" against flow from W to v. More formally, we will simultaneously prove the following invariants for all times t and any set $U \subseteq A$:

$$\sum_{w \in W} x_w(t) \leq \sum_{w \in W} x_w(0), \tag{4.9a}$$

$$\sum_{u \in U} x_u(t) \geq \sum_{u \in U} x_u(0) - \sum_{w \in W} (x_w(0) - x_w(t)) \quad \text{for any set } U \subseteq A. \tag{4.9b}$$

Let t be an arbitrary time step, and assume that the invariants hold at time t. We will establish that they still hold at time $t+1$. First, we notice some useful consequences of the invariants, including the desired fact that $|x_u(t) - x_u^*| \leq 2\delta$.

From Inequality (4.9a), we get that $\sum_{w \in W} x_w(t) \leq \sum_{w \in W} x_w(0) \leq \sum_{v \notin A} x_v(0) \leq \delta$. Substituting this bound into Inequality (4.9b) with $U = \{u\}$, and using that $x_u(0) \geq x_u^* - \delta$ gives us that $x_u(t) \geq x_u^* - 2\delta$. Similarly, using Inequality (4.9b) with $U = A \setminus \{u\}$ gives us an upper bound of $x_u(t) \leq x_u^* + 2\delta$. So we have shown that $|x_u(t) - x_u^*| \leq 2\delta$. It remains to establish the invariants for time $t+1$.

Let $(w, v), w \in W, v \notin W$ be an arbitrary edge, and u the unique active neighbor of v in $\mathcal{T}$. The flow on the edge (w, v) is $f_{w\to v}(t) = p\cdot\frac{x_w(t)\, x_v(t)(\mathcal{M}_v(t)-\mathcal{M}_w(t))}{\mathcal{M}_w(t)\, \mathcal{M}_v(t)}$. We have just seen that $x^*_u-2\delta \le x_u(t) \le x^*_u+2\delta$, so we can also bound $\mathcal{M}_v(t) \ge x^*_u-2\delta$. Applying Inequality (4.9b) with $U = A \setminus \{u\}$ also gives us an upper bound of $\mathcal{M}_v(t) \le 1 - \sum_{u'\in A, u'\neq u} x_{u'}(t) \le x^*_u + 2\delta$. Furthermore, using the definition of W and Inequality (4.9b) for $U = A$,

$$\mathcal{M}_w(t) \;\le\; \sum_{v\notin A} x_v(t) \;\le\; \sum_{v\notin A} x_v(0) + \sum_{w\in W} (x_w(0) - x_w(t)) \;\le\; 2\delta.$$

Substituting these bounds, we get that

$$f_{w\to v}(t) \ge p\cdot\frac{x_w(t)\, x_v(t)(x^*_u - 4\delta)}{2\delta(x^*_u + 2\delta)}. \tag{4.10}$$

By definition of δ, this quantity is always non-negative. In particular, this means that flow goes from w to v; since the edge (w, v) was arbitrary, we have established the invariant (4.9a).

Next, fix an arbitrary node pair $u \in A, v \in T_u$. The flow from u to v is

$$\begin{aligned} f_{u\to v}(t) &= p\cdot\frac{x_u(t)\, x_v(t)(\mathcal{M}_v(t) - \mathcal{M}_u(t))}{\mathcal{M}_u(t)\, \mathcal{M}_v(t)} \;\le\; p\cdot\frac{x_u(t)\, x_v(t) \sum_{w\in W\cap T_v} x_w(t)}{(x_u(t))^2} \\ &= p\cdot\frac{1}{x_u(t)} \sum_{w\in W\cap T_v} x_w(t)\, x_v(t) \;\le\; p\cdot\frac{1}{x^*_u - 2\delta} \sum_{w\in W\cap T_v} x_w(t)\, x_v(t). \end{aligned}$$

On the other hand, summing the bound (4.10) over all nodes $w \in W \cap T_v$, we get that

$$\sum_{w\in W\cap T_v} f_{w\to v}(t) \ge p\cdot\frac{x^*_u - 4\delta}{2\delta(x^*_u + 2\delta)} \cdot \sum_{w\in W\cap T_v} x_w(t)\, x_v(t).$$

Because $\delta \le x^*_u/8$, we get that $\frac{1}{x^*_u-2\delta} \le \frac{x^*_u-4\delta}{2\delta(x^*_u+2\delta)}$, so the flow from u to v is at most the total flow from all $w \in W \cap T_v$ to v. For any set $U \subseteq A$, summing this inequality over all $u \in U$ (and noticing that we never double-count the same edge) now establishes Invariant (4.9b) at time $t + 1$. $\square$

4.5 Conclusions and Open Questions

In this chapter, we presented a novel model of cultural dynamics that captures the essential aspect of several previously studied models: the interplay between selection and influence. We concentrated on two instances of this model. In the *global model*, each person selects another person from the entire population to interact with. In the *local model*, a person selects an interaction partner from a subset of the population consisting of similar people.

We provided a nearly complete treatment of the global model, showing that the system always converges from any initial mass vector, and providing a complete characterization of stable equilibria. An open question is to predict the equilibrium to which the system converges starting from a given initial mass vector. We suspect that with probability 1 over possible starting states, the system converges to an equilibrium in which the active nodes form an independent set.

The local model involves, at its heart, a dynamical system on the population fractions that is complicated even for small numbers of variables. As such, it raises many interesting and challenging questions, and we have made progress on some of these. In particular, we know that on paths of length 3 (for $\alpha > 1$) and at most 5 (for $\alpha = 1$), the system converges from any starting state. However, it is open whether convergence occurs for all graphs. On the stability frontier, for $\alpha > 1$, we conjecture that the only stable equilibria are those in which the active nodes have pairwise distance at least 3. We showed that such equilibria are indeed stable, and that a number of other equilibria are unstable — including ones in which the active nodes form any other independent set, or ones in which they form a *locally balanced graph* (a class that includes bipartite graphs). Finally, we would like to raise an even more challenging question: does the dynamical system defined by the general model always converge?

CHAPTER 5

SCIENTIFIC CREDIT ALLOCATION

5.1 Introduction

In this chapter and the next one, we turn to the second theme of this thesis: exploring the ways competition and credit create incentives. We will start our exploration with understanding credit allocation in the domain of science.

As a scientific community makes progress on its research questions, it also develops conventions for allocating *credit* to its members. Scientific credit comes in many forms; it includes explicit markers such as prizes, appointments to high-status positions, and publication in prestigious venues, but it also builds upon a broader base of informal reputational measures and standing within the community [19, 63, 71]. The mechanisms by which scientific credit is allocated have long been the subject of fascination among scientists, as well as a topic of research for scholars in the philosophy and sociology of science. A common theme in this line of inquiry has been the fundamental ways in which credit seems to be systematically *misallocated* by scientific communities over time — or at least allocated in ways that seem to violate certain intuitive notions of "fairness." Two categories of misallocation in particular stand out, as follows.

1. *Certain research questions receive an "unfair" amount of credit.* In other words, a community will often have certain questions on which progress is heavily rewarded, even when there is general agreement that other questions are equally important. Such issues, for example, have been at the heart of recent debates within the theoretical computer science community, focusing

on the question of whether conference program committees tend to overvalue progress on questions that display "technical difficulty" [1, 48].

2. *Certain people receive an "unfair" amount of credit.* Robert Merton's celebrated formulation of the *Matthew Effect* asserts, roughly, that if two (or more) scientists independently or jointly discover an important result, then the more famous one receives a disproportionate share of the credit, even if their contributions were equivalent [70, 71].[1] Other attributes such as affiliations or academic pedigree can play an analogous role in discriminating among researchers.

There is a wide range of potential explanations for these two phenomena, and many are rooted in hypotheses about human cognitive factors: a fascination with "hard" problems or the use of such problems to identify talented problem-solvers in the first case; the effect of famous individuals as focal points or the confidence imparted by endorsement from a famous individual in the second case [70, 99].

A model of competition and credit in science

One can read this state of affairs as a story of how fundamental human biases lead to inherent unfairness, but we argue in this chapter that it is useful to bring into the discussion an alternate interpretation, via a natural formal model for the process by which scientists choose problems and by which credit is allocated.

We begin by adapting a model proposed in influential work of Kitcher in the philosophy of science [62, 63, 98], and with roots in earlier work of Peirce, Arrow,

[1] This is a kind of rich-get-richer phenomenon, and Merton's use of the term "Matthew Effect" is derived from *Matthew 25:29* in the New Testament of the Bible, which says, "For unto every one that hath shall be given, and he shall have abundance: but from him that hath not shall be taken away even that which he hath."

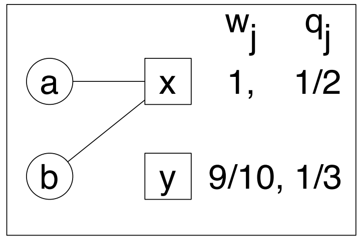
and Bourdieu [9, 21, 83]. Kitcher's model has some slightly complicated features that we do not need for our purposes, so we will focus the discussion in terms of the following closely related model; it is designed as a stylized abstraction of a community of n researchers who each choose independently among a set of m open problems to work on.

- The m open problems will also be referred to as *projects*. Each project j has an *importance* w_j (also called its *weight*), and a probability of success q_j (with a corresponding failure probability $f_j = 1 - q_j$). We assume these numbers are rational. The researchers will initially be modeled as identical, but we later consider generalizations to individuals with different problem-solving abilities.
- Each researcher must choose a single project to work on. We model researchers as working independently, so if k_j researchers work on project j, there is a probability of $(1 - f_j^{k_j})$ that at least one of them succeeds.
- In the event that multiple researchers succeed at project j, one of them is chosen uniformly at random to receive an amount of credit equal to the project's importance w_j. We can imagine there is a "race" to be the first to solve the problem, and the credit goes to the "winner" ; alternately, we get the same model if we imagine that all successful researchers divide the credit equally.

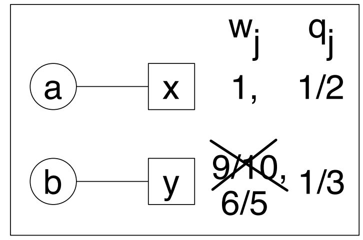
Suppose that researchers are motivated by the amount of credit they receive: each researcher chooses a project to work on to maximize her expected amount of credit, given the choices of all other researchers. The selection of projects is thus a game, in which the players are the researchers, the strategies are the choices of projects, and the payoffs are the expected amount of credit received. This game-theoretic view forms the basis of Kitcher's model of scientific competition; the view itself was perhaps first articulated explicitly in this form by the social scientist Pierre Bourdieu [19, 21], who wrote that researchers' motivations

> are organized by reference to – conscious or unconscious – anticipation of the average chances of profit ... Thus researchers' tendency [is] to concentrate on those problems regarded as the most important ones ... The intense competition which is then triggered off is likely to bring about a fall in average rates of symbolic profit, and hence the departure of a fraction of researchers towards other objects which are less prestigious but around which the competition is less intense, so that they offer profits of at least as great.

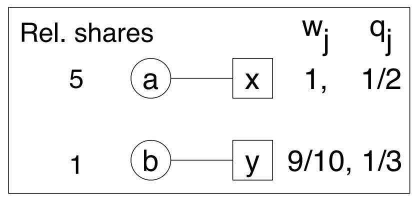
Like the frameworks of Bourdieu and Kitcher, our model is a highly simplified version of the actual process of selecting research projects and competing for credit. We are focusing on projects that can be represented as problems to be solved; we are not modeling the process of collaboration among researchers, the ways in which problems build on each other, or the ways in which new problems arise; and we are not trying to capture the multiple ways in which one can measure the importance or difficulty of a problem. These are all interesting extensions, but our point is to identify a tractable model that contains the fundamental ingredients in our discussion: a competition for credit among projects of varying difficulty, in a way that causes credit-seeking individuals to distribute themselves across different projects. We will see how phenomena that are complex but intuitively familiar can arise even when a community has a single, universally agreed-upon measure of importance and difficulty across projects.



(a) *Projects with original weights.*



(b) *Projects with modified weights.*



(c) *Players with modified credit shares.*

Figure 5.1: In (a), self-interested players do not reach a socially optimal selection of projects. However, if the weight of project y is increased (b), or if one of the players is guaranteed a sufficiently disproportionate share of the credit in the event of joint success (c), then a socially optimal assignment of players to projects arises.

Credit as a mechanism for allocating effort

Our main focus is to extend this class of models to consider the issues raised at the outset of the chapter, and in particular to the two sources of "unfairness" discussed there. The model we have described thus far is based on an intuitively fair allocation of credit that does not suffer from either of these two pathologies: all researchers are treated identically, and the credit a successful researcher receives is equal to the community's agreed-upon measure of the importance of the problem solved. In other words, no problems are overvalued relative to their true importance, and no researchers are *a priori* favored in the assignment of credit.

As a first thought experiment, suppose that we were allowed to design the rules by which credit was assigned in a research community; are these "fair" rules the ones we should use? The following very small example shows the difficulties we quickly run into. Suppose, for simplicity, that we are dealing with a community consisting of two players a and b, and two projects x and y. Project x is more important and also easier; it has $w_x = 1$ and $q_x = 1/2$. Project y is less important and more difficult; it has $w_y = 9/10$ and $q_y = 1/3$. Figure 5.1(a) shows the unique Nash equilibrium for this research community: both players work on x, each receiving an expected payoff $3/8$ (since project x will be solved with probability $3/4$, and a and b are equally likely to receive credit for it.)

If we were in charge of this research community, arguably the natural objective function for us to care about would be the *social welfare*, defined as the total expected importance of all projects successfully completed. And now here's the difficulty: the unique Nash equilibrium does not maximize social welfare. It produces a social welfare of $3/4$, whereas if the players divided up over the two different projects, we would obtain a social welfare of $1/2 + 3/10 = 4/5$.

Can we change the way credit is assigned so as to create incentives for the players under which the resulting Nash equilibrium maximizes social welfare? In fact, there are two natural ways to do this, and each should be recognizable given the discussion at the beginning of the introduction.

- First, we could declare that the credit received for succeeding at a project will not be proportional to its importance. Instead, in our example, we could decide that success at the harder project y will bring an amount of credit equal to $w'_y \neq w_y$. If $9/5 > w'_y > 9/8$, then the unique Nash equilibrium is socially optimal (Figure 5.1(b)).

- Alternately we could declare that if players a and b both succeed at the same project, they will not split the credit equally, but instead in a ratio of c to 1. (Equivalently, if they both succeed, player a is selected to receive all the credit with probability $c/(c+1)$ and player b with probability $1/(c+1)$.) If $c > 4$, then it is not worth it for b to try competing with a on project x, and b will instead work on project y, again leading to a socially optimal Nash equilibrium (Figure 5.1(c)).

This example highlights several points. First, we can think of the amount of credit associated with different projects as something malleable; by choosing to have certain projects confer more credit, the community can create incentives that cause effort to be allocated in different ways. Second, it is clearly the case that actual research communities engage in this shaping of credit, not just at an implicit level but through a variety of explicit mechanisms: the decisions of program committees and editorial boards about which papers to accept, the decisions of hiring committees about which people to interview and areas to recruit in, and the decisions of granting agencies about funding priorities all serve to shift the amounts of credit assigned to different kinds of activities. In this sense, a research community is, to a certain extent, a kind of "planned economy" — it is much more complex than our simple model, but many of its central institutions have the effect of deliberately implementing and publicizing decisions about the allocation of credit for different kinds of research topics.

What we see in the example is that the "fair" allocation of credit can be at odds with the goal of social optimality: if the community believes that as a whole it is being evaluated according to the total expected weight of successful projects, then by rewarding its participants according to these same weights, it produces a socially sub-optimal outcome. The two alternate ways of assigning credit above correspond to the two forms of "unfairness" discussed at the outset: overvaluing certain projects (in our example, the harder and less important project), and overvaluing the contributions of certain researchers. If done appropriately in this example, either of these can be used to achieve social optimality.

As a final point on the underlying motivation, we are not claiming that research communities are overtly trying to assign credit in a way that achieves social optimality, or arriving at credit allocations in general through explicit computation. It is clear that the human cognitive biases discussed earlier — in favor of certain topics and certain people — are a large and likely dominant contributor to this. What we do see, however, is that social optimality plays an important and surprisingly subtle role in the discussion about these issues: institutions such as program committees and funding agencies do take into account the goal of shaping the kind of research that gets done, and to the extent that these cognitive biases can sometimes — paradoxically — raise the overall productivity of the community, it arguably makes such biases particularly hard to eliminate from people's behavior.

Social optimality and misallocation of credit: General results

Our main results begin by establishing that the two kinds of mechanisms suggested by the example in Figure 5.1 are each sufficient to ensure social optimality in general — that is, in all instances. For any set of projects, it is possible to assign each project j a *modified weight* w'_j, potentially different from its real weight w_j, so that when players receive credit according to these modified weights, all Nash equilibria are socially optimal with respect to the real weights. It is also possible to assign each player i a weight z_i so that when players divide credit on successful projects in proportion to their weights z_i, all Nash equilibria are again socially optimal.

This makes precise the sense in which our two categories of credit misallocation can both be used to optimize social welfare.

These results in fact hold in a generalization of the basic model, in which the players are heterogeneous and have different levels of *ability* at solving problems. In this more general model, a player's success at a project depends on both her ability and the project's difficulty: each player i has a parameter $p_i \leq 1$ such that her probability of succeeding at project j is equal to the product $p_i q_j$. Beyond this, the remaining aspects of the model remain the same; in particular, if multiple players all succeed at the same project, then one is selected uniformly at random to receive the credit. (That is, their ability affects their chance of succeeding, but not their share of the credit.) For this more general game, there still always exist re-weightings of projects and also re-weightings of credit shares to players that lead to socially optimal Nash equilibria.

Our results make use of the fact that the underlying game, even in its more general form with heterogeneous players, is both a *congestion game* [72, 89] and a *monotone valid-utility game* [46, 102, 104]. However, given the motivating setting for our analysis, we have the ability to modify certain parameters of the game — as part of a research community's mechanism for allocating credit — that are not normally under the control of the modeler. As a result, our focus is on somewhat different questions, motivated by these credit allocation schemes. At the same time, there are interesting analogies to issues in congestion games from other settings. Re-weighting the amount of credit on projects can be viewed as a kind of "toll" system, interpreting the effort of the researchers as the "traffic" in the congestion game. The crux of our analysis for re-weighting the players is to begin by considering an alternate model in which an ordering is defined on the players, and the first player in this ordering to succeed receives all the credit. This suggests interesting potential connections with the theory of *priority algorithms* introduced by Borodin et al. [20]; although the context is quite different, we too are asking whether there is a "greedy ordering" that leads to optimality. A related set of questions was considered by Strevens in her model of sequential progress on a research problem, working within Kitcher's model of scientific competition [98].

We also consider some of the structural aspects of the underlying game; among other results, we show that the price of anarchy of the game is always strictly less than 2 (compared with a general upper bound of 2, which can sometimes be attained, for fully general monotone valid-utility games). For the case of identical players, we also show that the ratio of the price of anarchy to the price of stability (i.e. the welfare of the best Nash equilibrium relative to the worst) is at most 3/2. In particular, this implies that when there exists a Nash equilibrium that is optimal, there is no Nash equilibrium that is less than 2/3 times optimal.

Finally, we consider a still more general model, in which player success probabilities are arbitrary and unrelated: player i has a probability p_{ij} of succeeding on project j. We show that there exist instances of this general game in which no re-weighting of the projects yields a social optimal Nash equilibrium. However we do not rule out the possibility that there exists a re-weighting of the players that yields a socially optimal Nash equilibrium.

Interpreting the model

With any simple theoretical model of a social process — in this case, credit among researchers — it is important to ask whether the overall behavior of the model captures fundamental qualitative aspects of the real system's behavior. In this case we argue that it captures several important phenomena at a broad level. First, it is based on the idea that institutions within a research community can and do shift the amount of credit that different research topics receive, and in a number of cases with the goal of creating corresponding incentives toward certain research directions. Second, it argues that some of the typical ways in which credit is misallocated can interact in a complex fashion with social welfare, and that these misallocations can in fact play an important role in the maximization of welfare.

Moreover, there is a rapidly widening scope for the potential application of explicitly computational approaches to credit-allocation, as we see an increasing number of intentionally *designed* systems aimed at fostering massive Internet based-collaboration — these include large open-source projects, collaborative knowledge resources like Wikipedia, and collective problem-solving experiments such as the Polymath project [52]. For example, a number of credit-allocation conventions familiar from the scientific community have been built into Wikipedia, including the ways in which editors compete to have articles "featured" on the front page of the site [97], and the ways in which they go through internal review and promotion processes to achieve greater levels of status and responsibility [26, 68]. While the framework in this chapter is only an initial foray in this direction, the general issue of designing credit-allocation schemes to optimize collective productivity becomes an increasingly wide-ranging question.

There are also potential connections to work in an area termed the *economics of science*, which studies the allocation of resources by organizations across different research projects [19, 91], in the context of activities such as R&D (e.g. [33]). While the central issues in these models are somewhat different, connecting them more closely to the questions raised here is an interesting question.

Finally, the model offers a set of simple and, in the end, intuitively natural interpretations for the specific ways in which misallocation can lead to greater collective productivity. The re-weighting of projects not only follows the informal roadmap contained in Pierre Bourdieu's quote above, but sharpens it. Even without re-weighting of projects, the effect of competition does work to disperse some number of researchers out to harder and/or less attractive projects, which helps push the system toward states of higher social welfare. But the point is that this dispersion is not optimally balanced on its own; it needs to be helped along, and this is where the re-weighting of projects comes into play. The re-weighting of players is based on a different point — that when certain individuals are unfairly marginalized by a community, it can force them to embark on higher-risk courses of action, enabling beneficial innovation that would otherwise not have happened. In all these cases, it does not mean that such forms of misallocation are fair to the participants in the community, only that they can sometimes have the effect of increasing the community's overall output.

5.2 Identical Players

We first consider the case of the *project game* defined in the introduction when all players are identical, and then move on to the case in which players have different levels of ability. Recall that w_j denotes the weight (i.e. importance) of project j, and f_j denotes the probability that any individual player fails to succeed at it. Thus, when there are k players working on project j, the contribution of project j to the social welfare is $w_j(1 - f_j^k)$, and we denote this quantity by $\sigma_j(k)$.

We denote the choices of all players by a *strategy vector* $\vec{a}$, in which player

i chooses to work on project a_i. As is standard, we denote by a_{-i} the strategy vector $\vec{a}$ without the i^{th} coordinate and by j, a_{-i} the strategy vector $a_1, \ldots, a_{i-1}, j, a_{i+1}, \ldots, a_n$. We use $K_j(\vec{a})$ to denote the set of players working on project j in strategy vector $\vec{a}$, and we write $k_j(\vec{a}) = |K_j(\vec{a})|$. The social welfare obtained from strategy vector $\vec{a}$ is $u(\vec{a}) = \sum_{j \in M} \sigma_j(k_j(\vec{a}))$. Since each of the players working on the project is equally likely to receive the credit, the payoff, or utility, of player i under strategy vector $\vec{a}$ is $u_i(\vec{a}) = \dfrac{\sigma_{a_i}(k_{a_i}(\vec{a}))}{k_{a_i}(\vec{a})}$.

We make a few observations about these quantities. First, as noted in the introduction, $u_i(\vec{a})$ is the utility of i regardless of whether we interpret the credit as being assigned uniformly at random to one successful player on a project, or divided evenly over all successful players.

Moreover, since the players divide up the social welfare among themselves, we have $\sum_{i \in N} u_i(\vec{a}) = u(\vec{a})$. Since a player's utility depends solely on the number of other players choosing her project, it is not hard to verify that the game with identical players is a congestion game, and hence has pure Nash equilibria. Finally, we define the welfare improvement from increasing the number of players working on project j by 1; we denote this improvement by $r_j(k) = (1 - f_j)f_j^k$, where k is the number of players currently working on project j. The fact that $r_j(k)$ is decreasing in k would become useful later on.

We begin by developing some basic properties of the social optimum and of the set of Nash equilibria with identical players; we then build on this to prove bounds on the price of anarchy (the ratio of the social welfare of the social optimum to the worst Nash equilibrium) and the price of stability (the analogous ratio of the social optimum to the best Nash equilibrium). After this, we provide algorithms for re-weighting projects and re-weighting players so as to produce Nash equilibria that are socially optimal.

Before proceeding, we first state and prove three basic claims about the game with identical players. We begin by showing that the project game is a monotone valid-utility game.

Claim 5.2.1 *The project game with identical players is a monotone valid-utility game.*

Proof: We need a bit of additional notation: the quantity $u(a_i|a_{-i})$ will denote the marginal contribution of player i to the overall utility, given the choices made by all other players. By definition, this means that $u(a_i|a_{-i}) = r_{a_i}(k_{a_i}(\vec{a}) - 1)$.

The definition of a monotone valid-utility game [102, 104] requires verifying four properties of the utility functions, as follows.

1. $u(\vec{a})$ *is submodular:* Since $u(\vec{a})$ is the summation of the projects' separate utilities, it is enough to prove that the utility of every project is submodular. For identical players this is settled by the simple observation that $r_j(k)$ is decreasing in k.

2. $u(\vec{a})$ *is monotone:* Naturally, a project's success probability can only increase when more players are working on it, thus, the utility is monotone increasing.

3. $u_i(\vec{a}) \geq u(a_i|a_{-i})$: For this, we notice that $\sigma_j(k_j(\vec{a}))$ can be written as the sum of the marginal utilities contributed by the players on project j when they arrive to it in any order, and a player's utility is the average of such contributions over all arrival orders. Since the utility is submodular, the smallest of these contributions occurs when i arrives last. In this case it is equal to $u(a_i|a_{-i})$. Hence this quantity is at most $u_i(\vec{a})$ as required.

4. $u(\vec{a}) \geq \sum_i u_i(\vec{a})$: In this game, by the definition of a player's utility, they are equal.

□

Next, we show that the optimal solution can be computed in polynomial time by simple greedy algorithm.

Claim 5.2.2 *The optimal assignment can be computed by the following greedy algorithm: players are assigned to projects one at a time, and in each iteration a player is assigned to a project j with the greatest current marginal utility $r_j(k_j)$.*

Proof: Assume towards a contradiction that the assignment resulting from the greedy algorithm $\vec{a}$ is sub-optimal. Let $\vec{o}$ be an optimal assignment which is the most similar to $\vec{a}$ (i.e., minimizes $\sum_j |k_j(\vec{a}) - k_j(\vec{o})|$). Since assignments are insensitive to the identity of the players and $u(\vec{o}) > u(\vec{a})$, there exist two projects b and c such that:

- $k_b(\vec{a}) > k_b(\vec{o})$
- $k_c(\vec{a}) < k_c(\vec{o})$

Let i be the last player the algorithm assigned to project b. In the iteration when i was assigned to project b there were at most $k_c(\vec{a})$ players working on project c. As we noted at the beginning of Section 5.2, the function $r_j(k)$ is decreasing in k for all projects j, and hence $r_b(k_b(\vec{a})-1) \geq r_c(k_c(\vec{a}))$. By this decreasing property we also have that $r_b(k_b(\vec{o})) \geq r_b(k_b(\vec{a}) - 1)$ and that $r_c(k_c(\vec{a})) \geq r_c(k_c(\vec{o}) - 1)$. Because $\vec{o}$ is the optimal assignment then $r_b(k_b(\vec{o})) \leq r_c(k_c(\vec{o})-1)$. Hence we have that $r_b(k_b(\vec{o})) = r_c(k_c(\vec{o}) - 1)$. Therefore, we can get another optimal solution, $\vec{o}_1$ by starting with $\vec{o}$ and moving a single player from project c to project b. We can now reach a contradiction by observing that by construction $\vec{o}_1$ is more similar to $\vec{a}$ than $\vec{o}$ is.

□

We now show that a simple greedy algorithm can be also used from computing an optimal solution. A similar claim was also proved in [41] in the context of a related class of congestion games.

Claim 5.2.3 *A Nash equilibrium can be computed in polynomial time by the following algorithm: players choose projects one at a time in an arbitrary order, and in each iteration the current player i chooses a project that maximizes her utility in respect to the choices made by earlier players.*

Proof: Denote the assignment the algorithm computes by $\vec{a}$. Assume towards a contradiction that player i, who is currently assigned to project j, can increase her payoff by switching to project l: $u_i(j, a_{-i}) < u_i(l, a_{-i})$. Since all the players are identical we can assume without loss of generality that i was the last player who chose project j. Denote by $\vec{a'}$ the assignment vector at the iteration at which it was player i's turn to choose a project. Since player i chose project j we have that $u_i(j, a'_{-i}) \geq u_i(l, a'_{-i})$. Noticing that $k_l(\vec{a}) \geq k_l(\vec{a'})$ and that the utility function is submodular, we obtain a contradiction. □

We now prove that with identical players, any two Nash equilibria are very similar in their assignment of players to projects[2]:

[2]A similar claim was independently proven for a related class of congestion games in unpublished work of Kuniavsky and Smorodinsky.

Claim 5.2.4 *For every two different Nash equilibria $\vec{a}$ and $\vec{b}$ and for every two projects j,l such that $k_j(\vec{a}) > k_j(\vec{b})$ and $k_l(\vec{a}) < k_l(\vec{b})$, we have the following relationships: $k_j(\vec{a}) = k_j(\vec{b}) + 1$ and $k_l(\vec{b}) = k_l(\vec{a}) + 1$.*

Proof: As assignments are insensitive to the identities of the players, there exists a player i such that $a_i = j$ and $b_i = l$. Since $\vec{a}$ is a Nash equilibrium we have $u_i(j, a_{-i}) \geq u_i(l, a_{-i})$. On the other hand, $\vec{b}$ is also a Nash equilibrium, and hence $u_i(j, b_{-i}) \leq u_i(l, b_{-i})$. Recall that j and l are two projects such that $k_j(\vec{a}) > k_j(\vec{b})$ and $k_l(\vec{a}) < k_l(\vec{b})$. Therefore, because a player's utility is decreasing in the number of players working on the project, we have

$$u_i(j, a_{-i}) \leq u_i(j, b_{-i}) \leq u_i(l, b_{-i}) \leq u_i(l, a_{-i}) \leq u_i(j, a_{-i}).$$

Therefore, $u_i(j, a_{-i}) = u_i(j, b_{-i})$. This implies that $k_j(\vec{a}) = k_j(j, a_{-i}) = k_j(j, b_{-i}) = k_j(\vec{b}) + 1$. Similarly for $u_i(l, b_{-i}) = u_i(l, a_{-i})$ we have that $k_l(\vec{b}) = k_l(l, b_{-i}) = k_l(l, a_{-i}) = k_l(\vec{a}) + 1$. $\square$

5.2.1 The Price of Anarchy and Price of Stability

Observe that the price of anarchy (PoA) of the project game is at most 2. This is because our game is a monotone valid-utility game (Claim 5.2.1) and Vetta [104] showed that the price of anarchy (PoA) of monotone valid-utility games is at most 2. Here we provide a strengthened analysis of the price of anarchy that yields several consequences, all do not hold for monotone valid-utility games in general:

(i) a bound of $1 + \frac{c-1}{c}$ on the PoA for instances in which the worst Nash equilibrium has at most c players assigned to any single project.

(ii) as a corollary of (i), a general upper bound of $2 - \frac{1}{n}$ on the PoA for any instance.

(iii) a bound of $\frac{3}{2}$ between the price of anarchy and the price of stability (PoS) for any instance.

We first show that these bounds are tight, by means of the following example. Consider an instance with n players and n projects; all projects are guaranteed to succeed (i.e. $q_j = 1$ for all j); and the weights of the projects are defined so that $w_1 = 1$ and $w_j = 1/n$ for $j > 1$. The socially optimal assignment of players to projects in this game is for each player to work on a different project, yielding a social welfare of $2 - \frac{1}{n}$. On the other hand, it is a Nash equilibrium for all players to work on project 1, yielding a social welfare of 1. Furthermore, in the case of this example when $n = 2$, the social optimum is also a Nash equilibrium, establishing a gap of $\frac{3}{2}$ between the PoA and PoS in this case. (We also note that for general n, if we increase the weight of project 1 by arbitrarily little, then we obtain an example in which the PoS is arbitrarily close to $2 - \frac{1}{n}$.)

To prove the upper bounds in (i)-(iii), we use Roughgarden's notion of *smoothness* [90].

Definition 5.2.5 *A monotone valid-utility game is (λ, μ)-smooth if for every two strategy vectors $\vec{a}$ and $\vec{b}$, we have $\sum_{i \in N} u_i(b_i, a_{-i}) \geq \lambda u(\vec{b}) - \mu u(\vec{a})$.*

The following is a useful claim based on Roughgarden's paper:

Claim 5.2.6 *If a monotone valid-utility game is (λ, μ)-smooth then its price of anarchy is at most $\dfrac{1+\mu}{\lambda}$.*

Proof: We show that a stronger claim holds: for every Nash equilibrium $\vec{a}$ and every strategy vector $u(\vec{b}) \leq \frac{1+\mu}{\lambda} u(\vec{a})$. This, implies a bound on the PoA by taking $\vec{a}$ to be the worst Nash equilibrium and $\vec{b}$ an optimal solution. To see why the stronger claim holds first observe that since $\vec{a}$ is a Nash equilibrium then $\sum_{i \in N} u_i(b_i, a_{-i}) \leq \sum_{i \in N} u_i(\vec{a})$. Recall also that $\sum_{i \in N} u_i(\vec{a}) = u(\vec{a})$. Hence we have that $u(\vec{a}) \geq \lambda u(\vec{b}) - \mu u(\vec{a})$. By rearranging the terms we get that $u(\vec{b}) \leq \frac{1+\mu}{\lambda} u(\vec{a})$ as required. □

Theorem 5.2.7 *The project game with identical players is* (λ, μ)*-smooth for* $\lambda = 1$ *and*

$$\mu = \max_{\{l | \ k_l(\vec{a}) > k_l(\vec{b}) \geq 1\}} \frac{k_l(\vec{a}) - k_l(\vec{b})}{k_l(\vec{a}) - k_l(\vec{b}) + 1}.$$

Before proving the theorem we assert that (i)-(iii) can indeed be derived from it. First observe that by Claim 5.2.6 an instance maximizing the PoA is an instance for which the value of μ is maximized. It is not hard to see that if the number of players working on each project is at most c, then $\mu \leq \frac{c-1}{c}$ as $k_l(\vec{a}) \leq c$ and $k_l(\vec{b}) \geq 1$ for a project l maximizing the expression for μ. Thus, By applying Claim 5.2.6 we get consequence (i). Consequence (ii) is obtained by observing that the number of players working on a project is always bounded by n. To obtain consequence (iii), we call a game *weakly-*(λ, μ)*-smooth* provided the (λ, μ)-smoothness condition holds just for all Nash equilibria $\vec{a}$ and $\vec{b}$, rather than all arbitrary strategy vectors. Now, for any two Nash equilibria $\vec{a}$ and $\vec{b}$, Claim 5.2.4 implies that the number of players working on each project in $\vec{a}$ and $\vec{b}$ can differ by at most one. Hence, by Theorem 5.2.7 we have that the project game with identical players is weakly-(λ, μ)-smooth for $\mu = \frac{1}{2}$. We can now apply Claim 5.2.6 with $\vec{a}$ equal to the worst Nash equilibrium and $\vec{b}$ equal to the best Nash equilibria to get that $\dfrac{u(\vec{b})}{u(\vec{a})} \leq \dfrac{3}{2}$.

We are now ready to prove Theorem 5.2.7. For this proof it will be useful to require a stronger condition then the one in Definition 5.2.5, which is that the (λ, μ)-smoothness condition: $\sum_{i \in N} u_i(b_i, a_{-i}) \geq \lambda u(\vec{b}) - \mu u(\vec{a})$ will hold separately for each project – there exist λ and μ such that, for every strategy vectors $\vec{a}$ and $\vec{b}$:

$$\forall j \in M \quad \sum_{i \in K_j(\vec{b})} u_i(b_i, a_{-i}) \geq \lambda \sigma_j(k_j(\vec{b})) - \mu \sigma_j(k_j(\vec{a})). \tag{5.1}$$

The next two claims compare two strategy vectors $\vec{a}$ and $\vec{b}$ with respect to a given project j in two cases: when $k_j(\vec{a}) > k_j(\vec{b}) > 0$ (Claim 5.2.8) and when $k_j(\vec{a}) \leq k_j(\vec{b})$ (Claim 5.2.9). More specifically, they establish the following:

- If $k_j(\vec{a}) > k_j(\vec{b}) > 0$ then $\sum_{i \in K_j(\vec{b})} u_i(b_i, a_{-i}) \geq \sigma_j(k_j(\vec{b})) - \frac{k_j(\vec{a}) - k_j(\vec{b})}{k_j(\vec{a}) - k_j(\vec{b}) + 1} \sigma_j(k_j(\vec{a}))$.
- If $k_j(\vec{a}) \leq k_j(\vec{b})$ then $\sum_{i \in K_j(\vec{b})} u_i(b_i, a_{-i}) \geq \sigma_j(k_j(\vec{b}))$.

This completes the proof of Theorem 5.2.7 as we have that Condition 5.1 holds for the project game with identical players for $\lambda = 1$, $\mu = \max_{\{l \mid k_l(\vec{a}) > k_l(\vec{b}) \geq 1\}} \frac{k_l(\vec{a}) - k_l(\vec{b})}{k_l(\vec{a}) - k_l(\vec{b}) + 1}$ and hence the game is (λ, μ)-smooth for the appropriate values for Theorem 5.2.7.

Claim 5.2.8 *If* $k_j(\vec{a}) > k_j(\vec{b}) \geq 1$ *then* $\sum_{i \in K_j(\vec{b})} u_i(b_i, a_{-i}) \geq \sigma_j(k_j(\vec{b})) - \frac{k_j(\vec{a}) - k_j(\vec{b})}{k_j(\vec{a}) - k_j(\vec{b}) + 1} \sigma_j(k_j(\vec{a}))$.

Proof: Since, all players that work on project j in $\vec{b}$ also work on it in $\vec{a}$, we have that $\forall i \in K_j(\vec{b}) : \ u_i(b_i, a_{-i}) = u_i(\vec{a}) = \dfrac{\sigma_j(k_j(\vec{a}))}{k_j(\vec{a})}$. Thus, $\sum_{i \in K_j(\vec{b})} u_i(b_i, a_{-i}) =$

$k_j(\vec{b}) \cdot \dfrac{\sigma_j(k_j(\vec{a}))}{k_j(\vec{a})}$. Hence, we need to show that

$$\frac{k_j(\vec{b})}{k_j(\vec{a})}\sigma_j(k_j(\vec{a})) + \frac{k_j(\vec{a}) - K_j(\vec{b})}{k_j(\vec{a}) - k_j(\vec{b}) + 1}\sigma_j(k_j(\vec{a})) \geq \sigma_j(k_j(\vec{b})).$$

Since $\sigma_j(\cdot)$ is monotone, we have that $\sigma_j(k_j(\vec{a})) \geq \sigma_j(k_j(\vec{b}))$. Therefore, it is enough to show that for proving the claim

$$\frac{k_j(\vec{b})}{k_j(\vec{a})} + \frac{k_j(\vec{a}) - k_j(\vec{b})}{k_j(\vec{a}) - k_j(\vec{b}) + 1} \geq 1$$

Lastly, observe that the previous condition holds since $k_j(\vec{b}) > 0$ implies that $\dfrac{k_j(\vec{a}) - k_j(\vec{b})}{k_j(\vec{a}) - k_j(\vec{b}) + 1} \geq \dfrac{k_j(\vec{a}) - k_j(\vec{b})}{k_j(\vec{a})}$ and the claim follows. □

Claim 5.2.9 *If* $k_j(\vec{a}) \leq k_j(\vec{b})$ *then* $\sum_{i \in K_j(\vec{b})} u_i(b_i, a_{-i}) \geq \sigma_j(k_j(\vec{b}))$.

Proof: revised the proof. If $k_j(\vec{a}) = k_j(\vec{b})$ then clearly the claim holds. Else, $k_j(\vec{a}) < k_j(\vec{b})$. We first observe that $\forall i \in K_j(\vec{b}) : \ u_i(b_i, a_{-i}) \geq \dfrac{\sigma_j(k_j(\vec{a})+1)}{k_j(\vec{a})+1}$, since the utility of all player working on project j both in $\vec{a}$ and in $\vec{b}$ is $\dfrac{\sigma_j(k_j(\vec{a}))}{k_j(\vec{a})}$ and the utility of the rest of the players is $\dfrac{\sigma_j(k_j(\vec{a})+1)}{k_j(\vec{a})+1}$. Next, observe that $\dfrac{\sigma_j(k_j(\vec{a})+1)}{k_j(\vec{a})+1} \geq \dfrac{\sigma_j(k_j(\vec{b}))}{k_j(\vec{a})}$ since $\sigma_j(\cdot)$ is monotone and by assumption $k_j(\vec{a}) + 1 \leq k_j(\vec{b})$.

□

5.2.2 Re-weighting Projects to Achieve Social Optimality

We now describe a mechanism for re-weighting projects so as to achieve social optimality. As discussed in the introduction, we show that it is possible to assign new weights $\{w'_j\}$ to the projects so that when utilities are allocated according to these new weights, all Nash equilibria are socially optimal. Note that the re-weighting of projects only affects players' utilities, not the social welfare, as the latter is still computed using the true weights $\{w_j\}$.

The idea is to choose weights so that when players are assigned according to the social optimum, they all receive identical utilities. The following re-weighting accomplishes this: we compute a socially optimal assignment $\vec{o}$, and define $w'_j = \dfrac{k_j(\vec{o})}{(1 - f_j^{k_j(\vec{o})})}$ for $k_j(\vec{o}) > 0$ and $w'_j = 0$ otherwise.

Theorem 5.2.10 *With these weights, all Nash equilibria achieve the social welfare of assignment $\vec{o}$.*

Proof: We first show that $\vec{o}$ is a Nash equilibrium with these weights. Denote the utility of each player in $\vec{o}$ by x: that is, for every player i, we have $u_i(\vec{o}) = x$. We also have that for every project $j \neq o_i$ $u_i(j, o_{-i}) < x$. This holds for each j because either $k_j(\vec{o}) = 0$, in which case $w'_j = 0$, or else $k_j(\vec{o}) > 0$, in which case there are already players assigned to j, and for such projects j a player's utility function is strictly decreasing in the number of players working on j. Therefore, $\vec{o}$ is a Nash equilibrium.

Furthermore we also show that all Nash equilibria assign to every project j exactly $k_j(\vec{o})$ players. As a corollary of the proof of Claim 5.2.4 we have that if there exist Nash equilibria $\vec{o}$ and $\vec{a}$ assigning different numbers of players to some project, then there exists a player i such that $o_i \neq a_i$ and $u_i(o_i, o_{-i}) = u_i(a_i, o_{-i})$. But this is impossible since we have that $u_i(o_i, o_{-i}) = x$ and $u_i(a_i, o_{-i}) < x$. $\square$

It is interesting to reflect on the qualitative interpretation of these new weights for an instance with n players and a very large set of projects of equal weight and with success probabilities $q_1 \geq q_2 \geq q_3 \geq \cdots$ decreasing to 0. In this case, there will be a largest j^* for which the optimal assignment places any players on j^*, and computational experiments with several natural distributions of $\{q_j\}$ indicate that the number of players assigned to projects increases roughly monotonically toward a maximum approximately near j^*. This means that the credit assigned to projects must increase toward j^*, and then be chosen so as to discourage players from working on projects beyond j^*. Moreover, the value of j^* grows with n, the number of players. Hence we have a situation in which the research community can be viewed, roughly, as establishing the following coarse division of its projects into three categories: "too easy" (receiving relatively little credit), "just right" (near j^*, receiving an amount of credit that encourages extensive competition on these projects), and "too hard" (beyond j^*, receiving an amount of credit designed to dissuade effort on these projects). Moreover, smaller research communities reward easier problems (since j^* is smaller), while larger communities focus their rewards on harder problems.

5.2.3 Re-weighting Players to Achieve Social Optimality

We now discuss the companion to the previous analysis: a mechanism for re-weighting the players to achieve social optimality. Recall that this means we assign each player i a weight z_i, and when a set S of players succeeds at a project j, we choose player $i \in S$ to receive the credit w_j with probability $\frac{z_i}{\sum_{h \in S} z_h}$.

When players are identical, we can base the re-weighting mechanism on the optimality of the greedy algorithm expressed in Claim 5.2.2. That is, if we were to assign an absolute order to the players, and announce the convention that credit would go to the first player in the order to succeed at a project, then the players' simultaneous choices would simulate the greedy algorithm to achieve social optimality: the first player in the announced order would choose a project without regard to the choices of other players; the second player would choose as though the first player would win any direct competition, but without regard to the choices of any other players; and so forth. Now, instead of an order, we need to define weights on the players; but we can approximately simulate the order using sharply decreasing weights in which $z_i = \varepsilon^i$ for an $\varepsilon > 0$ chosen to be sufficiently small. The effect of these sharply decreasing weights is to ensure that a player i gets almost no utility from a project j if a player of higher weight also succeeds at j, and i gets almost all the utility from j if i is the player of highest weight to succeed at j. From this, we can show that each player's utility is roughly what it would be under an order on the players. We prove that we can indeed find such an ε as required.

Theorem 5.2.11 *With $\varepsilon > 0$ sufficiently small and the re-weighting of players defined by $z_i = \varepsilon^i$, all Nash equilibria of the resulting game are socially optimal.*

Even given the informal argument above, the proof is complicated by the fact that, with positive weights on all players, their strategic reasoning is more complex than it would be under an actual ordering. To prove Theorem 5.2.11, we consider the relationship between the actual utilities of the re-weighted players for a given strategy vector $\vec{a}$, denoted $\widetilde{u}_i(\vec{a})$:

$$\widetilde{u}_i(\vec{a}) = w_{a_i} q_{a_i} \sum_{S \subseteq \{K_{a_i}(\vec{a}) - i\}} \frac{z_i}{(\sum_{h \in S} z_h) + z_i} q_{a_i}^{|S|} (1 - q_{a_i})^{k_{a_i}(\vec{a}) - |S| - 1}$$

and their "ideal" utility under the order we are trying to simulate, denoted $\widehat{u}_i(\vec{a})$ — the utility function defined by having the first player in order to succeed at a project receive all the credit, which is formally defined next:

Definition 5.2.12

- $\pi_{<i}(S) = \{h \in S | z_i < z_h\}$ *where* $S \subseteq N$ *– the set of players before player* i*.*
- $\pi_{>i}(S) = \{h \in S | z_i > z_h\}$ *where* $S \subseteq N$ *– the set of players after player* i*.*
- $\widehat{u}_i(\vec{a}) = r_{a_i}(|\pi_{<i}(K_{a_i}(\vec{a}))|)$ *– the marginal contribution of player* i *to the social welfare.*

Recalling that the projects' weights and success probabilities are rational, let d be their common denominator. We first show that if these two different utilities are close enough with respect to d, then our approximate simulation of an order using weights will succeed:

Claim 5.2.13 *If for every player* i*, project* j *and strategy vector* $\vec{a}$ *we have*

$$\widehat{u}_i(j, a_{-i}) - \frac{1}{4d^{n+1}} \leq \widetilde{u}_i(j, a_{-i}) \leq \widehat{u}_i(j, a_{-i}) + \frac{1}{4d^{n+1}},$$

then any Nash equilibrium in the game with the weights $\{z_i\}$ *is also an optimal assignment.*

Proof: Consider some Nash equilibrium $\vec{a}$ in the game with the weights $\{z_i\}$ and let $\vec{o}$ be an optimal assignment that is the result of running the greedy algorithm from Claim 5.2.2 with the same order of player, that shares the longest prefix with $\vec{a}$: $\max_i \forall i' \leq i, a_{i'} = o_{i'}$. Assume towards a contradiction that $\vec{a}$ is not an optimal assignment. By definition, player i is the first player in the order that works on a different project in $\vec{a}$ (denote it by j) and in $\vec{o}$ (denote it by l). Since $\vec{a}$ is a Nash equilibrium with the weights $\{z_i\}$ we have that $\widetilde{u}_i(j, a_{-i}) \geq \widetilde{u}_i(l, a_{-i})$. By applying the Claim's assumption we have that:

$$\widehat{u}_i(j, a_{-i}) + \frac{1}{4d^{n+1}} \geq \widetilde{u}_i(j, a_{-i}) \geq \widetilde{u}_i(l, a_{-i}) \geq \widehat{u}_i(l, a_{-i}) - \frac{1}{4d^{n+1}}.$$

This implies that $\widehat{u}_i(j, a_{-i}) + \frac{1}{2d^{n+1}} \geq \widehat{u}_i(l, a_{-i})$. On the other hand, $\vec{o}$ is an optimal solution that shares the longest prefix with $\vec{a}$. Also, since $\vec{o}$ is the result of a greedy algorithm with the same order, we have that $\widehat{u}_i(l, o_{-i}) > \widehat{u}_i(j, o_{-i})$. Since, $\vec{a}$ and $\vec{o}$ are identical for all players prior to player i in the order, we have that $\widehat{u}_i(l, a_{-i}) > \widehat{u}_i(j, a_{-i})$.

To complete the proof, recall that d is the common denominator of all success probabilities and weights. As both $\widehat{u}_i(j, a_{-i})$ and $\widehat{u}_i(l, a_{-i})$ are products of at most $n+1$ terms of common denominator d, and they are not equal, so they must differ by at least $\frac{1}{d^{n+1}}$.

□

Next, we show that it is possible to choose ε sufficiently small such that for every player i and project j: $\widehat{u}_i(j, a_{-i}) - \frac{1}{4d^{n+1}} \leq \widetilde{u}_i(j, a_{-i}) \leq \widehat{u}_i(j, a_{-i}) + \frac{1}{4d^{n+1}}$ as required by Claim 5.2.13. We begin by presenting the following definitions that will be useful in simplifying the utility function:

Definition 5.2.14 *For a strategy vector $\vec{a}$, a project j and $S \subseteq K_j(a_{-i})$ we define*

$$\psi_i(j; S; \vec{a}) = \frac{z_i}{z_i + \sum_{h \in S} z_h} \cdot q_j^{|S|+1} \cdot (1 - q_j)^{k_j(a_{-i}) - |S|}.$$

This is the probability that the players in $S \cup \{i\}$ succeed at project j, the rest of the players working on j fail and player i gets the credit for succeeding the project.

Using the definition we now have $\widetilde{u}_i(\vec{a}) = w_{a_i} \sum_{S \subseteq K_{a_i}(a_{-i})} \psi_i(j; S; \vec{a})$.

By using $\pi_{<i}(S)$ and $\pi_{>i}(S)$ we can break up the player's utility in the following manner:

$$\widetilde{u_i}(j,a_{-i}) = w_j \cdot \left(\sum_{S\subseteq\pi_{>i}(K_j(\vec{a}))} \psi_i(j;S;\vec{a}) \quad + \sum_{\substack{S\subseteq K_j(a_{-i})\\ S\cap\pi_{<i}(K_j(\vec{a}))\neq\emptyset}} \psi_i(j;S;\vec{a}) \right) \tag{5.2}$$

This is a convenient representation of a player's utility since it partitions the successful player sets into two types:

1. $S \subseteq \pi_{>i}(K_j(\vec{a}))$: for such a set S, player i's weight is *dominant* in S, and hence she gets most of the utility.
2. $S \subseteq K_j(a_{-i})$ and $S \cap \pi_{<i}(K_j(\vec{a})) \neq \emptyset$: for such a set S, player i's weight is *dominated*, and hence she gets only a very small fraction of the utility.

In the next two Lemmas we bound player i's utility with respect to each of these types separately. The bounds are later combined to bound $\widetilde{u_i}(j,a_{-i})$.

Lemma 5.2.15 $\widehat{u}_i(j,a_{-i}) \geq w_j \cdot \sum_{S\subseteq\pi_{>i}(K_j(\vec{a}))} \psi_i(j;S;\vec{a}) \geq \dfrac{1}{1+2\varepsilon}\widehat{u}_i(j,a_{-i})$.

Proof: We first show that $\widehat{u}_i(j,a_{-i}) \geq w_j \cdot \sum_{S\subseteq\pi_{>i}(K_j(\vec{a}))} \psi_i(j;S;\vec{a})$. To do this, we write an alternative expression for $\widehat{u}_i(j,a_{-i})$:

$$\widehat{u}_i(j,a_{-i}) = \left(\underbrace{w_j q_j (1-q_j)^{|\pi_{<i}(K_j(\vec{a}))|}}_{=r_j(|\pi_{<i}(K_j(\vec{a}))|)} \right) \cdot \underbrace{\left(\sum_{S\subseteq\pi_{>i}(K_j(\vec{a}))} q_j^{|S|} \cdot (1-q_j)^{|\pi_{>i}(K_j(\vec{a}))|-|S|} \right)}_{=1}$$

$$= w_j \sum_{S\subseteq\pi_{>i}(K_j(\vec{a}))} q_j^{|S|+1} \cdot (1-q_j)^{k_j(a_{-i})-|S|}$$

The resulting expression is an upper bound on $w_j \sum_{S\subseteq\pi_{>i}(K_j(\vec{a}))} \psi_i(j;S;\vec{a})$ since it assumes that $z_i > 0$ and the rest of the weights are 0.

By the definition of the weights, we have that $z_h \le \varepsilon z_i$ for all $h \in \pi_{>i}(K_j(\vec{a}))$. This allows us to bound the weight coefficients $\dfrac{z_i}{z_i + \sum_{h\in S} z_h}$ in $\widetilde{u}_i(j, a_{-i})$:

$$\forall S \subseteq \pi_{>i}(K_j(\vec{a})) : \ \frac{z_i}{z_i + \sum_{h\in S} z_h} \ge \frac{z_i}{z_i + \sum_{h\in \pi_{>i}(K_j(\vec{a}))} z_h} \ge \frac{z_i}{z_i + \sum_{h=1}^{|\pi_{>i}(K_j(\vec{a}))|} \varepsilon^h z_i}$$
$$= \frac{1}{1 + \sum_{h=1}^{|\pi_{>i}(K_j(\vec{a}))|} \varepsilon^h} > \frac{1}{1+2\varepsilon}$$

where the last inequality holds for $\varepsilon < 0.5$. Hence we have that $\sum_{S\subseteq \pi_{>i}(K_j(\vec{a}))} \psi_i(j; S; \vec{a}) \ge \dfrac{1}{1+2\varepsilon} \widehat{u}_i(j, a_{-i})$. □

Lemma 5.2.16 $\sum_{\substack{S\subseteq K_j(a_{-i}) \\ S\cap \pi_{<i}(K_j(\vec{a}))\neq\emptyset}} \psi_i(j; S; \vec{a}) \le \dfrac{\varepsilon}{1+\varepsilon}$.

Proof: Observe that since for each of the sets S included in the sum, $S \cap \pi_{<i}(K_j(\vec{a})) \neq \emptyset$, then in each set S there exists at least a single player $h \in S$ such that $z_h > z_i$. This implies the following bound on the weight coefficients in $\widetilde{u}_i(j, a_{-i})$:

$$\frac{z_i}{z_i + \sum_{h\in S} z_h} \le \frac{z_i}{z_i + \min_{\{h\in S | z_h > z_i\}} z_h} \le \frac{z_i}{z_i + \dfrac{z_i}{\varepsilon}} = \frac{\varepsilon}{1+\varepsilon}$$

The proof is completed by observing that $\sum_{\substack{S\subseteq K_j(a_{-i}) \\ S\cap \pi_{<i}(K_j(\vec{a}))\neq\emptyset}} q_j^{|S|+1} \cdot (1 - q_j)^{k_j(a_{-i}) - |S|} < 1$. □

It is not hard to see that given Equation 5.2 and Lemmas 5.2.15 and 5.2.16 the following bounds hold:

Corollary 5.2.17 *For every strategy vector $\vec{a}$, player i and project j:*

$$\frac{1}{1+2\varepsilon} \widehat{u}_i(j, a_{-i}) \le \widetilde{u}_i(j, a_{-i}) \le \widehat{u}_i(j, a_{-i}) + w_j \frac{\varepsilon}{1+\varepsilon}$$

We can now use the previous corollary to compute a value of ε for which the assumptions of Lemma 5.2.13 hold:

Lemma 5.2.18 *For* $\varepsilon \leq \min_{l\in M}\left(\frac{1}{4d^{n+1}w_l}\right)$ *we have that:*

$$\widehat{u}_i(j,a_{-i}) - \frac{1}{4d^{n+1}} \leq \widetilde{u}_i(j,a_{-i}) \leq \widehat{u}_i(j,a_{-i}) + \frac{1}{4d^{n+1}}$$

Proof: By the lower bound on $\widetilde{u}_i(j,a_{-i})$ from Corollary 5.2.17 we have that

$$\widehat{u}_i(j,a_{-i}) - \frac{2\varepsilon}{1+2\varepsilon}\widehat{u}_i(j,a_{-i}) = \frac{1}{1+2\varepsilon}\widehat{u}_i(j,a_{-i}) \leq \widetilde{u}_i(j,a_{-i}).$$

Let l be the project with the maximal weight, observe that this is the same project minimizing $\frac{1}{4d^{n+1}}w_l$. Thus, we have that for every $\varepsilon \leq \min_{l\in M}\left(\frac{1}{4d^{n+1}w_l}\right)$ the following bound holds:

$$\frac{2\varepsilon}{1+2\varepsilon}\widehat{u}_i(j,a_{-i}) \leq 2\varepsilon \cdot w_l \leq \frac{1}{4d^{n+1}}.$$

Similarly, by the upper bound on $\widetilde{u}_i(j,a_{-i})$ from Corollary we have that $\widetilde{u}_i(j,a_{-i}) \leq \widehat{u}_i(j,a_{-i}) + w_j\frac{\varepsilon}{1+\varepsilon} \leq \widehat{u}_i(j,a_{-i}) + \frac{1}{4d^{n+1}}$ which completes the proof.

□

This completes our proof of Theorem 5.2.11 showing that there exists ε for which any equilibrium of the game with player-weights $\{z_i = \varepsilon^i\}$ is an optimal solution in the unweighted game.

5.3 Players of Heterogeneous Abilities

We now consider the case in which players have different levels of ability. Recall from the introduction that in this model, each player i has a parameter $p_i \leq 1$, and her probability of success at project j is $p_i q_j$. As before, player i receives credit for her selected project a_i if she succeeds at it and is chosen, uniformly at random, from among all players who succeed at it. Player i's utility is the expected amount of credit she receives in this process.

Recall that $K_j(\vec{a})$ is the set of players working on project j in strategy vector $\vec{a}$; we write $s_j(K_j(\vec{a})) = w_j(1 - \prod_{i \in K_j(\vec{a})}(1 - p_i q_j))$ for the contribution of project j to the social welfare, so that the overall social welfare of $\vec{a}$ is $u(\vec{a}) = \sum_{j \in M} s_j(K_j(\vec{a}))$. We denote the marginal utility of adding player i to project j by $s_j(i|K_j(\vec{a})) = s_j(K_j(\vec{a}) \cup \{i\}) - s_j(K_j(\vec{a})) = w_j p_i q_j \prod_{l \in K_j(\vec{a})}(1 - p_l q_j)$ and we use $u(j|a_{-i}) = s_j(i|K_j(a_{-i}))$ to denote the marginal utility of player i choosing project j when the rest of the players choose a_{-i}.

There is a useful closed-form way to write i's utility, as follows. First, suppose that in strategy vector $\vec{a}$, player i selects project j, and let S denote the other players who select j. Then in order for i to receive the credit of w_j for the project, she has to succeed (with probability $p_i q_j$); moreover, some subset S' of the other players on j will succeed (with probability $\prod_{h \in S'} p_h q_j$) while the rest will fail (with probability $\prod_{h \in \{S - S'\}}(1 - p_h q_j)$), and i must be selected from among the successful players (with probability $\dfrac{1}{|S'| + 1}$). Thus we have

$$u_i(\vec{a}) = w_j p_i q_j \sum_{S' \subseteq S} \left(\frac{1}{|S'| + 1} \prod_{h \in S'} p_h q_j \prod_{h \in \{S - S'\}} (1 - p_h q_j) \right).$$

This summation over all sets S' is a natural quantity that is useful to define separately for future use; we denote it by $c_j(S)$ and refer to it as the *competition function* for project j. The competition function represents the expected reduction in credit to a player on project j due to the competition from players in the set S; instead of the expected credit of $w_j p_i q_j$ that i would receive if she worked on

j in isolation, she gets $w_j p_i q_j c_j(S)$ when the players in S are also working on j. Thus, with a_i denoting the project chosen by i, and $K_{a_i}(\vec{a})$ denoting the set of all players choosing project a_i, we have $u_i(\vec{a}) = w_{a_i} p_i q_{a_i} c_{a_i}(K_{a_i}(\vec{a}) - i)$. We now state and prove a technical lemma giving an inductive form for the competition function that will be useful later on. For the sake of brevity we define:

Definition 5.3.1 *For $S' \subseteq S$ and project j, let $I_j(S', S) = \prod_{i \in S'} p_i q_j \prod_{i \in \{S - S'\}} (1 - p_i q_j)$.*

Lemma 5.3.2 *For any project j, set of players S, and player $h \notin S$, we have*

$$c_j(S+h) = c_j(S) - p_h q_j \sum_{S' \subseteq S} \left(\frac{1}{(|S'|+1)(|S'|+2)} I_j(S', S) \right).$$

Proof: By distinguishing between the case in which the new player h succeeds and that she fails we have that:

$$\begin{aligned} c_j(S+h) &= (1 - p_h q_j) \sum_{S' \subseteq S} \left(\frac{1}{(|S'|+1)} I_j(S', S) \right) + p_h q_j \sum_{S' \subseteq S} \left(\frac{1}{(|S'|+2)} I_j(S', S) \right) \\ &= c_j(S) - p_h q_j \sum_{S' \subseteq S} \left(\frac{1}{(|S'|+1)(|S'|+2)} I_j(S', S) \right). \end{aligned}$$

□

We can show the following basic facts about this general version of the game. We first show that the more general version of the game is still a monotone valid-utility game. The proof is very similar to the proof of Claim 5.2.1; the only part that changes in a non-trivial way is the proof that the utility (social welfare) function is submodular which we provide next:

Claim 5.3.3 *The social welfare function of the project game with different abilities is submodular.*

Proof: We show that $u(\vec{a})$ has decreasing marginal utility. Recall that $u(\vec{a})$ is the summation of the projects' separate utilities. Hence it is enough to prove that the utility of every project is submodular. More formally, We need to show that for every two sets of players $S \subseteq S'$ and for every project j and player i, we have $s_j(i|S) \geq s_j(i|S')$. To prove this, we observe that $w_j p_i q_j \prod_{l \in S}(1 - p_l q_j) \geq w_j p_i q_j \prod_{l \in S'}(1 - p_l q_j)$ simply becasue $1 \geq \prod_{l \in \{S' - S \cap S'\}}(1 - p_l q_j)$. □

Next, we show that the more general version of the game is a congestion game. This is less clear-cut than in the case of identical players, since now the payoffs depend not just on the number of players sharing a project but on their identities. To bypass this we prove that the utility functions for the project game with different abilities obey a certain structural property that, by results of Monderer and Shapley [72], implies that the game is a congestion game.

Claim 5.3.4 *The project game with different abilities is a congestion game.*

Proof: Recall that the utility of a player i depends not only on the number of other players working on i's project, but also on their identities. As a result, to establish that the game is a congestion game, we use a different characterization of congestion given by Monderer and Shapley in Corollary 2.9 of their paper [72]. Using the notation and terminology we have defined for the project game, the corollary can be written as follows.

Theorem 5.3.5 (Adapted from Monderer-Shapley) *The project game is an (exact) potential game if for every two players i, l, projects $x_i \neq y_i, x_l \neq y_l$ and strategy vector $a_{-i,l}$:*

$$u_i(y_i, x_l, a_{-i,l}) - u_i(x_i, x_l, a_{-i,l}) + u_l(y_i, y_l, a_{-i,l}) - u_l(y_i, x_l, a_{-i,l}) +$$

$$u_i(x_i, y_l, a_{-i,l}) - u_i(y_i, y_l, a_{-i,l}) + u_l(x_i, x_l, a_{-i,l}) - u_l(x_i, y_l, a_{-i,l}) = 0$$

We now use this to prove that the project game with different abilities is an exact potential game, from which the claim follows, since by another result of Monderer and Shapley, every finite exact potential game is isomorphic to a congestion game.

Recall that the utility of a player i is affected only by the players who are working on the same project as i. Hence, we should differentiate in the condition given in Theorem 5.3.5 between the cases in which players i and l are working on the same project and those in which they are not. By symmetry we can assume without loss of generality that $x_i \neq y_l$ and that $y_i \neq x_l$. Before proceeding with the case analysis, we present the following Lemma that will turn out useful for handling some of the cases. The proof is provided below.

Lemma 5.3.6 *For any three projects x, y, z such that $x \neq y$ and $x \neq z$:*

$$u_i(x, x, a_{-i,l}) - u_i(x, y, a_{-i,l}) = u_l(x, x, a_{-i,l}) - u_l(z, x, a_{-i,l}).$$

We distinguish between the following cases:

1. $x_i \neq x_l$ and $y_i \neq y_l$. By rearranging the terms we get:

$$\underbrace{u_i(y_i, x_l, a_{-i,l}) - u_i(y_i, y_l, a_{-i,l})}_{=0} + \underbrace{u_i(x_i, y_l, a_{-i,l}) - u_i(x_i, x_l, a_{-i,l})}_{=0} +$$
$$\underbrace{u_l(y_i, y_l, a_{-i,l}) - u_l(x_i, y_l, a_{-i,l})}_{=0} + \underbrace{u_l(x_i, x_l, a_{-i,l}) - u_l(y_i, x_l, a_{-i,l})}_{=0} = 0.$$

   For example, $u_i(y_i, x_l, a_{-i,l}) - u_i(y_i, y_l, a_{-i,l}) = 0$ since $K_{y_i}(y_i, x_l, a_{-i,l}) = K_{y_i}(y_i, y_l, a_{-i,l})$.

2. $x_i = x_l$ and $y_i \neq y_l$. By using Lemma 5.3.6 and the previous argument we have that:

$$\underbrace{u_i(x_i, y_l, a_{-i,l}) - u_i(x_i, x_l, a_{-i,l}) + u_l(x_i, x_l, a_{-i,l}) - u_l(y_i, x_l, a_{-i,l})}_{=0} +$$
$$\underbrace{u_i(y_i, x_l, a_{-i,l}) - u_i(y_i, y_l, a_{-i,l})}_{=0} + \underbrace{u_l(y_i, y_l, a_{-i,l}) - u_l(x_i, y_l, a_{-i,l})}_{=0} = 0.$$

3. $x_i \neq x_l$ and $y_i = y_l$. This case is symmetric to case 2.

4. $x_i = x_l$ and $y_i = y_l$. This case can be proved by using Lemma 5.3.6 twice, similar to case 2.

$\square$

Proof of Lemma 5.3.6 By using Lemma 5.3.2 we have that:

$$\begin{aligned} u_i(x, x, a_{-i,l}) &= w_x p_i q_x \cdot c_x(K_x(a_{-i,l}) + l) \\ &= w_x p_i q_x \cdot \left(c_l(K_x(a_{-i,l})) - p_l q_x \sum_{S \subseteq K_x(a_{-i,l})} \left(\frac{1}{(|S|+1)(|S|+2)} I_x(S, K_x(a_{-i,l})) \right) \right) \\ &= u_i(x, y, a_{-i,l}) - w_x p_i q_x \cdot p_l q_x \sum_{S \subseteq K_x(a_{-i,l})} \left(\frac{1}{(|S|+1)(|S|+2)} I_x(S, K_x(a_{-i,l})) \right) \end{aligned}$$

Similarly we have: $u_l(x, x, a_{-i,l}) =$

$$u_l(x, y, a_{-i,l}) - w_x p_l q_x \cdot p_i q_x \sum_{S \subseteq K_x(a_{-i,l})} \left(\frac{1}{(|S|+1)(|S|+2)} I_x(S, K_x(a_{-i,l})) \right)$$

Hence, $u_i(x, x, a_{-i,l}) - u_i(x, y, a_{-i,l}) = u_l(x, x, a_{-i,l}) - u_l(z, x, a_{-i,l})$. $\square$

Next, we show how one can compute a Nash equilibrium in the project game with different abilities. We will later see that the optimal solution for this more general version of the game can no longer be computed in polynomial time.

Claim 5.3.7 *A Nash equilibrium for the project game with different abilities can be computed in polynomial time.*

Proof: We begin by presenting a simple greedy algorithm for computing a Nash equilibrium. We then prove that the algorithm indeed computes a Nash equilibrium and that it runs in polynomial time. The algorithm is defined as follows:

1. Sort the players by their abilities.
2. Go over the players in descending order of ability and allocate each player to the project maximizing her utility with respect to the players previously allocated.

Denote the resulting allocation by $\vec{a}$. Assume towards a contradiction that $\vec{a}$ is not a Nash equilibrium. Let player i be the first player in the order for which there exists some project j such that $u_i(j, a_{-i}) > u_i(\vec{a})$. Let player l be the last player that joined project a_i. Note that for player l it has to be the case that $u_l(\vec{a}) \geq u_l(j, a_{-l})$, since its expected utility is exactly the same as it was in the time she made her choice and the expected utility from project j could have only decreased (this also implies that $l \neq i$). Recall that:

$$w_{a_i} p_l q_{a_i} c_{a_i}(K_{a_i}(\vec{a}) - l) = u_l(\vec{a}) \geq u_l(j, a_{-l}) = w_j p_l q_j c_j(K_j(\vec{a})).$$

This implies that $w_{a_i} q_{a_i} c_{a_i}(K_{a_i}(\vec{a}) - l) \geq w_j q_j c_j(K_j(\vec{a}))$. We now observe that $c_{a_i}(K_{a_i}(\vec{a}) - i) \geq c_{a_i}(K_{a_i}(\vec{a}) - l)$ since $q_i > q_l$ and thus $u_i(j, a_{-i}) \leq u_i(\vec{a})$ in contradiction to the assumption that $\vec{a}$ is not a Nash equilibrium.

We now show that the algorithm runs in polynomial time. To do this we should show that the players utilities can be computed in polynomial time, this amounts to computing the computation functions:

Claim 5.3.8 *The competition function can be computed in poly time.*

Proof: We provide an alternative inductive formula for the competition function that can be evaluated in polynomial time. We define the probability that the set of successful players is of size k by $l(S,k)$. With this notation in mind we have that:

$$c_j(S) = \sum_{S' \subseteq S} \left(\frac{1}{|S'|+1} \prod_{h \in S'} p_h q_j \prod_{h \in \{S - S'\}} (1 - p_h q_j) \right) = \sum_{k=0}^{k=|S|} \frac{1}{k+1} l(S,k)$$

We now consider the effect that adding an additional player $h \notin S$ to the set of players working has on a project on the value of $l(S+h,k)$. It is not hard to see that $l(S+h,k) = (1-p_h q_j) l(S,k) + p_h q_j l(S,k-1)$, where $l(S,-1) = 0$ and $l(S,|S|+1) = 0$.

This formula readily admits a simple algorithm for computing $c_j(S)$. We first fix some order on the players in S, for simplicity we rename the players according to this order. The algorithm performs $|S|$ steps, where in step i it computes $l(\{1,\dots,i\},k)$ for every $k \in \{1,\dots,i\}$. Observe that for each k computing $l(\{1,\dots,i\},k)$ can be done in constant time by using the values of $l(\{1,\dots,i-1\},k)$ and $l(\{1,\dots,i-1\},k-1)$ that where computed in the previous step. Therefore, we have that step i takes $O(i)$ time implying that the whole computation of $c_j(S)$ takes $O(n^2)$.

□ □

Claim 5.3.9 *Computing the social optimum for the project game with different abilities is NP-hard.*

Proof: We use a reduction from the *Subset Product* problem, whose NP-completeness is established in Garey and Johnson [43]. The Subset Product problem is defined as follows: given a set of n natural numbers $X = \{x_1, ..., x_n\}$ and a target number Q^*, does there exist $S \subseteq X$ such that $\prod_{x_i \in S} x_i = Q^*$?

As a first step, we show that the closely related *Multiplicative Number Partition* problem (MNP) is NP-complete. In MNP, we are again given a set of n natural numbers $X = \{x_1, ..., x_n\}$, but now we are asked whether there is a partition (S, T) of X such that $\prod_{x_i \in S} x_i = \prod_{x_j \in T} x_j$. We can show that MNP is NP-complete by a reduction from Subset Product, by analogy with the corresponding reduction from Subset Sum to (Additive) Number Partition. That is, given an instance of Subset Product with a set X and a target Q^*, we define $P = \prod_{x_i \in X} x_i$. Notice that if P is not divided by Q^* without a remainder then there is no subset as needed. Hence, we can assume without loss of generality that Q^* divides P. We then show that we can solve MNP for $X' = X \cup \{x_{n+1} = P^2/Q^*, x_{n+2} = P \cdot Q^*\}$ if and only if we can solve Subset Product.

Notice that x_{n+1} and x_{n+2} should be in different sets because $x_{n+1} \cdot x_{n+2} > P$. We assume without loss of generality that $x_{n+1} \in S'$. Define $Y = \prod_{x_i \in \{S' - x_{n+1}\}} x_i$. By the definition of Y we have that $\prod_{x_j \in \{T' - x_{n+2}\}} x_j = \frac{P}{Y}$. By substituting in the equality $\prod_{x_i \in S'} x_i = \prod_{x_j \in T'} x_j$ we get that:

$$\frac{P^2}{Q^*} \cdot Y = P \cdot Q^* \cdot \frac{P}{Y} \iff \frac{Y}{Q^*} = \frac{Q^*}{Y}$$

Since both Y and Q^* are positive we get that $Y = Q^*$. Thus, we have proven the following lemma:

Lemma 5.3.10 *For a partition (S', T') of X' : $\prod_{x_i \in S'} x_i = \prod_{x_j \in T'} x_j \iff \prod_{x_i \in \{S' - x_{n+1}\}} x_i = Q^*$*

We can now conclude that for $S = \{S' - \{x_{n+1}\}\}$, which by the construction is a subset of X, we have that $\prod_{x_i \in S} = Q^*$. It follows that MNP is NP-complete.

Finally, we prove that the socially optimal assignment in the project game with different abilities is NP-hard. We do this by a reduction from MNP to the special case of the optimal assignment in which we have n players and 2 identical projects. In this special case we assume both projects have a weight of 1 and success probability 1, and player i has a failure probability $\bar{p}_i$.

Given an instance of MNP, we create an instance of this special case of the optimal assignment problem by defining for every number x_i a player i with failure probability $\bar{p}_i = \frac{1}{x_i}$. The optimal solution to the assignment of players to projects is a partition (S, T) that maximizes the social welfare: $(1 - \prod_{i \in S} \bar{p}_i) + (1 - \prod_{j \in T} \bar{p}_j)$. This implies that the optimal partition actually minimizes $\prod_{i \in S} \bar{p}_i + \prod_{j \in T} \bar{p}_j$. By plugging in the values of $\bar{p}_i$ and $\bar{p}_j$. we have that the optimal solution minimizes:

$$\prod_{i \in S} \frac{1}{x_i} + \prod_{j \in T} \frac{1}{x_j} = \frac{\prod_{i \in S} x_i + \prod_{j \in T} x_j}{P}.$$

The following lemma completes the proof by establishing the connection between social welfare maximization and MNP: (the proof is provided below)

Lemma 5.3.11 *A partition (S, T) minimizes $\prod_{i \in S} x_i + \prod_{j \in T} x_j$ if and only if it minimizes $|\prod_{i \in S} x_i - \prod_{j \in T} x_j|$*

Thus, given a partition (S, T) that is an optimal solution to the identical projects variant, we can determine the answer to MNP by checking whether $\prod_{i \in S} x_i = \prod_{j \in T} x_j$.

□

Proof of Lemma 5.3.11: In this lemma, we use ΠS as a shorthand for $\prod_{i \in S} x_i$. Let (S, T) and (S', T') be two partitions of X. We show that if $\Pi S + \Pi T < \Pi S' + \Pi T'$ then $|\Pi S - \Pi T| < |\Pi S' - \Pi T'|$. We begin by observing that the following three inequalities are equivalent, since all terms are products of natural (and hence non-negative) numbers:

$$\Pi S + \Pi T < \Pi S' + \Pi T'$$

$$(\Pi S + \Pi T)^2 < (\Pi S' + \Pi T')^2$$

$$(\Pi S)^2 + 2\Pi S \cdot \Pi T + (\Pi T)^2 < (\Pi S')^2 + 2\Pi S' \cdot \Pi T' + (\Pi T')^2.$$

Since both (S, T) and (S', T') are partitions of X, we have that: $\Pi S \cdot \Pi T = \Pi S' \cdot \Pi T'$, and hence we can subtract four times this common product from both sides of the previous inequality to get three more equivalent inequalities:

$$(\Pi S)^2 - 2\Pi S \cdot \Pi T + (\Pi T)^2 < (\Pi S')^2 - 2\Pi S' \cdot \Pi T' + (\Pi T')^2$$

$$(\Pi S - \Pi T)^2 < (\Pi S' - \Pi T')^2$$

$$|\Pi S - \Pi T| < |\Pi S' - \Pi T'|.$$

$\square$

5.3.1 Re-weighting Projects to Achieve Social Optimality

We now describe how to re-weight projects, creating new weights $\{w'_j\}$, so as to make a given social optimum $\vec{\sigma}$ a Nash equilibrium. First, since the relative values of the project weights are all that matters, we can choose any project x arbitrarily and set its new weight w'_x equal to 1. We will set the weights w'_j of the other projects so that every player's favorite alternate project (and hence the target of any potential deviation) is x.

Now, among all the players working on another project $j \neq x$, which one has the greatest incentive to move to x? It is the player $i \in K_j(\vec{o})$ with the lowest ability p_i, since all players $i' \in K_j(\vec{o})$ experience the same competition function $c_x(K_x(\vec{o}))$, but i experiences the strongest competition from the other players in $K_j(\vec{o})$. This is because they all have ability at least as great as i, so i has the most to gain by moving off j.

Motivated by this, for a strategy vector $\vec{a}$ and a project j, we define $\delta_j(\vec{a})$ to be the player $i \in K_j(\vec{a})$ of minimum ability p_i. We define $w'_x = 1$ and for every other project $j \neq x$, we define

$$w'_j = \frac{q_x c_x(K_x(\vec{o}))}{q_j c_j(K_j(\vec{o}) - \delta_j(\vec{o}))}. \tag{5.3}$$

Theorem 5.3.12 *The optimal assignment $\vec{o}$ is a Nash equilibrium in the game with the given weights $\{w'_j\}$.*

Proof: To prove this, we will show that if a player did want to move to another project, she would choose to move to project x. After establishing this, it is enough to show that all the players working on project x in the optimal assignment do not want to move to another project, and that the rest of the players do not want to move to project x.

We now show that a player i working on a project other than x views x as her best alternate project.

Lemma 5.3.13 *For any player i such that $o_i \neq x$, and for every project $j \neq o_i$, we have that $u_i(x, o_{-i}) \geq u_i(j, o_{-i})$*

Proof: We need to show that $w'_x p_i q_x c_x(K_x(\vec{o})) \geq w'_j p_i q_j c_j(K_j(\vec{o}))$. By setting the weights to their values according to Formula (5.3), we get that:

$$p_i q_x c_x(K_x(\vec{o})) \geq \frac{q_x c_x(K_x(\vec{o}))}{q_j c_j(K_j(\vec{o}) - \delta_j(\vec{o}))} p_i q_j c_j(K_j(\vec{o}))$$

By rearranging the terms we have that: $c_j(K_j(\vec{o}) - \delta_j(\vec{o})) \geq c_j(K_j(\vec{o}))$. Intuitively, this inequality follows from the fact that as more players work on a project, it is less likely that a specific player will be the one to succeed at it. Formally, it follows from Lemma 5.3.2 above. □

Finally, we show that players working on project x do not want to leave project x, and players not working on x do not want to move to x (and hence, by Lemma 5.3.13, do not want to move anywhere else either).

Lemma 5.3.14

1. *All players who are working in the optimal assignment on project x do not want to move to a different project.*
2. *All players who are working in the optimal assignment on project different than x do not want to move to project x.*

Proof: Assume towards a contradiction that there exists a player i who prefers to work on project $j \neq o_i$. This means that

$$w'_{o_i} p_i q_{o_i} c_{o_i}(K_{o_i}(\vec{o}) - i) < w'_j p_i q_j c_j(K_j(\vec{o}))$$

For each of the two statements we set w'_{o_i} and w'_j to their values according to Formula (5.3) and get to a contradiction by rearranging the terms.

1. We set $w'_{o_i} = 1$ and $w'_j = \dfrac{q_x c_x(K_x(\vec{o}))}{q_j c_j(K_j(\vec{o}) - \delta_j(\vec{o}))}$ and get the following inequality:

$$p_i q_x c_x(K_x(\vec{o}) - i) < \frac{q_x c_x(K_x(\vec{o}))}{q_j c_j(K_j(\vec{o}) - \delta_j(\vec{o}))} p_i q_j c_j(K_j(\vec{o}))$$

After rearranging the inequality we get that:

$$\frac{c_x(K_x(\vec{o})-i)}{c_x(K_x(\vec{o}))} < \frac{c_j(K_j(\vec{o}))}{c_j(K_j(\vec{o})-\delta_j(\vec{o}))}$$

The contradiction follows by noticing that $c_x(K_x(\vec{o})-i) > c_x(K_x(\vec{o}))$ by Lemma 5.3.2; however, by the same lemma we also have that $c_j(K_j(\vec{o})) < c_j(K_j(\vec{o})-\delta_j(\vec{o}))$.

2. We set $w'_{o_i} = \dfrac{q_x c_x(K_x(\vec{o}))}{q_{o_i} c_{o_i}(K_{o_i}(\vec{o})-\delta_{o_i}(\vec{o}))}$ and $w'_j = 1$ and get the following inequality:

$$\frac{q_x c_x(K_x(\vec{o}))}{q_{o_i} c_{o_i}(K_{o_i}(\vec{o})-\delta_{o_i}(\vec{o}))} p_i q_{o_i} c_{o_i}(K_{o_i}(\vec{o})-i) < p_i q_x c_x(K_x(\vec{o}))$$

After rearranging the inequality we get that:

$$\frac{c_{o_i}(K_{o_i}(\vec{o})-i)}{c_{o_i}(K_{o_i}(\vec{o})-\delta_{o_i}(\vec{o}))} < 1$$

$$\frac{c_{o_i}(K_{o_i}(\vec{o})-i)}{c_{o_i}(K_{o_i}(\vec{o})-\delta_{o_i}(\vec{o}))} = \frac{c_{o_i}(\{K_{o_i}(\vec{o})-i-\delta_{o_i}(\vec{o})\}+\delta_{o_i}(\vec{o}))}{c_{o_i}(\{K_{o_i}(\vec{o})-i-\delta_{o_i}(\vec{o})\}+i)}$$

By Lemma 5.3.2 we have that as p_h is greater the amount we subtract from $c(S)$ is greater. Therefore since by definition $p_i \geq p_{\delta_{o_i}(\vec{o})}$, we have

$$c_{o_i}(\{K_{o_i}(\vec{o})-i-\delta_{o_i}(\vec{o})\}+\delta_j(\vec{o})) > c_{o_i}(\{K_{o_i}(\vec{o})-i-\delta_{o_i}(\vec{o})\}+i)$$

and this is a contradiction.

□

Since this establishes that all players want to stay with their current projects, it follows that $\vec{o}$ is a Nash equilibrium under the modified weights, and hence the proof of Theorem 5.3.12 is complete. □ □

5.3.2 Re-weighting Players to Achieve Social Optimality

It is also possible to re-weight the players so as to make the social optimum a Nash equilibrium. Because the greedy algorithm no longer computes the social optimum, it is no longer enough to use weights to approximately simulate an arbitrary ordering on the players. However, we can use an extension of this plan that incorporates two additional ingredients: first, we base the greedy ordering on the socially optimal assignment, and second, we do not use a strict ordering but rather one that groups the players into *stages* of equal weight.

The algorithm for assigning weights is as follows. In the beginning, we fix an optimal assignment $\vec{o}$ and a sufficiently small value of $\varepsilon > 0$ (to be determined below), and we declare all players to be *unassigned*. The algorithm then operates in a sequence of *stages* $c = 1, 2, \ldots$. At the start of stage c, some players have been given weights and been assigned to projects, resulting in a partial strategy vector $\vec{a^c}$ consisting only of players assigned before stage c. We show that at the start of stage c, each unassigned player would maximize her payoff by choosing a project from the set

$$X_c = \{j \mid w_j \prod_{h \in K_j(\vec{a^c})} (1 - p_h q_j) q_j = \max_l w_l \prod_{h \in K_l(\vec{a^c})} (1 - p_h q_l) q_l\}$$

Thus in stage c, the algorithm does the following. It first computes this set of projects X_c. Then, for each project $j \in X_c$ for which there exists a player i such that $o_i = j$ and i is unassigned, it assigns i to project j, and sets $z_i = \varepsilon^c$.

It would be natural to try proving that with these weights, the assignment $\vec{o}$ is a Nash equilibrium. However, this is not necessarily correct. In the final stage c^* of the algorithm, it may be that the number of unassigned players is less than $|X_{c^*}|$, and in this case some of the unassigned players might go to projects other than the ones corresponding to $\vec{o}$. However we show that the following defined assignment $\vec{o'}$, which is derived from $\vec{o}$, is both an optimal assignment and a Nash equilibrium with these weights.

Definition 5.3.15 *Assignment $\vec{o'}$ is constructed as follows:*

1. *For every player i that was not assigned in the last stage of the algorithm, we define $o'_i = o_i$.*
2. *For every project $j \in X_{c^*}$ we compute the value*

$$c_j(\vec{o'}) = \sum_{S \subseteq K_j(\vec{o'})} \frac{z^*}{(\sum_{l \in S} z_l) + z^*} I_j(S, K_j(\vec{o'}))$$

   where z^* is the weight defined for players that were assigned last and $I_j(S, K_j(\vec{o'})$ is the same as defined in Definition 5.3.1: $I_j(S, K_j(\vec{o'})) = \prod_{l \in S} p_l q_j \prod_{l \in K_j(\vec{o'})-S}(1 - p_l q_j)$.
3. *We sort all the projects in X_{c^*} by their value for $w_j c_j(\vec{o'})$.*
4. *We assign each unassigned player to one of the top projects in X_{c^*} according to the sorting.*

Theorem 5.3.16 *The previously defined assignment $\vec{o'}$ is an optimal assignment and a Nash equilibrium in the game with weights $\{z_i\}$.*

We begin by showing that $\vec{o'}$ is indeed an optimal assignment:

Claim 5.3.17 *$\vec{o'}$ is an optimal assignment (i.e., $u(\vec{o'}) = u(\vec{o})$).*

Proof: By the construction of $\vec{o'}$ the only players that might not work on the same project as in $\vec{o}$ are those that were assigned last. All these players are assigned to projects in X_{c^*}. Notice that all projects in X_{c^*} maximize $w_j \prod_{\{l \in K_j(\vec{a}^{c^*-1})\}}(1 - p_l q_j) q_j$. Hence, the contribution of the players assigned last is the same regardless of which specific project in X_{c^*} they are working on. Therefore $\vec{o'}$ is an optimal assignment. $\square$

Next, we show that $\vec{o'}$ is a Nash equilibrium in the game with weights $\{z_i\}$. The proof resembles the proof of Theorem 5.2.11. As in Theorem 5.2.11, the actual utilities of the re-weighted players for a given strategy vector $\vec{a}$ are denoted by $\widetilde{u_i}(\vec{a})$, and their "ideal" utilities under the partial order we are trying to simulate are denoted $\widehat{u_i}(\vec{a})$:

$$\widehat{u_i}(\vec{a}) = w_{a_i} p_i q_{a_i} \prod_{l \in \pi_{<i}(K_{a_i}(\vec{a}))} (1 - p_l q_{a_i})$$

where $\pi_{<i}(K_{a_i}(\vec{a}))$, as before, is the set of players working on project a_i which are strictly before player i in the order.

Recalling that the projects' weights and success probabilities are rational, let d be the common denominator of all the terms in the sets $\{w_j : j \in M\}$ and $\{p_i q_j : i \in N, j \in M\}$.

Claim 5.3.18 *If for every player i and project j such that the weight of player i is* unique *among players working on project j:*

$$\widehat{u_i}(j, o'_{-i}) - \frac{1}{4d^{n+1}} \leq \widetilde{u_i}(j, o'_{-i}) \leq \widehat{u_i}(j, o'_{-i}) + \frac{1}{4d^{n+1}},$$

then $\vec{o'}$ is a Nash equilibrium.

Proof: Assume towards a contradiction that $\vec{o'}$ is not a Nash equilibrium. Thus, there exists a player i and a project $j \neq o'_i$ such that $\widetilde{u_i}(j, o'_{-i}) > \widetilde{u_i}(\vec{o'})$. By the weighting algorithm we have that $\widehat{u_i}(\vec{o'}) \geq \widehat{u_i}(j, o'_{-i})$. To see this, assume player

i was assigned in stage c. If $c < c^*$, then $o'_i = o_i$ and o_i was one of the projects maximizing the marginal contribution to social welfare; if $c = c^*$, then by the definition of $\vec{o}$, the project o'_i must have been one of the projects maximizing this marginal contribution. So in either case we have

$$w_{o'_i} \prod_{l \in K_{o'_i}(\vec{a^c})} (1 - p_l q_{o'_i}) q_{o'_i} \geq w_j \prod_{l \in K_j(\vec{a^c})} (1 - p_l q_j) q_j.$$

By multiplying both sides with p_i we have that $\widehat{u}_i(\vec{o'}) \geq \widehat{u}_i(j, o'_{-i})$. We now distinguish between two cases:

1. $\widehat{u}_i(\vec{o'}) > \widehat{u}_i(j, o'_{-i})$ – this means that player i was assigned at a different stage than all the players working on project j were. Hence, player i has a unique weight on project j. Since every player always has a unique weight on the project she is allocated to by using the assumption of the claim, we get that:

$$\widehat{u}_i(j, o'_{-i}) + \frac{1}{4d^{n+1}} \geq \widetilde{u}_i(j, o'_{-i}) > \widetilde{u}_i(\vec{o'}) \geq \widehat{u}_i(\vec{o'}) - \frac{1}{4d^{n+1}}$$

   This implies that $\widehat{u}_i(\vec{o'}) - \widehat{u}_i(j, o'_{-i}) < \frac{1}{2d^{n+1}}$. But by the definition of d, since $\widehat{u}_i(\vec{o'})$ and $\widehat{u}_i(j, o'_{-i})$ are not equal, they must differ by at least $\frac{1}{d^{n+1}}$, a contradiction.

2. $\widehat{u}_i(\vec{o'}) = \widehat{u}_i(j, o'_{-i})$ – let c be the stage in which player i was assigned. The fact that $\widehat{u}_i(\vec{o'}) = \widehat{u}_i(j, o'_{-i})$ implies that $j \in X_c$. Now, Lemma 5.3.19 below guarantees that if $c < c^*$ then there is another player l of the same stage as i's ($z_i = z_l$) that is allocated to project j ($o_l = j$). Similarly, for $c = c^*$, by the construction of $\vec{o'}$ player i there has to be a player l of the same weight as i's such that $o'_j = l$. Since player i can only improve her utility by switching to one of the "top" projects in X_{c^*}. On each of these projects there is already a player working with the same weight as player i.

By the definition of the marginal utility we have that $\widehat{u}_i(\vec{o'}) = \widehat{u}_i(j, o'_{-i}) = \widehat{u}_i(j, o'_{-i,l})$. In Lemma 5.3.20 below, we show that $\widetilde{u}_i(j, o'_{-i}) \leq \widehat{u}_i(j, o'_{-i,l}) - \frac{1}{2d^{n+1}} = \widehat{u}_i(\vec{o'}) - \frac{1}{2d^{n+1}}$. This contradicts with the assumption that $\widetilde{u}_i(j, o'_{-i}) > \widetilde{u}_i(\vec{o'})$ since we have that $\widetilde{u}_i(\vec{o}) \geq \widehat{u}_i(\vec{o'}) - \frac{1}{4d^{n+1}}$.

□

Lemma 5.3.19 *In every stage $c < c^*$ of the algorithm, for every project $j \in X_c$ there exists an* unassigned *player i such that $o_i = j$.*

Proof: Assume towards a contradiction that in some stage c there exists a project $j \in X_c$ for which all the players working on it in $\vec{o}$ have already been assigned. Let player l be a player left *unassigned* after stage c then $u(l, o_{-l}) > u(\vec{o})$. This is because in each stage the projects in the set X_c maximize the marginal contribution. Since the utility is submodular, the marginal contribution of the projects can only decrease in every stage. Hence, player l's marginal contribution to project j is greater than her contribution to project o_l. Also, by removing player l from project o_l the marginal contribution of the rest of the players working on o_l can only increase. From this we conclude that $u(l, o_{-l}) > u(\vec{o})$, in contradiction to $\vec{o}$ being an optimal assignment. □

Lemma 5.3.20 *For every two players i and l that have the same weight, $\widetilde{u}_i(o'_l, o'_{-i}) \leq \widehat{u}_i(o'_l, o'_{-i,l}) - \frac{1}{2d^{n+1}}$.*

Proof: Let $z_i = z_l = z^*$. We have

$$\widetilde{u}_i(o'_l, o'_{-i}) = (1 - p_l q_{o'_l}) \cdot w_{o_l} p_i q_{o'_l} \sum_{S \subseteq \{K_{o'_l}(o'_{-i,l})\}} \frac{z^*}{(\sum_{h \in S} z_h) + z^*} \cdot I_{o'_l}(S, K_{o'_l}(o'_{-i,l}))$$
$$+ p_l q_{o'_l} \cdot w_{o'_l} p_i q_{o'_l} \sum_{S \subseteq \{K_{o'_l}(o'_{-i,l})\}} \frac{z^*}{(\sum_{h \in S} z_h) + 2z^*} \cdot I_{o'_l}(S, K_{o'_l}(o'_{-i,l})).$$

By rearranging the terms we have that $\widetilde{u}_i(o'_l, o'_{-i}) = \widetilde{u}_i(o'_l, o'_{-i,l}) -$

$$w_{o'_l} p_l q_{o'_l} p_i q_{o'_l} \sum_{S \subseteq \{K_{o'_l}(o'_{-i,l}\}} \left(\frac{z^*}{(\sum_{h \in S} z_h) + z^*} - \frac{z^*}{(\sum_{h \in S} z_h) + 2z^*} \right) \cdot I_{o'_l}(S, K_{o'_l}(o'_{-i,l})).$$

By only considering the empty set in the summation we get that:

$$\widetilde{u}_i(o_l, o_{-i}) \leq \widetilde{u}_i(o_l, o_{-i,l}) - \frac{1}{2} w_{o_l} p_l q_{o_l} p_i q_{o_l} \prod_{\{h \in K_{o_l}(o_{-i,l})\}} (1 - p_h q_{o_l})$$
$$= \widetilde{u}_i(o_l, o_{-i,l}) - \frac{1}{2} p_l q_{o_l} \cdot \widehat{u}_i(o_l, o_{-i,l})$$

By the definition of the common denominator we have that $\frac{1}{2} p_l q_{o_l} \cdot \widehat{u}_i(o_l, o_{-i,l}) \geq \frac{1}{d^{n+1}}$ and hence $\widetilde{u}_i(o_l, o_{-i}) \leq \widetilde{u}_i(o_l, o_{-i,l}) - \frac{1}{2d^{n+1}}$ as required. $\square$

The proof of Theorem 5.3.16 is completed by observing that a very similar proof to the one we provided for identical players (Lemma 5.2.18) shows how for $\varepsilon <$ the bounds for Claim 5.3.18 hold, That is: for every player i and project j: $\widehat{u}_i(j, o'_{-i}) - \frac{1}{4d^{n+1}} \leq \widetilde{u}_i(j, o'_{-i}) \leq \widehat{u}_i(j, o'_{-i}) + \frac{1}{4d^{n+1}}$. We omit the proof since it is very similar to the case of identical players. To see this, we present the adjusted definition for $\psi_i(j; S; \vec{a})$:

Definition 5.3.21 $\psi_i(j; S; \vec{a}) = \frac{z_i}{z_i + \sum_{l \in S} z_l} p_i q_j \prod_{l \in S} p_l q_j \prod_{l \in \{K_j(a_{-i}) - S\}} (1 - p_l q_j)$

Using this definition it is easy to derive a proof similar to the proof of Lemma **??**. The lemma follows by using Lemma 5.2.18 as is.

5.4 A Further Generalization: Arbitrary Success Probabilities

Finally, we consider a further generalization of the model, in which player i has an arbitrary success probability p_{ij} when working on project j. The strategies and payoffs remain the same as before, subject to this modification. Also, this generalization is a monotone valid-utility game and congestion game; however, we omit the proofs since they are very similar to the proofs for the case from the previous section.

An interesting feature of this generalization is that one can no longer always make the social optimum a Nash equilibrium by re-weighting projects. To see why, consider an example in which there are two players 1 and 2, and two projects a and b. We have $w_a = w_b = 1$ and success probabilities $p_{1a} = 1$, $p_{1b} = 0.5$, $p_{2a} = 0.5$, and $p_{2b} = 0.1$. Now, the social optimum is achieved if player 1 is assigned to a and player 2 is assigned to b. But this gives too little utility to player 2, and in order to keep player 2 on b, we need to re-weight the projects so that $w'_b \geq 2.5w'_a$. In this case, however, player 1 also has an incentive to move to b, proving that no re-weighting can enforce the social optimum.

The case of re-weighting players is an open question. In Sections 5.2 and 5.3, we used the re-weighting of players in a limited way, to simulate an ordering. It is possible that a similar tactic can also be used in the general model — that is, there may always exist a partial ordering on the players yielding a socially optimal Nash equilibrium. If this is not the case, one can potentially make use of weights on the players in more complex ways.

As one interesting partial result on the re-weighting of players in this model, we can show the following.

Theorem 5.4.1 *If there exists a social optimum $\vec{o}$ that assigns each player to a distinct project, then it is possible to re-weight the players so that $\vec{o}$ is a Nash equilibrium.*

Proof: The proof, uses an analysis of the alternating-cycle structure of a bipartite graph on players and projects, combined with ideas from the proof of Theorem 5.3.16. As in other results on re-weighting players, we use the weights to simulate an ordering on the players. That is, we arrange the players in some specific order, and then we announce that all the credit on a project will be allocated to the first player in the order to succeed at it. We first describe how to construct such an ordering for which every Nash equilibrium in the resulting game is socially optimal, and then we show how to approximately simulate this order using weights.

Let $\vec{o}$ be an optimal assignment of players to projects in which there is at most one player working on each project. The following lemma establishes that there must be some player i who would choose her own project o_i if he were placed first in the order.

Lemma 5.4.2 *If in the optimal assignment there is at most one player working on each project then there exists a player i such that* $\max_j w_j p_{i,j} \leq w_{o_i} p_{i,o_i}$.

Proof: Assume towards a contradiction that such a player does not exist. Then for every player i there exists a project g_i such that $w_{g_i} p_{i,g_i} > w_{o_i} p_{i,o_i}$. Since in the optimal assignment there is at most one player working on each project, we can picture the assignment as a matching between the projects and the players. Consider the bipartite graph which has the players on the left side, the projects on the right side and both the edges of the optimal matching $\{(i, o_i)\}$ and edges from each player to her preferred project $\{(i, g_i)\}$. We color the edges in the first of these sets blue and the edges in the second of these sets red. This bipartite graph has $2n$ nodes and $2n$ edges, and it therefore contains a cycle C. The cycle C has interleaving red and blue edges, because each player on C has exactly one incident blue edge and one incident red edge. Hence, we can form a new perfect matching between players and projects by re-matching each player on C with the project to which he is matched using her red edge rather than her blue edge. Since all the players strictly prefer the projects to which they are connected by red edges, the social welfare of this new matching is greater than the social welfare of the blue matching, which contradicts the optimality of the blue matching. □

Given this lemma, we can construct the desired ordering by induction. We identify a player i with the property specified in Lemma 5.4.2 and place him first in the order. Since he knows he will receive all the credit from any project he succeeds at, he will choose her own project in the optimal solution o_i. We now remove i and o_i from consideration and proceed inductively; the structure of the optimum on the remaining players is unchanged, so we can apply Lemma 5.4.2 on this smaller instance and continue in this way, thus producing an ordering.

The remainder of the proof is similar to the analysis for the case of identical players: we simulate the ordering $i_1, i_2, \ldots, i_n$ using weights by choosing a sufficiently small $\varepsilon > 0$ and assigning player i_c (the c^{th} player in the order) a weight of $z_{i_c} = \varepsilon^c$. □

It is interesting to note that once we are in the regime of this more general model we can construct instances in which a Braess's paradox appears – adding an additional player reduces the welfare of the best Nash equilibrium. Consider the following 2-player (1 and 2) and 2-project (a and b) instance as an example. The weights of the projects are $w_a = 1$, $w_b = 3/8$ and the success probabilities are: $p_{1a} = p_{2a} = 0.5$, $p_{1b} = p_{2b} = 1$. For this 2-player instance the social welfare of the best Nash equilibrium, in which each player works on a different project is 7/8. If we add another player,3, such that $p_{3a} = 0, p_{3b} = \varepsilon$. We get that in any Nash equilibrium player 3 works on project b. However if player 3 is working on project b then both players 1 and 2 prefer to work on project a which leads to a social welfare of 3/4 of the best Nash equilibrium instead of social welfare of 7/8 without player 3.

CHAPTER 6

DYNAMIC MODELS OF REPUTATION AND COMPETITION

6.1 Introduction

In the previous chapter we considered how to allocate credit when players are myopic and only care about maximizing their current utility. However, sometimes players' reasoning process is more complex and also takes into account the effects of their current actions on the future. In this chapter we present a model capturing the effects of this long-range reasoning. We study this question in the context of markets for employment.

Markets for employment have been the subject of several large bodies of research, including the long and celebrated line of work on bipartite matching of employers to job applicants [88], sociological and economic approches to the process of finding a job [53, 74, 86], and many other frameworks. Recent work in theoretical computer science has modeled issues such as the competition among employers for applicants [57, 58] and hiring policies that take a firm's reputation into account [25].

Despite this history of research, there remain a number of fundamental issues in job-market matching that have gone largely unmodeled. One of these, familiar to anyone who has dealt with job markets in academia or related professions, is the feedback loop over multiple hiring cycles between the job candidates that a firm (or academic department) pursues and the evolution of its overall reputation. In particular, there is a basic trade-off at work: successfully recruiting higher-quality candidates can raise a firm's reputation, which in turn can make it more attractive to candidates in future hiring cycles; on the other hand, competing for these higher-quality candidates comes with a greater risk of emerging from a given hiring cycle empty-handed.

Here we formulate and study a strategic model that captures these issues; the firms in our model make recruiting decisions in a way that takes into account the probabilistic effect of these decisions on their reputations and hence their effectiveness at recruiting in future periods. We ask what effect these types of long-range strategies have, in the model, on outcomes for job candidates — do more or fewer people get employed when firms make use of this long-range reasoning? As the model demonstrates, there are two natural opposing forces at work here:

(i) A firm's desire to increase its reputation for the sake of future periods may cause it to compete for a stronger candidate and lose, when it could instead have hired a weaker candidate who now goes unemployed.

(ii) As one firm's reputation evolves, it may choose to stop competing with another firm of higher reputation, leading to implicit coordination that results in job offers to a larger set of people.

A firm's utility is the total quality of all the candidates it hires, and so our measure of social welfare — the sum of the firms' utilities — is simply the total quality of all candidates hired by any of them. We consider the natural performance guarantee question in this model: how does the social welfare under multi-period strategic behavior compare to the maximum social welfare attainable, where the maximum corresponds to a central authority that is able to impose a matching of candidates to firms? Essentially, this is a measure of how much talent the system is collectively able to employ, when it is governed by competition.

Our model, based on competition between two firms, is simple to state but leads to complex phenomena based on the trade-off between forces (i) and (ii) above. We obtain a tight bound of $\dfrac{2}{1+\sqrt{1.5}} \approx 0.898$ on the ratio of the social welfare under the canonical Nash equilibrium to the optimal social welfare in this model, as the number of periods goes to infinity. Studying this *performance ratio*[1] as a function of the number of periods, we find that for some settings of the parameters, the performance ratio is worse for instances with a "medium" number of periods, rather than those with very few (where force (i) does not have enough time to generate unemployment) or those with very many (where force (ii) takes over and ensures a high level of employment). The analysis develops interesting connections between multi-step strategic interaction with competition and Polya urn processes [84].

Formulating the Model

The subtleties discussed above emerge already in a very simple model of multi-period competition; we therefore focus on a highly reduced formulation that captures these issues yet permits a tight analysis. In particular, we study these effects in the case of just two players, and with a candidate pool that has the same structure in each time period. Our model can clearly be extended in ways that add complexity in a number of dimensions, and this suggests natural directions for further work on multi-period matching games with this structure. We discuss some of these directions briefly in the conclusions section (Section 6.6).

We set up the model as a game with two players over k rounds. We can think

[1] We use the neutral term "performance ratio" rather than *price of anarchy* or *price of stability* because — as we will see — our game has a natural equilibrium, and we are interested in the relative performance of this natural equilibrium, rather than necessarily focusing on the best or worst equilibrium.

of each player as representing an academic department that is able to try hiring one new faculty candidate in each of the next k hiring seasons. In each round $t \in \{1, 2, \ldots, k\}$, the players are presented with a set of job candidates with fixed numerical *qualities*. Since we have only two firms in our model, we will assume that the firms' hiring will only involve considering the two strongest candidates; we therefore assume that there are only two candidates available. Normalizing the quality of the stronger candidate, we define the qualities of the two candidates to be 1 and $q < 1$ respectively.

We want to be able to talk separately about a department's *utility* — the total quality of all candidates it has hired — and its *reputation* — its ability to attract new candidates based on the quality of the people it has hired. A number of studies of academic rankings have emphasized that departments are judged in large part by their strongest members; intuitively, this is why a smaller department with several "star" members can easily rank higher than a much larger department, and ranking schemas often include measures that focus on this distinction.

Given this, a natural way to define reputation in our model is to say that the reputation of firm i in round t, denoted $x_i(t)$, is equal to the number of higher-quality candidates (i.e. those of quality 1 rather than q) that it has hired so far. This is distinct from the utility of firm i in round t, denoted $u_i(t)$, which is simply the sum of the qualities of all the candidates it has hired.

We assume that a firm is seeking to maximize its utility over the full k rounds; *however*, note that since this is a multi-period game, and reputation determines success in future rounds of hiring, a firm's equilibrium strategy will in fact involve actions that are effectively seeking to increase reputation even at the expense of short-term sacrifices to expected utility. This, indeed, is exactly the type of behavior we hope to see in a model of recruiting.

Building on this discussion, we therefore structure the game as follows.

- Each player i has a numerical *reputation* $x_i(t)$ and *utility* $u_i(t)$ in round t. We will focus mainly on the case in which the two players each start with reputation equal to 1, though in places we will consider variations on this initial condition.
- In each round $t \in \{1, 2, \ldots, k\}$, player i chooses one of the candidates j to try recruiting; this choice of j constitutes the player's *strategy* in round t.
- If player i is the only one to try recruiting j, then j accepts the offer. If both players compete for the same candidate j, then j accepts player i's offer with probability proportional to player i's reputation. This follows the *Tullock contest function* that is standard in economic theory for modeling competition [94, 103], thus we have : player 1 hires j with probability $\dfrac{x_1(t)}{x_1(t) + x_2(t)}$ and player 2 hires j with probability $\dfrac{x_2(t)}{x_1(t) + x_2(t)}$. The player who loses this competition for candidate j hires no one in this round.
- Finally, each player receives a payoff in round t equal to the quality of the candidate hired in the round (if any). The player's utility is increased by the quality of the candidate it has hired; the player's reputation is increased by 1 if it has hired the stronger candidate in round t, and remains the same otherwise.

Thus the model captures the basic trade-off inherent in recruiting over multiple rounds — by competing for a stronger candidate, a player has the opportunity to increase its reputation by a larger amount, but it also risks hiring no one. The model is designed to arrive at this trade-off using very few underlying parameters;

but we believe that the techniques developed for the analysis suggest approaches to more complex variants, and we discuss some of these in the conclusions section (Section 6.6).

The maximum possible social welfare is achieved if the two players hire the top two candidates respectively in each round, achieving a social welfare of $k(1+q)$. The key question we consider here is what social welfare can be achieved in equilibrium for this k-round game, and how it compares to the welfare of the social optimum. In effect, how much does the struggle for reputation leave candidates unemployed?

The subgame perfect equilibria in this multi-round game are determined by backward induction — essentially, in a given round t, a player evaluates the possible values its utility and reputation can take in round $t+1$, after the (potentially probabilistic) outcome of its recruiting in round t. There are multiple equilibria, but there is a single natural class of *canonical equilibria* for the model, in which the higher-reputation player always goes after the stronger candidate, and — predicated on the equilibrium having this form in future rounds — the lower-reputation player makes an optimal decision to either compete for the stronger candidate or make an offer to the weaker candidate. (When the lower-reputation player is indifferent between these two options, we break the symmetry using the assumption that the lower player hires the weaker candidate.) Proving that this structure in fact produces an equilibrium is non-trivial; in part this is because reasoning about subgame perfect equilibria always involves some complexity due to the underlying tree of possibilities, but the present model adds to this complexity because the randomization involved in the outcome causes the possible trajectories of the game to "explore" most of this tree.

We study the behavior of this canonical equilibrium, and we define the *performance ratio* of an instance to be the ratio of total welfare between the canonical equilibrium and the social optimum.

Overview of Results

We first consider the performance ratio as a function of the number of rounds k. As an initial question, which choice of k yields the worst performance ratio? When $q < \frac{1}{2}$, the answer is simple: for $k = 1$, the players necessarily compete in the one round they have available, and this yields a performance ratio of $1/(1+q)$ — as small as possible. When $q > \frac{1}{2}$, however, the situation becomes more subtle. For $k = 1$, the players do not compete in the canonical equilibrium, and so the performance ratio for $k = 1$ is 1. At the other end of the spectrum, when $q \geq \frac{1}{2}$, the two players will eventually stop competing with probability 1 and the performance ratio converges up to 1 when k becomes large. But in between, the performance ratio can be larger than at both extremes; in particular, when the quantity $\dfrac{q}{1-q}$ approaches an integer value k from below, we show that the performance ratio is maximized when the number of rounds takes this intermediate value k.

We then turn to the main result of the chapter, which is to analyze the performance ratio in the limit as the number of rounds k goes to infinity. When $q \geq \frac{1}{2}$, as just noted, we show that the two players will eventually stop competing with probability 1 and the performance ratio converges to 1. But when $q < \frac{1}{2}$, something more complex happens: there is a positive probability, strictly between 0 and 1, that the players compete forever. This has a natural interpretation — as reputations evolve, the two players can settle into relative levels of reputation under which it is worthwhile for the lower player to compete for the stronger candidate; but it may also happen that after a finite number of rounds, one player decides that it is too weak to continue competing for the stronger candidate, and it begins to act on its second-tier status. What is interesting is that each of these outcomes has a positive probability of occurring.

The possibility of indefinite competition leads to a non-trivial performance ratio; we show that the worst case occurs when $q = \sqrt{1.5} - 1 \approx .2247$, with a performance ratio of $\frac{2}{1+\sqrt{1.5}} \approx 0.898$. We also show that the performance ratio converges to 1 as q goes either to 0 or to 1. Our analysis proceeds by defining an urn process that tracks the evolution of the players' reputations; this is a natural connection to develop, since urn processes are based on models in which probabilities of outcomes in a given step — the result of draws from an urn — are affected by the realized outcomes of draws in earlier steps. We provide more background about urn processes in the next section. Informally speaking, the fact that a player might compete for a while and then permanently give up in favor of an alternative option is also reminiscent of strategies in the multi-armed bandit problem, where an agent may experiment with a risky option for a while before permanently giving up and using a safer option; later in the chapter, we make this analogy more precise as well. To make use of these connections, we study a sequence of games that begins with players who are constrained to follow a set sequence of decisions for a long prefix of rounds, and we then successively relax this constraint until we end up with the original game in which players are allowed to make strategic decisions from the very beginning.

In Section 6.5, we also consider variants of the model in which one changes the function used for the success probabilities in the competition between the two players for a candidate. Note that the way in which competition is handled is an implicit reflection of the way candidates form preferences over firms based on their reputations, and hence varying this aspect of the model allows us to explore different ways in which candidates can behave in this dimension. In particular, we consider a variation on the model in which — when the two players compete for a candidate — the lower-reputation player succeeds with a fixed probability $p < \frac{1}{2}$ and the higher-reputation player succeeds with probability $1-p$. This model thus captures the long-range competition to become the higher-reputation player using an extremely simple model of competition within each round. The main result here is that for all $p < q$, the performance ratio converges to 1; the analysis makes use of biased random walks in place of urn processes to analyze the long-term competition between the players.

Further Related Work

As noted above, there has been recent theoretical work studying the effect of reputation and competition in job markets. Broder et al. consider hiring strategies designed to increase the average quality of a firm's employees [25]. Our focus here is different, due to the feedback effects from future rounds that our model of competition generates: a few weak initial hires can make it very difficult for a player to raise its quality later, while a few strong initial hires can make the process correspondingly much easier. Immorlica et al. consider competition between employers, though in a quite different model where candidates are presented one at a time as in the *secretary problem* [57, 58], and each player's goal is to hire a candidate that is stronger than the competitor's. They do not incorporate the spillover of this competition into future rounds.

Our work can also be viewed as developing techniques for analyzing the performance ratio and/or price of anarchy in settings that involve dynamic matchings — when nodes on one side of a bipartite graph must make strategic decisions about matchings to nodes that arrive dynamically to the other side of the graph. In the context of job matching, Shimer and Smith consider a dynamic matching model of a labor market in which the central constraint is the cost of searching for potential partners [93]. Haeringer and Wooders apply dynamic matching to the problem of sequential job offers over time [54], but in a setting that considers the sequencing of offers in a single hiring cycle; this leads to different questions, since the consequence for reputation in future hiring cycles is not in the scope of their investigation. Dynamic matchings have also been appearing in a number of other recent application contexts (e.g. [37, 105]), and there are clearly many unresolved questions here about the cost of strategic behavior.

Finally, our model can be abstracted beyond the setting of job-market matching to capture essentially any context in which two firms must decide over multiple rounds whether to compete or to make use of a private outside option. There are many domains that exhibit this general structure, and it would be interesting to see whether our techniques can be adapted to some of these other situations. For example, this issue has been explored — by different means — in the context of product compatibility [28]. The issue of whether a weaker competitor decides to directly compete or give up in favor of an alternative option is also implicit in studies of the branding and advertising decisions firms make — including whether to explicitly acknowledge a second-place status, as for example the Avis car rental company did in its "We Try Harder" campaign [75].

6.2 Preliminaries

An instance of the recruiting game, as described in the introduction, is defined by the initial reputations x_1 and x_2 of the two players; the relative quality q of the weaker candidate compared to the stronger one; and the number of rounds k. Accordingly, we denote an instance of the game by $G_{k,q}(x_1, x_2)$. Generally q will be clear from context, and so we will also refer to this game as simply $G_k(x_1, x_2)$. We will refer to the player of higher reputation as the *higher player*, and the player of lower reputation as the *lower player*. In case the players have the same reputation we will refer to player 1 as the higher player.

The game as defined is an extensive-form game, and as such it can admit many subgame perfect equilibria. For example, it is easy to construct a single-round game in which it is an equilibrium for the lower player to try to recruit the stronger candidate and for the higher player to go after the weaker candidate. This equilibrium clearly has a less natural structure than one in which the higher player goes after the stronger option; to avoid such pathologies, as noted in the introduction, we will study multi-round strategies $s_k(x_1, x_2)$ that are defined as follows:

Definition 6.2.1 *Denote by $s_k(x_1, x_2)$ the following strategies for the players over the k rounds: in every round the higher player goes for the stronger candidate and the lower player best-responds by choosing the candidate that maximizes its utility, taking into account the current round and all later rounds by induction. For $s_k(x_1, x_2)$ to be well-defined we make the following two assumptions: (1) If the lower player is indifferent between going for the stronger candidate and the weaker candidate we assume it chooses to go for the weaker candidate. (2) If the two* *players have the same reputations we break ties in favor of player* 1.

The strategies $s_k(x_1, x_2)$ can be summarized essentially by saying that in every round of the game, first the higher player gets to make an offer to its preferred candidate, and given this decision the lower player makes the choice maximizing its utility. To show that the strategies $s_k(x_1, x_2)$ form a sub-game perfect equilibrium we will show inductively that in every round it is in the higher player's best interest to make an offer to the stronger candidate. More formally we denote the strategy of making an offer to the stronger candidate in some round by $+$ and to the weaker candidate by $-$. We define $f(s_k(x_1, x_2))$ to be the pair of strategies that the players use in the first round of $s_k(x_1, x_2)$.

We denote player i's utility when the two players play the strategies prescribed by $s_k(x_1, x_2)$ by $u_i(s_k(x_1, x_2))$. We now formally write down the utility of the players in $s_k(x_1, x_2)$ based on the value of $f(s_k(x_1, x_2))$:

- If $f(s_k(x_1, x_2)) = \langle +, + \rangle$ then

$$u_1(s_k(x_1, x_2)) = \frac{x_1}{x_1 + x_2}(1 + u_1(s_{k-1}(x_1 + 1, x_2))) + \frac{x_2}{x_1 + x_2} u_1(s_{k-1}(x_1, x_2 + 1))$$
$$u_2(s_k(x_1, x_2)) = \frac{x_1}{x_1 + x_2} u_2(s_{k-1}(x_1 + 1, x_2)) + \frac{x_2}{x_1 + x_2}(1 + u_2(s_{k-1}(x_1, x_2 + 1)))$$

- If $f(s_k(x_1, x_2)) = \langle +, - \rangle$ then

$$u_1(s_k(x_1, x_2)) = 1 + u_1(s_{k-1}(x_1 + 1, x_2))$$
$$u_2(s_k(x_1, x_2)) = q + u_2(s_{k-1}(x_1 + 1, x_2))$$

- If $f(s_k(x_1, x_2)) = \langle -, + \rangle$ then

$$u_1(s_k(x_1, x_2) = q + u_1(s_{k-1}(x_1, x_2 + 1))$$
$$u_2(s_k(x_1, x_2)) = 1 + u_2(s_{k-1}(x_1, x_2 + 1))$$

We denote the social welfare when the players are using strategies $s_k(x_1, x_2)$ by

$$u(s_k(x_1, x_2)) = u_1(s_k(x_1, x_2)) + u_2(s_k(x_1, x_2)).$$

Even though it is natural to suspect that the strategies $s_k(x_1, x_2)$ are indeed a sub-game perfect equilibrium, proving that this is the case is not such a simple task. The first step in showing that the strategies $s_k(x_1, x_2)$ are a sub-game perfect equilibrium, and a useful fact by itself, is the monotonicity of the players' utilities $u_i(s_k(x_1, x_2))$. More formally, in Section 6.7 of the appendix we show that:

Claim 6.2.2 *For any x_1, x_2, and $\varepsilon > 0$:*

1. $u_1(s_k(x_1 + \varepsilon, x_2)) \geq u_1(s_k(x_1, x_2)) \geq u_1(s_k(x_1, x_2 + \epsilon))$.
2. $u_2(s_k(x_1, x_2 + \varepsilon)) \geq u_2(s_k(x_1, x_2)) \geq u_2(s_k(x_1 + \epsilon, x_2))$.

Next, we prove that the three following statements hold.

Proposition 6.2.3 *For any integers x_1, x_2 and k the following holds for the strategies $s_k(x_1, x_2)$.*

1. *$s_k(x_1, x_2)$ is a sub-game perfect equilibrium in the game $G_k(x_1, x_2)$.*
2. *If a player does not compete in the first round of the game $G_k(x_1, x_2)$, then it does not compete in all subsequent rounds.*
3. *The utility of the higher player in the game $G_k(x_1, x_2)$ is at least as large as the utility of the lower player.*

Essentially, we prove all three properties simultaneously by induction on the number of rounds of the game; to make the inductive argument easier to follow, we prove each of the statements separately in the Appendix.

Let us mention two more claims that we will make use of later on and are proven in the Appendix:

Claim 6.2.4 *If $u_i(s_k(x_1, x_2)) = kq$, for some player i, then player i never competes in the game $G_k(x_1, x_2)$.*

Claim 6.2.5 *If player i competes in the first round of the game $G_k(x_1, x_2)$ and wins, then in the next round of the game it also makes an offer to the stronger candidate.*

Connections to Urn Processes

Note that since each player's reputation is equal to the number of stronger candidates it has hired, the reputations are always integers (assuming they start from integer values). These integer values evolve while the players are competing; and once they stop competing, we know by statement (2) of Proposition 6.2.3 the exact outcome of the game. This brings us to the close resemblance between our recruiting game and a Polya Urn process [84].

First, let us define what the Polya Urn process is:

Definition 6.2.6 (Polya Urn process) *Consider an urn containing b blue balls and r red balls. The process is defined over discrete rounds. In each round a ball is sampled uniformly at random from the urn; hence the probability of drawing a blue* *ball is $\frac{b}{b+r}$ and the probability of drawing a red ball is $\frac{r}{b+r}$. Then, the ball together with another ball of the same color are returned to the urn.*

There is a clear resemblance between our recruiting game and the urn model. As long as the players compete, their reputations evolve in the same way as the number of blue and red balls in the urn, since the probabilistic rule for a candidate to select which firm to join is the same as the rule for choosing which color to add to the urn, and by assumption the reputation of the winning player is increased by the stronger candidate's quality, which is 1.

A striking fact about urn models is that the fraction of the blue (or red) balls converges in distribution as the number of rounds goes to infinity. More specifically, if initially the urn contains a single red ball and a single blue ball then the fraction of blue balls converges to a uniform distribution on $[0, 1]$ as the number of rounds goes to infinity. More generally, the fraction of blue balls converges to the β distribution $\beta(b, r)$. Understanding urn processes is useful for understanding our proofs; however we should stress that our model and its analysis have added complexity due to the fact that players stop competing at a point in time that is strategically determined.

Connections to Bandit Problems

It is interesting to note that as long as the lower player stays lower our equilibrium selection rule makes this effectively a one-player game. In a sense, the lower player's strategy in this phase resembles the optimal strategy in a mulit-armed bandit problem [45], and more specifically in a one-armed bandit problem [16]. In a one-armed bandit a single player is repeatedly faced with two options (known as "arms" following the terminology of slot machines): the player can pull arm 1, which gives a reward sampled from some *unknown* distribution, or pull arm 2 which gives him a reward from a *known* distribution. Informally speaking, by pulling arm 1 the player gets both a reward and some information about the distribution from which the reward is drawn. The player's goal is to maximize its expected reward possibly under some discounting of future rounds. A celebrated result establishes that for some types of discounting (for example geometric) one can compute a number called the *Gittins index* for each arm (based on one's observations and the prior) and the strategy maximizing the player's expected reward is to pull the arm with the highest Gittins index in each round [45]. Since by definition the Gittins index of the fixed arm is fixed, this implies that once the Gittins index of the unknown arm drops below the one of the known arm, the player should only pull the known arm. This also means that the player stops collecting information on the distribution of the unknown arm and hence from this round onwards it always chooses the fixed arm.

There are analogies as well as differences between our game and the one-armed bandit problem. In our game, the lower player is also faced with a choice between a risky option (competing) and a safe option (going for the weaker candidate). On the other hand, an important difference between our model and the one-armed bandit problem is that our game is in fact a two-player game and at any point the lower-reputation player can become the higher-reputation one; this property contributes additional sources of complexity to the analysis of our game. Moreover, it is important to note that for many distributions and discount sequences (including the ones most similar to our game) a closed-form expression of the Gittins index is unknown.

6.3 Analyzing the Game with a Fixed Number of Rounds

We begin by analyzing the game played over a fixed number of rounds k and study the dependence of the performance ratio on k. In the next section, we turn to the main result of the chapter, which is to analyze the limit of the performance ratio as the number of rounds k goes to infinity.

Our first result is a simple but powerful bound of $\frac{2q}{1+q}$ on the performance ratio, which holds for all k. This is done by relating the performance ratio to players' decision whether to compete in the first round. The argument underlying this relationship is quite robust, in that it is essentially based only on the reasoning that the players can always decide to stop competing and go for the weaker candidate. Note that this bound also implies that as q goes to 1 the performance ratio also goes to 1.

Claim 6.3.1 *The performance ratio of any game $G_{k,q}(x_1, x_2)$ is at least $\frac{2q}{1+q}$.*

Proof: We begin with the simple observation that the expected social welfare equals the sum of the expected utilities of the two players in the beginning of the game. To get a lower bound on the performance ratio it is enough to compute an upper bound on the expected social welfare. This is done by observing that $u_i(s_k(x_1, x_2)) \geq kq$, since a player can always secure a utility of kq by always making an offer to the weaker candidate. Hence, the following is a bound on the performance ratio: $\dfrac{u_1(s_k(x_1,x_2)) + u_2(s_k(x_1,x_2))}{k(1+q)} \geq \dfrac{2kq}{k(1+q)} = \dfrac{2q}{1+q}$. $\square$

Corollary 6.3.2 *The performance ratio of any game $G_{k,q}(x_1, x_2)$ is at least $2/3$. This holds for $q > 1/2$ since $\dfrac{2q}{1+q} > 2/3$ and for $q \leq 1/2$ since the performance ratio is trivially lower-bounded by $1/(1+q) \geq 2/3$.*

Next, we ask what is the length of a game for which the worst performance ratio is achieved. For $q < 1/2$, this is simply a single-round game. However, for $q > 1/2$ the answer is not so simple. We show that when $\frac{q}{1-q} + \varepsilon$ is an integer for an arbitrarily small $\varepsilon > 0$, a game of $k_q = \frac{q}{1-q} + \varepsilon$ rounds exhibits a performance ratio arbitrarily close to $\frac{2q}{1+q}$. It is interesting that the players' strategies in the games achieving this maximum performance ratio have a very specific structure – the players compete just for the first round and then the player who lost goes for the weaker candidate for the rest of the game.

Proposition 6.3.3 *Let* $\varepsilon = \lceil \frac{q}{1-q} \rceil - \frac{q}{1-q}$ *and* $k_q = \frac{q}{1-q} + \varepsilon$. *Then, as* ϵ *approaches* 0 *from above (remaining strictly positive), the performance ratio of the game* $G_{k_q,q}(x,x)$ *converges to* $\frac{2q}{1+q}$.

Proof: Observe that by Claim 6.3.4 below the players in the game $G_{k_q,q}(x,x)$ compete for the first round (since $\varepsilon > 0$) and then completely stop competing. Thus the expected social welfare of the canonical equilibrium is $k + (k-1)q$ and its performance ratio is:

$$\frac{1+(k-1)(1+q)}{k(1+q)} = \frac{1+((\frac{q}{1-q}+\varepsilon)-1)(1+q)}{(\frac{q}{1-q}+\varepsilon)(1+q)} = \frac{2q^2+\varepsilon-\varepsilon q^2}{q+q^2+\varepsilon-\varepsilon q^2}.$$

It is not hard to see now that as ϵ approaches 0 the performance ratio approaches $\dfrac{2q}{1+q}$. □

We now prove for the k_q's discussed in the previous proposition the players indeed compete only for the first round and then stop competing. More formally we prove:

Claim 6.3.4 *In the game* $G_{k,q}(x,x)$ *for* $\frac{q}{1-q} < k \leq \frac{1}{1-q}$ *the players compete in the first round and then completely stop competing.*

Proof: Player 2 (which is the lower player in the game) competes in the game $G_{k,q}(x,x)$ if:

$$\frac{1}{2}(1+u_2(s_{k-1}(x,x+1)))+\frac{1}{2}u_2(s_{k-1}(x+1,x)) > q+u_2(s_{k-1}(x+1,x)).$$

After some rearranging we get that this implies that player 2 competes if:

$$1+u_2(s_{k-1}(x,x+1)) > 2q+u_2(s_{k-1}(x+1,x)).$$

Note that $k \leq \frac{1}{1-q} \implies q \geq \frac{k-1}{k}$. Thus, by Claim 6.3.5 below we have that for any x_1, x_2 the players in the game $G_{k-1,q}(x_1,x_2)$ do not compete. This implies that $u_2(s_{k-1}(x,x+1)) = k-1$ and $u_2(s_{k-1}(x+1,x)) = (k-1)q$. Thus, the players in the game $G_{k,q}(x,x)$ compete if $k > (k+1)q$ implying $\frac{q}{1-q} < k$ as required. □

Finally we prove:

Claim 6.3.5 *If $q \geq \frac{k}{k+1}$ then the players in the game $G_{k,q}(x_1,x_2)$ never compete.*

Proof: Let $x_2 \leq x_1$. Player 2 competes in the game $G_{k,q}(x_1,x_2)$ if:

$$\frac{x_2}{x_1+x_2}(1+u_2(s_{k-1}(x_1,x_2+1)))+\frac{x_1}{x_1+x_2}u_2(s_{k-1}(x_1+1,x_2)) > q+u_2(s_{k-1}(x_1+1,x_2)).$$

After some rearranging we get that player 2 competes if:

$$1+u_2(s_{k-1}(x_1,x_2+1)) > \frac{x_1+x_2}{x_2}q+u_2(s_{k-1}(x_1+1,x_2)).$$

Observe that $u_2(s_{k-1}(x_1,x_2+1)) \leq k-1$ as this is the maximum utility a player can get in a $(k-1)$-round game. Also observe that $u_2(s_{k-1}(x_1+1,x_2)) \geq (k-1)q$ and that by assumption $\frac{x_1+x_2}{x_2} \geq 2$. Thus, we have that a necessary condition for player 2 to compete is that $k > (k+1)q$. This implies that for $q \geq \frac{k}{k+1}$ player 2 does not compete in the first round of the game $G_{k,q}(x_1,x_2)$. By part (2) of Proposition 6.2.3 we have that if a player does not compete in the first round of the game it also does not compete in all subsequent rounds which completes the proof. A very similar proof works for the case that player 1 is the lower player. □

6.4 Analyzing the Long-Game Limit

We now turn to the main question in the chapter, which is the behavior of the performance ratio in the limit as the number of rounds goes to infinity.

Our main result here is that as k goes to infinity the performance ratio of the game $G_k(1,1)$ goes to $\dfrac{1+2qr}{1+q}$, where $r = \min\{q, \frac{1}{2}\}$. In particular for $q < 1/2$ this implies that as k goes to infinity the performance ratio goes to $\dfrac{1+2q^2}{1+q}$. This function attains its minimum when $q = \sqrt{1.5} - 1 \approx .2247$ and at this point it has a value of $\frac{2}{1+\sqrt{1.5}} \approx 0.898$. For $q \geq 1/2$, on the other hand, this simply implies that as k goes to infinity the performance ratio of the game $G_k(1,1)$ goes to 1. Defining $r = \min\{q, \frac{1}{2}\}$ helps us to present a single unified proof both for $q < 1/2$ and for $q \geq 1/2$.

The proof of this theorem becomes somewhat involved even though its main idea is quite natural. Intuitively speaking, we know that as long as the players compete, our game proceeds the way an urn process does. This means that the probability that player 2, for example, is the one to hire the stronger candidate converges to a uniform distribution as the number of rounds k the players compete goes to infinity. Henceforth, we will also refer to this probability as player's 2 *relative reputation*. We show that if the relative reputation of one of the players converges to a number smaller than r, then after a fairly small number of rounds – specifically $\theta(\ln(k))$ – the players stop competing. The probability that the relative reputation of one of the players converges to something less than r is simply $2r$. Therefore, the expected social welfare of our canonical equilibrium converges to $k + 2qr(k - \theta(\ln(k)))$ and the performance ratio converges to $\frac{1+2qr}{1+q}$.

We divide the proof to four subsections. In Subsection 6.4.1 we introduce t-

binding games, which give us a formal way to study games in which the two players compete for at least the first t rounds. By showing that the utilities of the players in our game are at least as large as their utilities in the t-binding game we reduce our problem to showing that the expected utility in a t-binding game is "large enough". This is done in Subsection 6.4.3. The proof relies on Subsection 6.4.2 which, loosely speaking, shows that if after t rounds of competition the relative reputation of the lower player is non-trivially smaller than r then the lower player stops competing. Finally, in Subsection 6.4.4 we state the formal theorem and wrap up the proof.

6.4.1 t-Binding Games

A recruiting game is t-*binding* if in the first t rounds the two players are required to compete for the stronger candidate. We denote a t-binding game by $G_k^t(x_1, x_2)$. We also denote by $s_k^t(x_1, x_2)$ the canonical equilibrium of the game $G_k^t(x_1, x_2)$ in which the players compete for the first t rounds and then follow the strategies $s_{k-t}(x'_1, x'_2)$ in the resulting game. Denote by $u(s_k^t(x_1, x_2))$ the expected social welfare of the canonical equilibrium in the game $G_k^t(x_1, x_2)$. It is intuitive to suspect that making the players compete for the first t rounds can only decrease their utility. In the next lemma we prove that this intuition is indeed correct:

Lemma 6.4.1 *The expected social welfare of the game $G_k(1,1)$ is greater than or equal to the expected social welfare of the game $G_k^t(1,1)$; that is, $u(s_k(1,1)) \geq u(s_k^t(1,1))$.*

Proof: We prove the lemma by proving a stronger claim:

Claim 6.4.2 *The expected utility of each of the players in the game* $G_k^t(1,1)$ *for* $0 \leq t < k$ *is greater than or equal to their expected utility in the game* $G_k^{t+1}(1,1)$.

Proof: For simplicity we prove the claim for player 2; however the claim holds for both players. By definition, in the game $G_k^t(1,1)$ the players compete for at least the first t rounds. During this phase of competition, the two players' reputations evolve according to the update rule for a standard Polya urn process, as described in Section 6.2. A standard result on that process implies that at the end of these t rounds with probability $\frac{1}{t+1}$ player 1 has a reputation of $1+t-i$ and player 2 has a reputation of $1+i$ for $0 \leq i \leq t$. Thus, we have that:

$$u_2(s_k^t(1,1)) = \frac{1}{t+1}\sum_{i=0}^{t} u_2(s_{k-t}(1+t-i, 1+i))$$

Let $I_\delta = \{i | f(s_{k-t}(1+t-i, 1+i)) = \delta\}$ for $\delta \in \{\langle +,+\rangle, \langle +,-\rangle, \langle -,+\rangle\}$. For example, $I_{\langle +,+\rangle}$ is the set of all indices i for which the players compete in the first round of the game $G_{k-t}(1+t-i, 1+i)$.

We can now write the sum, usefully, as

$$\begin{aligned} u_2(s_k^t(1,1)) = & \frac{1}{t+1}\sum_{i \in I_{\langle +,+\rangle}} u_2(s_{k-t}(1+t-i, 1+i)) + \frac{1}{t+1}\sum_{i \in I_{\langle +,-\rangle}} u_2(s_{k-t}(1+t-i, 1+i)) \\ & + \frac{1}{t+1}\sum_{i \in I_{\langle -,+\rangle}} u_2(s_{k-t}(1+t-i, 1+i)) \end{aligned}$$

By this partition:

- For $i \in I_{\langle +,+\rangle}$, we have $u_2(s_{k-t}(1+t-i, 1+i)) = u_2(s_{k-t}^1(1+t-i, 1+i))$ – since in both of these games the two players compete in the first round.

- For $i \in I_{\langle +,- \rangle}$, we have $u_2(s_{k-t}(1+t-i, 1+i)) \geq u_2(s^1_{k-t}(1+t-i, 1+i))$ – since $u_2(s^{\langle +,- \rangle}_{k-t}(1+t-i, 1+i)) \geq u_2(s^{\langle +,+ \rangle}_{k-t}(1+t-i, 1+i))$. (in the first round of the game player 2 prefers going after the weaker candidate over competing).

- For $i \in I_{\langle -,+ \rangle}$, we have $u_2(s_{k-t}(1+t-i, 1+i)) > u_2(s^1_{k-t}(1+t-i, 1+i))$ – since $u_2(s^{\langle -,+ \rangle}_{k-t}(1+t-i, 1+i)) > u_2(s^{\langle +,+ \rangle}_{k-t}(1+t-i, 1+i))$ by monotonicity.

Thus, we have $u_2(s^t_k(1,1)) \geq \frac{1}{t+1} \sum_{i=0}^{t} u_2(s^1_{k-t}(1+t-i, 1+i)) = u_2(s^{t+1}_k(1,1))$.

$\square$ $\square$

6.4.2 When does the lower player stop competing?

This next phase of our analysis is composed of two parts: in the first part we show that the utility of the lower player in a k-round game is upper bounded by $\max\{b_q(k,t,x), kq\}$ for some function b to be later defined. In the second part we compute the conditions under which $b_q(k,t,x) < kq$ which implies that under the same conditions the lower player in the game stops competing.

For this subsection we denote player 1's reputation by $t - x$ and player 2's reputation by x. Both statements below also hold for player 1 and the game $G_k(x, t-x)$.

The following notation will be useful for our proofs:

- $f_q(i,t) = \binom{t}{i} q^i (1-q)^{t-i}$ – probability mass function for the binomial distribution with t trials.

- $F_q(x,t) = \sum_{i=0}^{x} \binom{t}{i} q^i (1-q)^{t-i}$ – cumulative distribution function for an integer x.

To understand the intuition behind the upper bound function $b_q(k,t,x)$ it is useful to look at an alternative description of the urn process. Under this description, we have a coin whose bias is sampled from a uniform distribution on $[0,1]$; then in each round the coin is tossed. If the coin turns up heads a blue ball is added to the urn; otherwise a red ball is added to the urn. Under this alternative description we can think of our lower player as trying to toss this coin (i.e. competing) in the hope that its bias is greater than r (recall that $r = \min\{q, \frac{1}{2}\}$). We refer to the event in which the bias of the coin is greater than r as a good event, and the event it is not a bad event. To upper-bound the player's utility we assume that if the good event happens the player wins the stronger candidate for all subsequent rounds and hence its utility is k. If the bad event happens then the player completely stops competing and thus its utility is $(k-1)q$.

Hence, the function that we use to upper bound the player's utility is:

$$\begin{aligned} b_q(k,t,x) &= \frac{x}{t} + 3F_r(x,t)k + (1 - 3F_r(x,t))(k-1)q \\ &= \frac{x}{t} + (k-1)q + 3F_r(x,t) \cdot \big((k-1)(1-q) + 1\big) \end{aligned}$$

We show that $\max\{b_q(k,t,x), kq\}$ is indeed an upper bound on the players' utility as the previous intuition suggests.

Lemma 6.4.3 *For any k, x and $t > \frac{4\ln(1/12)}{\ln(1-r)}$, we have $u_2(s_k(t-x,x)) \leq \max\{b_q(k,t,x), kq\}$.*

Proof: We divide the proof into two cases. When, $r \leq \dfrac{x+1}{t+1}$ the bound we need to prove is very loose and hence we can prove it directly. However, for $r > \dfrac{x+1}{t+1}$

proving this bound is more tricky and for this we use an induction that some times relies on the first case. Next we provide proofs for these two cases:

Claim 6.4.4 *For any k, x and $t > \frac{4\ln(1/12)}{\ln(1-r)}$ such that $r \leq \dfrac{x+1}{t+1}$, $u_2(s_k(t-x,x)) \leq \max\{b_q(k,t,x), kq\}$*

Proof: We actually prove a stronger claim, which is that for $t > \frac{4\ln(1/12)}{\ln(1-r)}$, $b_q(k,t,x) \geq k$. The utility of any player in a game of k rounds is at most k and hence this will be enough to complete the proof. To do this we need to show that $3F_r(x,t) \geq 1$ for $t > \frac{4\ln(1/12)}{\ln(1-r)}$. By [55] we have that the median of a binomial distribution is at distance of at most $\ln(2)$ from its mean. Thus, if $x \geq rt + \ln(2)$ we are done, as in this case $3F_r(x,t) \geq \frac{3}{2}$. Else, $rt + r - 1 \leq x < rt + \ln(2)$. This implies that $x + 2 \geq rt + \ln(2)$, which in turn implies that $F_r(x+2,t) \geq \frac{1}{2}$. Since $F_r(x+2,t) = F_r(x,t) + f_r(x+1,t) + f_r(x+2,t)$, what left to show is that $f_r(x+1,t) + f_r(x+2,t) < \frac{1}{6}$:

$$\begin{aligned} f_r(x+i,t) &= \binom{t}{x+i} r^{x+i}(1-r)^{t-(x+i)} \leq \big(\frac{x+i}{t}\big)^{-(x+i)} r^{x+i}(1-r)^{t-x-i} \\ &\leq r^{-(x+i)} r^{x+i}(1-r)^{t-x-i} = (1-r)^{t-x-i} \leq (1-r)^{\frac{1}{2}t-3} \end{aligned}$$

The last transition is due to the fact that $x < rt + \ln(2)$, $r \leq 1/2$ and $i \leq 2$. Thus we have that $f_r(x+1,t) + f_r(x+2,t) \leq 2(1-r)^{\frac{1}{2}t-3}$. To compute when $2(1-r)^{\frac{1}{2}t-3} < \frac{1}{6}$ we take a natural logarithm and get that: $(\frac{1}{2}t - 3) \cdot \ln(1-r) < \ln(\frac{1}{12})$, hence it is not hard to see that the claim holds for $t > \frac{4\ln(1/12)}{\ln(1-r)}$. □

Claim 6.4.5 *For any k, x and $t > \frac{4\ln(1/12)}{\ln(1-r)}$ such that $r > \dfrac{x+1}{t+1}$, $u_2(s_k(t-x,x)) \leq \max\{b_q(k,t,x), kq\}$*

Proof: Recall that $b_q(k,t,x) = \frac{x}{t} + (k-1)q + 3F_r(x,t) \cdot \big((k-1)(1-q) + 1\big)$. We prove the claim by induction. We first observe that the claim holds for $k = 1$.

Notice that by the assumption that $\frac{x+1}{t+1} < r \leq \frac{1}{2}$ we have that player 2 is the lower player. Thus, $u_2(s_1(t-x,x)) \leq \max\left\{\frac{x}{t}, q\right\}$ and the claim holds. Next, we assume correctness for $k-1$ rounds and prove for k. If $u_2(s_k(t-x,x)) = kq$, then the induction hypothesis holds and we are done. Otherwise, we have that $u_2(s_k(t-x,x)) > kq$, this immediately implies that $f(s_k(t-x,x)) = \langle +,+\rangle$.

By Claim 6.2.5 this implies that $u_2(s_{k-1}(t-x,x+1)) > (k-1)q$. Hence, either by the induction hypothesis (if $r > \frac{x+2}{t+2}$) or by Claim 6.4.4 (if $r \leq \frac{x+2}{t+2}$) we have that $u_2(s_{k-1}(t-x,x+1)) \leq b_q(k-1,t+1,x+1)$. Since $\frac{x+1}{t+2} < r$. We can also use the induction hypothesis to get that $u_2(s_{k-1}(t-x+1,x)) \leq \max\{b_q(k-1,t+1,x), kq\}$.

To show that $u_2(s_k(t-x,x)) \leq \max\{b_q(k,t,x), kq\}$ we now distinguish between two cases:

1. $u_2(s_{k-1}(t-x+1,x)) \leq b_q(k-1,t+1,x)$:

$$
\begin{aligned}
u_2(s_k(t-x,x)) &\leq \frac{x}{t}\left(1 + b_q(k-1,t+1,x+1)\right) + \frac{t-x}{t} b_q(k-1,t+1,x) \\
&= \frac{x}{t} + \frac{x}{t}\cdot\frac{x+1}{t+1} + \frac{t-x}{t}\cdot\frac{x}{t+1} + (k-2)q + 3F_r(x,t+1)\cdot\big((k-2)(1-q)+1\big) \\
&\quad + 3\frac{x}{t}\cdot f_r(x+1,t+1)\cdot\big((k-2)(1-q)+1\big) \\
&\leq^{(1)} \frac{2x}{t} + (k-2)q + 3F_r(x,t)\cdot\big((k-2)(1-q)+1\big) \\
&\leq^{(2)} \frac{x}{t} + (k-1)q + 3F_r(x,t)\cdot\big((k-1)(1-q)+1\big) = b_q(k,t,x)
\end{aligned}
$$

Transition (1) is obtained by applying Claim 6.4.6 (below) and some rearranging. For transition (2) we use the fact that $\frac{x}{t} < \frac{x+1}{t+1} < r \leq q$.

2. $u_2(s_{k-1}(t-x+1,x)) = (k-1)q$:

$$
\begin{aligned}
u_2(s_k(t-x,x)) &\leq \frac{x}{t}\left(1+b_q(k-1,t+1,x+1)\right) + \frac{t-x}{t}(k-1)q \\
&= \frac{x}{t} + \frac{x}{t}\cdot\frac{x+1}{t+1} + \frac{x}{t}(k-2)q \\
&\quad + \frac{x}{t}\cdot 3F_r(x+1,t+1)\cdot\big((k-2)(1-q)+1\big) + \frac{t-x}{t}(k-1)q \\
&= \frac{x}{t} + \frac{x}{t}\cdot\frac{x+1}{t+1} + \frac{t-x}{t}q + (k-2)q + \frac{x}{t}\cdot 3F_r(x+1,t+1)\cdot\big((k-2)(1-q)+1\big) \\
&\leq \frac{x}{t} + (k-1)q + \frac{x}{t}\cdot 3F_r(x+1,t+1)\cdot\big((k-2)(1-q)+1\big)
\end{aligned}
$$

For the last transition we use the fact that $\frac{x}{t}\cdot\frac{x+1}{t+1} < \frac{x}{t}\cdot q$ since by assumption we have that $\frac{x+1}{t+1} < r \leq q$.

Notice that $\frac{x}{t}F_r(x+1,t+1) < F_r(x,t+1) + \frac{x}{t}f_r(x+1,t+1)$. Hence by applying Claim 6.4.6 (below) we get that $\frac{x}{t}F_r(x+1,t+1) < F_r(x,t)$ which completes the proof.

□

Claim 6.4.6 *If $x < t$ then, $F_q(x,t+1) + \frac{x}{t}f_q(x+1,t+1) \leq F_q(x,t)$*

Proof:

$$
\begin{aligned}
F_r(x,t+1) &= \sum_{i=0}^{x} \binom{t+1}{i} q^i (1-q)^{t+1-i} \\
&= (1-q) \sum_{i=0}^{x} \frac{t+1}{t+1-i} \binom{t}{i} q^i (1-q)^{t-i} \\
&= (1-q) \sum_{i=0}^{x} (1 + \frac{i}{t+1-i}) \binom{t}{i} q^i (1-q)^{t-i} \\
&= (1-q) F_q(x,t) + (1-q) \sum_{i=1}^{x} \frac{i}{t+1-i} \cdot \frac{t!}{i!(t-i)!} q^i (1-q)^{t-i} \\
&= (1-q) F_q(x,t) + (1-q) \sum_{i=1}^{x} \frac{t!}{(i-1)!(t+1-i)!} q^i (1-q)^{t-i} \\
&= (1-q) F_q(x,t) + (1-q) \sum_{i=1}^{x} \binom{t}{i-1} q^i (1-q)^{t-i} \\
&= (1-q) F_q(x,t) + q \sum_{i=1}^{x} \binom{t}{i-1} q^{i-1} (1-q)^{t-i+1} \\
&= (1-q) F_q(x,t) + q \sum_{i=0}^{x-1} \binom{t}{i} q^i (1-q)^{t-i} \\
&= (1-q) F_q(x,t) + q \cdot F_q(x-1,t) \\
&= F_q(x,t) - q \cdot f_q(x,t)
\end{aligned}
$$

It remains to show that $qf_q(x,t) > \frac{x}{t} qf_q(x+1,t+1)$ which is done by noticing that since $x < t$ then:

$$
\begin{aligned}
\frac{x}{t} f_q(x+1,t+1) &= \frac{x}{t} \binom{t+1}{x+1} q^{x+1} (1-q)^{t-x} \\
&= \frac{x}{t} \cdot \frac{t+1}{x+1} \cdot q \cdot \binom{t}{x} q^x (1-q)^{t-x} \\
&= \frac{x}{t} \cdot \frac{t+1}{x+1} \cdot q \cdot f_q(x,t) \\
&< q \cdot f_q(x,t)
\end{aligned}
$$

□ □

We can now use the previous bound to compute the conditions under which the lower player prefers to stop competing.

Theorem 6.4.7 *In the game $G_k(t - p \cdot t, p \cdot t)$ for $p = r - \epsilon$, $\varepsilon > 0$ and $t = \max\{\frac{4\ln(1/12)}{\ln(1-r)}, \frac{3\ln(k)-\ln(q-p)}{(r-p)^2}\}$ player 2 does not compete at all.*

Proof: By Lemma 6.4.3 we have that $u_2(s_k(t-p\cdot t, p\cdot t)) \leq \max\left\{b_q(k,t,p\cdot t), kq\right\}$ for $t > \frac{4\ln(1/12)}{\ln(1-r)}$. Since we have that $u_2(s_k(t - p \cdot t, p \cdot t)) \geq kq$, if we show that $b_q(k,t,p\cdot t) \leq kq$, then we will have $u_2(s_k(t - p \cdot t, p \cdot t)) = kq$. It will then follow from Claim 6.2.4 that the lower player (player 2) does not compete at all. The theorem will thus follow if we show that for $t = \max\{\frac{4\ln(1/12)}{\ln(1-r)}, \frac{3\ln(k)-\ln(q-p)}{(r-p)^2}\}$, we have $b_q(k,t,p\cdot t) \leq kq$.

By Hoeffding's inequality with $\varepsilon = r - p$, we get that $F_r(p \cdot t, t) \leq e^{-2t(r-p)^2}$. Now, to compute the value of t for which $u_2(s_k(t-p\cdot t, p\cdot t)) = kq$, we simply find the value of t for which the following inequality holds:

$$p + (k-1)q + 3e^{-2t(r-p)^2}\big((k-1)(1-q)+1\big) \leq kq$$

After some rearranging we get that:

$$3e^{-2t(r-p)^2}\big((k-1)(1-q)+1\big) \leq q - p$$
$$3(k-1)(1-q)+1 \leq e^{2t(r-p)^2}(q-p)$$

Taking natural logarithms we get:

$$\ln(3(k-1)(1-q)+1) \leq 2t(r-p)^2 + \ln(q-p)$$
$$\frac{\ln(3(k-1)(1-q)+1) - \ln(q-p)}{2(r-p)^2} \leq t$$

In particular this implies that the claim holds for $t \geq \frac{3\ln(k)-\ln(q-p)}{(r-p)^2}$. $\square$

6.4.3 The Expected Social Welfare of a t-Binding Game

We show that for large enough k the social welfare of the t-binding game $G_k^t(1,1)$ is relatively high. This is done by showing that there exists some t, such that after competing for t rounds, with probability $2(r-\epsilon)-\frac{4}{t+1}$ the players reach a game in which the lower player (either player 1 or player 2) does not want to compete any more.

Lemma 6.4.8 *For every $\varepsilon > 0$ and $k \geq e^{\frac{8(r-\epsilon)}{\epsilon^3}} + e^{\frac{4\ln(1/12)}{\ln(1-r)}}$, there exists t such that the expected social welfare of the t-binding game $G_k^t(1,1)$ is at least $k \cdot (1+2q(r-3\epsilon-\epsilon^2))$.*

Proof: By the assumption that the game is t-binding we have that both players compete over the stronger candidate for the first t rounds. This implies that at the end of these t rounds with probability $\frac{1}{t+1}$ player 1 has a reputation of $1+t-i$ and player 2 has a reputation of $1+i$ for $0 \leq i \leq t$. Or, in other words, the relative reputation of player 2 is $\frac{1+i}{t+2}$ with probability $\frac{1}{t+1}$.

Notice that for any $0 \leq i \leq \lfloor (r-\epsilon)(t+2) \rfloor - 2$ it holds that $\frac{1+i}{t+2} < r-\epsilon$. Thus, the probability that the relative reputation of player 2 is smaller than $(r-\epsilon)$ is

$$\frac{1}{t+1} \cdot (\lfloor (r-\epsilon)(t+2) \rfloor - 2 + 1) \geq \frac{1}{t+1} \cdot ((r-\epsilon)(t+2) - 2) \geq (r-\epsilon) - \frac{2}{t+1}$$

This implies that with probability of at least $(r-\epsilon)-\frac{2}{t+1}$ after t rounds the current game is $G_{k-t}\left((t+2)(1-p), p\cdot(t+2)\right)$ for $p < r-\varepsilon$. Notice that by symmetry the same holds for player 1. By choosing t that obeys the requirements of Theorem 6.4.7 we get that the lower player in this game does not compete. Therefore, the probability that one of the players stops competing after t rounds is at least $2(r-\epsilon)-\frac{4}{t+1}$. To bound the expected social welfare we make the conservative assumption that with probability $1-(2(r-\epsilon)-\frac{4}{t+1})$ the players compete till the end of the game and get that:

$$u(s_k^t(1,1)) \geq k+2q\left((r-\epsilon)-\frac{2}{t+1}\right)(k-t) \geq k+2q(r-\epsilon)k-2q(r-\epsilon)t-\frac{4kq}{t}$$

Next, we show that for $k \geq e^{\frac{8(r-\epsilon)}{\epsilon^3}}+e^{\frac{4\ln(1/12)}{\ln(1-r)}}$ and $t=\frac{4\ln(k)-\ln(\varepsilon)}{\epsilon^2}$ the conditions for both Theorem 6.4.7 and this Lemma hold. Indeed, if Theorem 6.4.7 holds, we have that for every $0<p<r-\epsilon$ the players in the game $G_{k-t}\left((t+2)(1-p),p\cdot(t+2)\right)$ for $p<r-\varepsilon$ do not compete, as required. Recall that Theorem 6.4.7 requires $t+2$ to be at least $\max\{\frac{4\ln(1/12)}{\ln(1-r)},\frac{3\ln(k)-\ln(q-p)}{(r-p)^2}\}$. Observe that $\frac{4\ln(k)-\ln(\varepsilon)}{\epsilon^2} \geq \frac{3\ln(k)-\ln(q-p)}{(r-p)^2}$ as by definition $q-p \geq r-p > \varepsilon$; and that since $\ln(k) > \frac{4\ln(1/12)}{\ln(1-r)}$ we also have that $t \geq \frac{4\ln(1/12)}{\ln(1-r)}$.

Next, we show that $u(s_k^t(1,1)) \geq k\cdot(1+2q(r-3\epsilon-\epsilon^2))$. We begin by plugging in $t=\frac{4\ln(k)-\ln(\varepsilon)}{\epsilon^2}$:

$$\begin{aligned}
u(s_k^t(1,1)) &\geq k+2kq(r-\epsilon)-2q(r-\epsilon)\cdot\frac{4\ln(k)-\ln(\epsilon)}{\epsilon^2}-\frac{4kq}{\frac{4\ln(k)-\ln(\epsilon)}{\epsilon^2}} \\
&> k+2kq(r-\epsilon)-2q(r-\epsilon)\cdot\frac{4\ln(k)-\ln(\varepsilon)}{\epsilon^2}-\frac{kq\epsilon^2}{\ln(k)} \\
&\geq k+2kq(r-\epsilon-\epsilon^2)-2q(r-\epsilon)\cdot\frac{4\ln(k)}{\epsilon^2}+2q(r-\epsilon)\frac{\ln(\varepsilon)}{\varepsilon^2}
\end{aligned}$$

To prove the Lemma we show that for $k \geq e^{\frac{8(r-\epsilon)}{\epsilon^3}}+e^{\frac{4\ln(1/12)}{\ln(1-r)}}$ the following two inequalities hold:

1. $(r-\epsilon)\cdot\frac{8\ln(k)}{k\epsilon^2}<\epsilon$: For this we do a variable substitution and denote $\ln(k)=z$, so that $k=e^z$. Now we find z such that $4(r-\epsilon)z<\epsilon^3\cdot e^z$. By Taylor expansion we have that $e^z>\frac{z^2}{2}$. Thus, we can instead compute when $4(r-$

$\epsilon)z < \epsilon^3 \cdot \frac{z^2}{2}$ and get that the inequality holds for $z > \frac{8(r-\epsilon)}{\epsilon^3}$. This implies that the inequality holds for $k > e^{\frac{8(r-\epsilon)}{\epsilon^3}}$.

2. $|(r-\epsilon)\frac{\ln(\varepsilon)}{k\varepsilon^2}| < \varepsilon$: This condition also holds for $k > e^{\frac{8(r-\epsilon)}{\epsilon^3}}$ since if $k > e^{\frac{8(r-\epsilon)}{\epsilon^3}}$ by Taylor expansion we have that $k > ((\frac{8(r-\epsilon)}{\epsilon^3})^2)/2 = \frac{16(r-\varepsilon)^2}{\varepsilon^6} > \frac{r-\varepsilon}{\varepsilon^3} \cdot |\ln(\varepsilon)|$ and therefore $|(r-\epsilon)\frac{\ln(\varepsilon)}{k\varepsilon^2}| < \varepsilon$.

Thus, for $k \geq e^{\frac{8(r-\epsilon)}{\epsilon^3}} + e^{\frac{4\ln(1/12)}{\ln(1-r)}}$ and $t = \frac{4\ln(k)-\ln(\varepsilon)}{\epsilon^2}$ we have that $u(s_k^t(1,1)) \geq k \cdot (1 + 2q(r - 3\epsilon - \epsilon^2))$ as required. □

6.4.4 Wrapping up the Proof

Theorem 6.4.9 *For $\varepsilon > 0$ and $k \geq e^{\frac{8(r-\epsilon)}{\epsilon^3}} + e^{\frac{4\ln(1/12)}{\ln(1-r)}}$, the performance ratio of the game $G_k(1,1)$ is at least $\frac{1+2q(r-3\epsilon-\epsilon^2)}{1+q}$.*

Proof: By Lemma 6.4.1 we have that for any t, $u(s_k(1,1)) \geq u(s_k^t(1,1))$. By Lemma 6.4.8 we have that there exists a t such that $u(s_k^t(1,1)) \geq k\,(1 + 2q(r - 3\epsilon - \epsilon^2))$. By combining the two we get that $u(s_k(1,1)) \geq k\,(1 + 2q(r - 3\epsilon - \epsilon^2))$. This means that the performance ratio of the game $G_k(1,1)$ is at least $\frac{k\left(1+2q(r-3\epsilon-\epsilon^2)\right)}{k(1+q)} = \frac{1+2q(r-3\epsilon-\epsilon^2)}{1+q}$. □

Corollary 6.4.10 *As k goes to infinity, the performance ratio of the game $G_k(1,1,)$ goes to $\frac{1+2rq}{1+q}$.*

6.5 Other Competition Functions: Fixed Probability

One of the key components of our model is the underlying *competition function*: when players of reputation x_1 and x_2 respectively compete for the same candidate in a given round, the competition function specifies the probability that the candidate selects each player, in terms of x_1 and x_2. In this section we explore the effect that using other competition functions has on the performance ratio. An extreme example is when the higher player deterministically wins the competition (and if both players have the same reputation, then each wins with probability 1/2). Using this competition rule in the game $G_k(x, x)$ the players only compete for the first round to "discover" who is the higher player and then stop competing. Thus the performance ratio of this game is very close to 1. In this section we study a natural generalization of this function.

Consider a competition function specifying that the lower player wins with a fixed probability $p < 1/2$, and the higher player wins with probability $(1 - p)$. In case the two players have the same reputations, ties are broken in favor of player 1. Clearly if $p > q$, then the players compete forever, since the lower player gains more from competing than from going for the weaker candidate. Therefore, we assume from now on that $p < q$. We observe that this competition function belongs to the set of competition functions defined in the Appendix and hence we can make use of all claims specified there. For example, we have that the strategies $s_k(x_1, x_2)$ form a subgame perfect Nash equilibrium in this game (Proposition 6.7.6), the players' utilities are monotone (Claim 6.7.3) and that once a player decides to go after the lower candidate it will do so in all subsequent rounds (Claim 6.7.7).

We first show that once the absolute value of the difference between the players' reputations reaches $\theta(\log(k))$, the lower player stops competing (Claim 6.5.1

and Lemma 6.5.2). Then, in Lemma 6.5.3 we show that the expected number of rounds it takes the players to reach such a difference in reputations, starting from equal reputations, is also $\theta(\log(k))$. This implies that as k goes to infinity the performance ratio goes to 1, as we prove in Theorem 6.5.4.

Our first step is similar in spirit to the proof for the Tullock competition function; we show that the utility of the lower player is bounded by $\max\{b_q^p(k,d), kq\}$, where

$$\begin{aligned} b_q^p(k,d) &= p + (\frac{p}{1-p})^d k + (1 - (\frac{p}{1-p})^d)(k-1)q \\ &= p + (\frac{p}{1-p})^d \big((k-1)(1-q)+1\big) + (k-1)q \end{aligned}$$

This is obtained by induction over the difference in the reputations of the two players. The intuition for the upper bound function is also similar. In the good event, the lower player becomes the higher player and wins all subsequent rounds; hence its utility is k. In the bad event the lower player stays the lower player and loses the reward for competing this round; hence its utility is at most $(k-1)q$. To compute the probability of the good event, we can imagine that d (the difference between the players' reputations) is the initial location of a particle performing a biased random walk that goes left with probability p and right with probability $1-p$. Under this view, the probability that this particle ever reaches 0 — and hence that the difference d ever reaches 0 — is $\frac{p}{1-p}$ [40]. We formalize this intuition in the next claim. The claim is stated and proved for player 2 but a similar claim also holds for player 1.

Claim 6.5.1 *For any $d \geq 0$ and any k: $u_2(s_k(x, x-d)) \leq \max\{b_q^p(k,d), kq\}$.*

Proof: First observe that the claim clearly holds for any k and $d = 0$, since for this case $b_q^p(k,d) \geq k$ and by definition we have that $u_2(s_k(x,x)) \leq k$. We now prove by induction on the number of rounds k that the claim holds for $d \geq 1$. Note that the claim holds for the base case, $k=1$, since $u_2(s_1(x_1, x_2 - d)) \leq \max\{p, q\}$ for every $d \geq 1$. We assume the claim holds for any $0 < k' \leq k-1$ and prove it for k. If $u_2(s_k(x, x-d)) = kq$ we are done. Else,

$$u_2(s_k(x,x-d)) = p(1 + u_2(s_{k-1}(x, x-d+1))) + (1-p)(u_2(s_{k-1}(x+1, x-d)))$$

By the assumption that $u_2(s_k(x, x-d)) > kq$ and using Claim 6.7.15 we have that $u_2(s_{k-1}(x, x-d+1)) > (k-1)q$. Thus, by the induction hypothesis (or our observation for $d=0$ in case d was 1), we have that $u_2(s_{k-1}(x, x-d+1)) \leq b_q^p(k-1, d-1)$. By the induction hypothesis we also have that $u_2(s_{k-1}(x+1, x-d)) \leq \max\{b_q^p(k-1, d+1), (k-1)q\}$. We distinguish between two cases depending on the two possible upper bounds on $u_2(s_{k-1}(x+1, x-d))$:

If $u_2(s_{k-1}(x+1, x-d)) \leq b_q^p(k-1, d+1)$:

$$\begin{aligned}
u_2(s_k(x,x-d)) &\leq p\left(1 + b_q^p(k-1, d-1)\right) + (1-p) \cdot b_q^p(k-1, d+1) \\
&= 2p + (k-2)q + (\frac{p}{1-p})^{d-1} \cdot \left(p + \frac{p^2}{1-p}\right) \cdot \big((k-2)(1-q)+1\big) \\
&= 2p + (\frac{p}{1-p})^{d}\big((k-2)(1-q)+1\big) + (k-2)q \\
&< p + (\frac{p}{1-p})^{d}\big((k-1)(1-q)+1\big) + (k-1)q = b_q^p(k,d)
\end{aligned}$$

Else, $u_2(s_{k-1}(x+1, x-d)) = (k-1)q$:

$$\begin{aligned}
u_2(s_k(x,x-d)) &\leq p\left(1 + b_q^p(k-1, d-1)\right) + (1-p)(k-1)q \\
&< p + p^2 - pq + (k-1)q + p(\frac{p}{1-p})^{d-1}\big((k-1)(1-q)+1\big) \\
&< p + (\frac{p}{1-p})^{d}\big((k-1)(1-q)+1\big) + (k-1)q = b_q^p(k,d)
\end{aligned}$$

□

We can now use the previous claim to identify the magnitude of the difference between the players' reputations for which they stop competing. We state the claim for player 2 but a similar claim also holds for player 1.

Lemma 6.5.2 *If* $d > \dfrac{\log(k) - \log(p-q)}{\log(\frac{1-p}{p})} = d_q^p(k)$ *then the lower player in the games* $G_k(x, x-d)$, $G_k(x-d, x)$ *does not compete.*

Proof: Claim 6.5.1 reduces the problem of finding when does the lower player quits to solving for d such that $b_q^p(k,d) < kq$. This would be enough to conclude that player 2 stops competing as we know by Claim 6.7.14 that once $u_2(s_k(x_1, x_1)) = kq$ player 2 does not compete at all. After some rearranging of $b_q^p(k,d) < kq$ we have that:

$$(\frac{p}{1-p})^d\big((k-1)(1-q)+1\big) < q-p$$

We now take logarithms and get that:

$$d \cdot \log(\frac{p}{1-p}) + \log((k-1)(1-q)+1) < \log(q-p)$$

Therefore, $d > \dfrac{\log((k-1)(1-q)+1)) - \log(p-q)}{\log(\frac{1-p}{p})}$ and the claim follows. □

Next, we compute for how long the players are expected to compete until the absolute value of the difference between their reputations becomes greater than the previously computed bound. To do this, we study this difference as it performs a biased random with a reflecting barrier at 0:

Lemma 6.5.3 *In the game* $G_k(x_1, x_2)$*, the expected number of rounds the players compete until the absolute value of the difference between their reputations is at least* d *is at most* $\frac{d}{1-2p}$.

Proof: We consider a particle undergoing a biased random walk in which the probability of moving to the left is p and the probability of moving to the right is $1-p$, as before. Since this particle tracks the absolute value of the difference between the players' reputations, the walk we are studying has a reflecting barrier on 0. This implies that when the particle reaches 0, in the next step it always goes to 1.

Our analysis will thus be based on studying the expected time it takes for the particle to reach the value d, starting from a value below d. Clearly this expected time is maximized when the particle starts at 0, corresponding to an initial reputation difference of 0. Thus, we compute a bound on the expected number of rounds it takes players with identical reputations to reach a difference of d in their reputations.

The expected time it takes the particle to reach d starting at 0 when there is a reflective barrier is upper bounded by the expected time it takes it to reach d starting at 0 when there is no such barrier. To see why, we invoke a standard argument in which we imagine both walks being governed by the random flips of a coin with bias p, and we compare between the trajectory of the particle in these two walks for the same random sequence of coin-flip outcomes. We note that if the particle reaches d in the walk without the barrier it has to be the case that it also reached d using a prefix of the same sequence of coin-flip outcomes in the walk with the barrier.

Finally, we use the fact that the expected number of rounds required for a particle performing this walk to reach d starting from 0, without a barrier at 0, is $\dfrac{d}{1-2p}$ ([40]). To see why this is the case, let E_i be the expected time for the walk to reach i. If E_1 is well defined, then $E_i = i \cdot E_1$. Also, $E_1 = 1 + (1-p) \cdot 0 + p \cdot E_2 =$

$1 + 2p \cdot E_1$. Therefore, $E_1 = \dfrac{1}{1-2p}$ and $E_d = \dfrac{d}{1-2p}$. $\square$

We are now ready to prove the following theorem.

Theorem 6.5.4 *For fixed $p < q$ the performance ratio of the game $G_k(x_1, x_2)$ is at least $1 - \theta(\frac{\log(k)}{k})$.*

Proof: Let R be a random variable equals to the number of rounds for which the players compete in the game $G_k(x_1, x_2)$. We can use it to compute the social welfare as follows:

$$u(s_k(x_1, x_2)) = \sum_{r=1}^{k} Pr(R = r)(k + (k - r)q) = k(1 + q) - q \sum_{r} Pr(R = r) \cdot r$$

Now, to compute a lower bound on the social welfare, we should compute an upper bound on $\sum_{r=1}^{k} Pr(R = r) \cdot r$. We claim that $\sum_{r=1}^{k} Pr(R = r) \cdot r < \dfrac{d_q^p(k)}{1-2p}$. The reason is that either the lower player quits when the difference between the players' reputation is $d_q^p(k)$, as we proved in Claim 6.5.1, or the lower player might decide to quit earlier in the game. In any case, the expected number of rounds the players compete until the lower player drops is at most $\dfrac{d_q^p(k)}{1-2p}$, by Lemma 6.5.3. Therefore, the performance ratio is at least $\dfrac{k(1+q) - \frac{d_q^p(k)}{1-2p} q}{k(1+q)} = 1 - \theta(\frac{\log(k)}{k})$. $\square$

6.6 Conclusions

When firms compete for job applicants over many hiring cycles, there is a basic strategic tension inherent in the process: trying to recruit highly sought-after job candidates can build up a firm's reputation, but it comes with the risk that firm will fail to hire anyone at all. In this chapter, we have shown how this tension can arise in a simple dynamic model of job-market matching. Although our model is highly stylized, it contains a number of interesting effects that we analyze, including the way in which competition can lead to inefficiency through underemployment (quantified in our analysis of the performance ratio at equilibrium) and the possibility of different modes of behavior, in which a weaker firm may end up competing forever, or it may give up at some point and accept its second-place status.

The model and analysis also suggest a number of directions for further investigation. One direction is to vary the *competition function* that determines the outcome of a competition between the two firms when they make offers to the same candidate. As noted above, this can be viewed as varying the way in which candidates make decisions between firms based on their reputations. In Section 6.5 of the appendix, we explore this issue by considering an alternate rule for competition in which the lower-reputation player wins with a fixed probability $p < \frac{1}{2}$ (independent of the difference in reputation) and the higher-reputation player wins with probability $1 - p$.

This fixed-probability competition function is simpler in structure than the Tullock function, and it is illuminating in that it cleanly separates two different aspects of the strategic decision being made about future rounds. With the Tullock function, when the lower player competes, it has the potential for a short-term gain in its success probability even in the next round (since the ratio of reputations will change), and it also has the potential for a long-term gain by becoming the higher player. With the fixed-probability competition function, the short-term aspect is effectively eliminated, since as long as a player remains the lower party, it has the same probability of success; we are thus able to study strategic behavior about competing when the only upside is the long-range prospect of becoming the higher player. We show that the performance is generally much better with this fixed-probability rule than with the Tullock function, providing us with further insight into the specific way in which competition leads to inefficiency through a reduced performance ratio.

Other directions that lead quickly to interesting questions are to consider the case of more than two firms, and to consider models in which the candidates have different characteristics in different time periods. For both of these general directions, our initial investigations suggest that the techniques developed here will be useful for shedding light on the properties of more complex models that take these issues into account.

6.7 Appendix: The Canonical Equilibrium and its Properties

Our main goal in this section is to prove that the strategies $s_k(x_1, x_2)$ form a subgame perfect equilibrium in the game $G_k(x_1, x_2)$, and to present the proofs of some useful properties of this equilibrium. The arguments can be carried out in a setting more general than that of the Tullock competition function, and we present them for a broader class of competition functions, specifying the probability that a candidate chooses each firm in the event of competition between them. We work at this greater level of generality for two reasons. First, it makes clear what properties of the competition function are necessary for the equilibrium results. Second, and more concretely, we study a variant of the model in Section 6.5 that involves a different competition function, in which a candidate picks the lower-reputation firm with a fixed probability $p < \frac{1}{2}$ regardless of the actual numerical values of the reputations. This fixed-probability competition function satisfies our more general assumptions, and thus we can apply all the results of this section to it.

For ease of exposition, the results in Section 6.2 of the main text are presented specifically for the Tullock competition function; as a result, to complete the link back to this section, we state which claims here generalize each claim from Section 6.2.

Our results hold for the following general definition of a competition function $c : \mathbb{R} \times \mathbb{R} \to (0, 1)$, capturing the intuitive notion that $c(x_1, x_2)$ should represent the probability that player 1 wins a competition when the two players' strengths are x_1 and x_2 respectively.

Definition 6.7.1 *A function $c : \mathbb{R} \times \mathbb{R} \to (0, 1)$ is a competition function if:*

- $c(x_1, x_2 + \varepsilon) \leq c(x_1, x_2) \leq c(x_1 + \varepsilon, x_2)$.
- *For every $x_1 \neq x_2$: $c(x_1, x_2) = (1 - c(x_2, x_1))$.*
- $c(x, x) \geq (1 - c(x, x))$.

With this notation in mind the utility of player 2 for competing is now:

$$(1 - c(x_1, x_2)) \cdot (1 + u_2(s_{k-1}(x_1, x_2 + 1))) + c(x_1, x_2) \cdot u_2(s_{k-1}(x_1 + 1, x_2)).$$

Observe that the following two properties hold for any competition function. These will be useful for later proofs:

- Let $\varepsilon > 0$. $(1 - c(x_1, x_2 + \varepsilon)) \geq c(x_2, x_1)$.
- Let $\varepsilon > 0$. $(1 - c(x_2, x_1)) \geq c(x_1, x_2 + \varepsilon)$.

To see why the first statement holds, observe that if $x_2 \neq x_1 + \varepsilon$ we have that:

$$(1 - c(x_1, x_2 + \varepsilon)) = c(x_2 + \varepsilon, x_1) \geq c(x_2, x_1).$$

Else, we have that $x_2 \neq x_1$ and in this case we have that:

$$(1 - c(x_1, x_2 + \varepsilon)) \geq (1 - c(x_1, x_2)) = c(x_2, x_1).$$

The second statement also holds for similar reasons.

We begin by showing that since the lower player in $s_k(x_1, x_2)$ chooses the strategy maximizing its utility it can always guarantee itself a utility of at least kq by always going for the weaker candidate.

Claim 6.7.2 *Let player i be the lower player in the game $G_k(x_1, x_2)$. Then $u_i(s_k(x_1, x_2)) \geq kq$:*

Proof: We prove the claim for the case that $x_2 \leq x_1$ but a similar proof can be easily devised for the case that $x_1 < x_2$. We prove the claim by induction on the number of rounds k. For the base case $k = 1$, it is easy to see that the utility of player 2 is $\max\{(1 - c(x_1, x_2)), q\}$, and therefore the claim holds. We assume correctness for $(k-1)$-round games and prove for k-round games. Observe that: $u_2(s_k(x_1, x_2)) = \max\{u_2(s_k^{\langle +,+ \rangle}(x_1, x_2)), q + u_2(s_{k-1}(x_1 + 1, x_2))\}$. Since player 2 is also the lower player in the game $G_{k-1}(x_1 + 1, x_2)$ we can use the induction hypothesis and get that $u_2(s_{k-1}(x_1 + 1, x_2)) \geq (k-1)q$ which completes the proof. $\square$

Next we show that the utilities $u_i(s_k(x_1, x_2))$ are monotone increasing in player i's reputation and monotone decreasing in its opponent's reputation.

Claim 6.7.3 (generalizes Claim 6.2.2) *For any x_1, x_2, and $\varepsilon > 0$:*

1. $u_1(s_k(x_1 + \varepsilon, x_2)) \geq u_1(s_k(x_1, x_2))$ *and* $u_2(s_k(x_1 + \varepsilon, x_2)) \leq u_2(s_k(x_1, x_2))$
2. $u_1(s_k(x_1, x_2 - \varepsilon)) \geq u_1(s_k(x_1, x_2))$ *and* $u_2(s_k(x_1, x_2 - \varepsilon)) \leq u_2(s_k(x_1, x_2))$.

Proof: We prove both properties concurrently by induction over k, the number of rounds in the game. Since the proofs for both properties are very similar, we only present here the proof for the first property. For the base case, $k = 1$, we distinguish between the following cases cases:

1. $s_1(x_1, x_2) = s_1(x_1 + \epsilon, x_2)$: if they compete in in both $s_1(x_1, x_2)$ and $s_1(x_1 + \epsilon, x_2)$, then the claim holds simply because $c(x_1 + \epsilon, x_2) \geq c(x_1, x_2)$ and $(1 - c(x_1 + \epsilon, x_2)) \leq (1 - c(x_1, x_2))$. Else, in both games the players have the exact same utility (either 1 or q).

2. $s_1(x_1, x_2) \neq s_1(x_1 + \epsilon, x_2)$, $s_1(x_1, x_2) \neq \langle +, + \rangle$ and $s_1(x_1 + \epsilon, x_2) \neq \langle +, + \rangle$: observe that this is only possible if $x_1 < x_2 \leq x_1 + \epsilon$ and in this case we have that $s_1(x_1, x_2) = \langle -, + \rangle$ and $s_1(x_1 + \epsilon, x_2) = \langle +, - \rangle$ so it is not hard to see that the claim holds.

3. $s_1(x_1, x_2) = \langle +, + \rangle$ and $s_1(x_1 + \epsilon, x_2) \neq \langle +, + \rangle$: this implies that $x_1 \geq x_2$ since the utility player 1 can get for competing in $G_1(x_1 + \epsilon, x_2)$ is greater than the utility it can get for competing in $G_1(x_1, x_2)$: $c(x_1 + \epsilon, x_2) \geq c(x_1, x_2) > q$. Therefore, $u_1(s_1(x_1 + \varepsilon, x_2)) \geq u_1(s_1(x_1, x_2))$. For player 2, $u_2(s_k(x_1 + \varepsilon, x_2)) \leq u_2(s_k(x_1, x_2))$, since we have that $q < (1 - c(x_1, x_2))$.

4. $s_1(x_1, x_2) \neq \langle +, + \rangle$ and $s_1(x_1+\epsilon, x_2) = \langle +, + \rangle$: this implies that $x_2 > x_1$. As for player 2 its not hard to see that: $u_2(s_1^{\langle +,+\rangle}(x_1, x_2)) \geq u_2(s_1^{\langle +,+\rangle}(x_1+\epsilon, x_2))$. Thus the only reason that the players do not compete in $s_1(x_1, x_2)$ is that player 1 prefers to go for the weaker candidate and it is entitled to make this choice in $s_1(x_1, x_2)$ only is $x_1 < x_2$. It is not hard to see that the claim holds for this case as well.

We now assume that both statements 1 and 2 hold for $(k-1)$-round games and prove they also hold for k-round games. The proof takes a very similar structure to the proof for the base case, except now we shall use the induction hypothesis instead of first principles. The following observation will be useful for the proof:

Observation 6.7.4 *By applying the induction hypothesis we get that the following two statements hold for any $\delta \in \{\langle +, + \rangle, \langle +, - \rangle, \langle -, + \rangle\}$:*

1. $u_1(s_k^\delta(x_1 + \epsilon, x_2)) \geq u_1(s_k^\delta(x_1, x_2))$
2. $u_2(s_k^\delta(x_1, x_2)) \geq u_2(s_k^\delta(x_1 + \epsilon, x_2))$

Take for example the first statement and consider $\delta = \langle +, + \rangle$. To see why it is indeed the case that $u_1(s_k^{\langle +,+\rangle}(x_1 + \epsilon, x_2)) \geq u_1(s_k^{\langle +,+\rangle}(x_1, x_2))$, observe that by the induction hypothesis we have that: $u_1(s_{k-1}(x_1 + 1 + \epsilon, x_2)) \geq u_1(s_{k-1}(x_1 + 1, x_2))$, $u_1(s_{k-1}(x_1 + \epsilon, x_2 + 1)) \geq u_1(s_{k-1}(x_1, x_2 + 1))$, $u_1(s_{k-1}(x_1 + \varepsilon + 1, x_2)) \geq u_1(s_{k-1}(x_1 + \varepsilon, x_2 + 1))$ and that $c(x_1 + \epsilon, x_2) \geq c(x_1, x_2)$. Similarly, we can use the induction hypothesis to prove that the two statements are correct for any $\delta \in \{\langle +, + \rangle, \langle +, - \rangle, \langle -, + \rangle\}$.

Just as in the base case, we now distinguish between the following cases:

1. $f(s_k(x_1, x_2))) = f(s_k(x_1 + \epsilon, x_2))$: the claim holds by Observation 6.7.4.

2. $f(s_k(x_1, x_2)) \neq f(s_k(x_1 + \epsilon, x_2))$, $f(s_k(x_1, x_2)) \neq \langle +, + \rangle$ and $f(s_k(x_1 + \epsilon, x_2)) \neq \langle +, + \rangle$: observe that this is only possible if $x_1 < x_2 \leq x_1 + \epsilon$ and in this case we have that $f(s_k(x_1, x_2)) = \langle -, + \rangle$ and $f(s_k(x_1 + \epsilon, x_2)) = \langle +, - \rangle$ so it is not hard to see that the claim holds.

3. $f(s_k(x_1, x_2)) = \langle +, + \rangle$ and $f(s_k(x_1 + \epsilon, x_2)) \neq \langle +, + \rangle$: Similar to the corresponding case for $k = 1$, observe that this implies that $x_1 \geq x_2$. As by Observation 6.7.4 we have that $u_1(s_k^{\langle +,+ \rangle}(x_1 + \epsilon, x_2)) \geq u_1(s_k^{\langle +,+ \rangle}(x_1, x_2))$. Thus, if the players do not compete in $s_k(x_1 + \epsilon, x_2)$ it can only be because the lower player prefers not to compete and this lower player has to be player 2. Now, by applying the induction hypothesis for player 1 we get that $u_1(s_{k-1}(x_1 + \epsilon + 1, x_2)) \geq u_1(s_{k-1}(x_1 + 1, x_2)) \geq u_1(s_{k-1}(x_1, x_2 + 1))$. Thus, it is not hard to see that $u_1(s_k(x_1 + \varepsilon, x_2)) \geq u_1(s_k(x_1, x_2))$. For player 2, since as the lower player it chooses to compete in $f(s_k(x_1, x_2))$ but not in $f(s_k(x_1 + \epsilon, x_2))$ we have that:

$$\begin{aligned} u_2(s_k(x_1, x_2) = u_2(s_k^{\langle +,+ \rangle}(x_1, x_2)) > u_2(s_k^{\langle +,- \rangle}(x_1, x_2)) \geq u_2(s_k^{\langle +,- \rangle}(x_1 + \epsilon, x_2)) \\ = u_2(s_k(x_1 + \epsilon, x_2) \end{aligned}$$

where the last transition is by Observation 6.7.4.

4. $f(s_k(x_1, x_2)) \neq \langle +, + \rangle$ and $f(s_k(x_1 + \epsilon, x_2)) = \langle +, + \rangle$: this is similar to the previous case only now we have that $x_1 < x_2$. The reason is that by Observation 6.7.4 we have that $u_2(s_k^{\langle +,+ \rangle}(x_1, x_2)) \geq u_2(s_k^{\langle +,+ \rangle}(x_1 + \epsilon, x_2))$. Therefore the lower player in the game $G_k(x_1, x_2)$ is player 1. After establishing this, it is easy to verify that the claim holds by applying the induction hypothesis in a very similar manner to the previous case.

$\square$

6.7.1 Subgame Perfection of the Strategies $s_k(x_1, x_2)$

We prove the following three statements simultaneously by induction on the number of rounds in the game:

Proposition 6.7.5 (generalizes Claim 6.2.3) *For any integers x_1, x_2 and k the following holds for the strategies $s_k(x_1, x_2)$.*

1. *$s_k(x_1, x_2)$ is a sub-game perfect equilibrium in the game $G_k(x_1, x_2)$.*
2. *If a player does not compete in the first round of the game $G_k(x_1, x_2)$, then it does not compete in all subsequent rounds.*
3. *The utility of the higher player in the game $G_k(x_1, x_2)$ is at least as large as the utility of the lower player.*

We separate the simultaneous induction into three numbered statements above. Proving these three statements by simultaneous induction on k means that in studying properties of $s_k(x_1, x_2)$ and $G_k(x_1, x_2)$, we can assume that all three parts hold for $s_{k'}(x_1, x_2)$ and $G_{k'}(x_1, x_2)$ for every $0 < k' \leq k - 1$.

We begin by presenting the proof for part 1 of the proposition showing that $s_k(x_1, x_2)$ is a sub-game perfect equilibrium. As part of the induction the proof relies on the correctness of parts 2 and 3 for games of less than k rounds. Next, we prove parts 2 and 3 in Claim 6.7.7 and Proposition 6.7.9 respectively. Both proofs assume that $s_{k'}(x_1, x_2)$ is a subgame perfect equilibrium for every $0 < k' \leq k - 1$.

Proposition 6.7.6 *The strategies described by $s_k(x_1, x_2)$ are a subgame perfect equilibrium in the game $G_k(x_1, x_2)$ for every two integers x_1, x_2.*

Proof: We prove the claim by induction on k the number of rounds. We only present the proof for $x_1 \geq x_2$ as the proof for the case that $x_1 < x_2$ is very similar. For the base case $k = 1$, if $c(x_1, x_2) > q$, then it is clearly the case that player 1 competes. Else, since $x_1 \geq x_2$ we have that $(1 - c(x_1, x_2)) \leq c(x_1, x_2) \leq q$, therefore player 2 does not want to compete as well. As clearly player 1 cannot benefit from competing over the weaker candidate, this implies that there exists an equilibrium in which player 1 goes for the stronger candidate.

Next, we assume correctness for k'-round games for any $0 < k' \leq k - 1$ and prove for k-round games. This means we can apply Corollary 6.7.8 and get that once the lower player prefers to go for the weaker candidate it completely stops competing and Proposition 6.7.9 to get that the higher player always has greater utility than the lower player.

Assume towards contradiction that in the *unique* equilibrium for the first round of the game $G_k(x_1, x_2)$ player 1 goes after the weaker candidate and player 2 goes after the stronger candidate. Thus, player 1's utility is $u_1(s_k^{\langle -,+\rangle}(x_1, x_2)) = q + u_1(s_{k-1}(x_1, x_2 + 1))$.

We first observe that by monotonicity if player 1 prefers to go for the weaker candidate over competing, it has to be the case that $c(x_1, x_2) < q$. We also observe that it has to be the case that $f(s_k(x_1, x_2)) = \langle +, + \rangle$. As it is easy to see that in the case of $f(s_k(x_1, x_2)) = \langle +, - \rangle$ player 1 prefers to go for the stronger candidate over competing for the *weaker* candidate.[2]

We now distinguish between 3 possible scenarios in $s_{k-1}(x_1, x_2 + 1)$ which is by the induction hypothesis a subgame perfect equilibrium in $G_{k-1}(x_1, x_2 + 1)$:

[2]Observe that this holds since $q + u_1(s_{k-1}(x_1 + q, x_2)) < 1 + u_1(s_{k-1}(x_1 + 1, x_2))$ by monotonicity. This is a well defined use of Claim 6.7.3 since it does not assume the players' reputations to be integers.

1. $f(s_{k-1}(x_1, x_2+1)) = \langle +, + \rangle$: observe that in this case player 1 prefers to compete in the first round of the game $G_k(x_1, x_2)$ since by monotonicity $u_1(s_{k-2}(x_1+1, x_2+1)) \geq u_1(s_{k-2}(x_1, x_2+2))$[3]; by the induction hypothesis we have that $s_{k-1}(x_1, x_2+1)$ and $s_{k-1}(x_1+1, x_2)$ are subgame perfect equilibria we have that $u_1(s_{k-1}(x_1+1, x_2)) \geq q + u_1(s_{k-2}(x_1+1, x_2+1))$ and $u_1(s_{k-1}(x_1, x_2+1)) \geq q + u_1(s_{k-2}(x_1, x_2+2))$; and $c(x_1, x_2) \geq c(x_1, x_2+1)$. Thus, we have that:

$$\begin{aligned} u_1(s_k^{\langle +,+ \rangle}(x_1, x_2)) &= c(x_1, x_2) \cdot (1 + u_1(s_{k-1}(x_1+1, x_2))) \\ &\quad + (1 - c(x_1, x_2)) \cdot u_1(s_{k-1}(x_1, x_2+1)) \\ &\geq c(x_1, x_2) \cdot (1 + (q + u_1(s_{k-2}(x_1+1, x_2+1)))) \\ &\quad + (1 - c(x_1, x_2)) \cdot (q + u_1(s_{k-2}(x_1, x_2+2))) \\ &\geq q + c(x_1, x_2+1) \cdot (1 + u_1(s_{k-2}(x_1+1, x_2+1))) \\ &\quad + (1 - c(x_1, x_2+1)) \cdot u_1(s_{k-2}(x_1, x_2+2)) \\ &= q + u_1(s_{k-1}^{\langle +,+ \rangle}(x_1, x_2+1)) \\ &= u_1(s_k^{\langle -,+ \rangle}(x_1, x_2)). \end{aligned}$$

Therefore $\langle -, + \rangle$ is not an equilibrium for the first round of the game $G_k(x_1, x_2)$.

2. $f(s_{k-1}(x_1, x_2+1)) = \langle +, - \rangle$: this implies that $x_1 \geq x_2 + 1$. We can apply Corollary 6.7.8 and get that for player 2, $u_2(s_{k-1}(x_1, x_2+1)) = (k-1)q$. Since by monotonicity we have that $(k-1)q = u_2(s_{k-1}(x_1, x_2+1)) \geq u_2(s_{k-1}(x_1 + 1, x_2))$ we get that:

$$u_2(s_k^{\langle +,+ \rangle}(x_1, x_2)) \leq (1 - c(x_1, x_2)) + (k-1)q < kq$$

Where the last transition is due to the fact that $(1 - c(x_1, x_2)) \leq c(x_1, x_2) < q$. Hence, by Claim 6.7.2 the lower player (player 2) does not want to compete

[3]To handle the case of $k = 2$ we define $u_i(s_0(x_1, x_2)) = 0$.

in the game $G_k(x_1, x_2)$. Therefore $\langle -, + \rangle$ is not the unique equilibrium for the first round of the game $G_k(x_1, x_2)$.

3. $f(s_{k-1}(x_1, x_2 + 1)) = \langle -, + \rangle$: This implies that $x_1 = x_2$ therefore we can apply Corollary 6.7.8 for player 1 in the game $G_{k-1}(x_1, x_2 + 1)$ and get that $u_1(s_k^{\langle -,+ \rangle}(x_1, x_2)) = q + u_1(s_{k-1}(x_1, x_2 + 1)) = kq$. We now have the following chain of inequalities, by applying Proposition 6.7.9:

$$u_1(s_k^{\langle -,+ \rangle}(x_1, x_2)) > u_1(s_k^{\langle +,+ \rangle}(x_1, x_2)) = u_1(s_k(x_1, x_2)) \geq u_2(s_k(x_1, x_2)) \geq kq$$

The last transition is due to the fact that player 2 is the lower player in the game $G_k(x_1, x_2)$ and thus we can use Claim 6.7.2. This is of course in contradiction to the fact that $u_1(s_k^{\langle -,+ \rangle}(x_1, x_2)) = kq$.

□

Proof of Part (2) of Proposition 6.7.5

We show that if a player prefers not compete in the first round of the game $G_k(x_1, x_2)$, then it does not compete in all subsequent rounds. This is done by showing that if a player does not compete in the first round of a game, then it does not compete in the second round. The proof assumes that $s_{k-1}(x_1, x_2)$ is a subgame perfect equilibrium as it used as part of the induction.

Formally we show:

Claim 6.7.7 *If $s_{k-1}(x_1, x_2)$ is a subgame perfect equilibrium for every x_1 and x_2, then*

- $f(s_k(x_1, x_2)) = \langle +, - \rangle \implies f(s_{k-1}(x_1 + 1, x_2)) = \langle +, - \rangle$.

- $f(s_k(x_1, x_2)) = \langle -, + \rangle \implies f(s_{k-1}(x_1, x_2 + 1)) = \langle -, + \rangle$.

Proof: We prove the first statement of the claim as the proof of the second statement is very similar. Assume towards a contradiction that $f(s_k(x_1, x_2)) = \langle +, - \rangle$ but $f(s_{k-1}(x_1 + 1, x_2)) \neq \langle +, - \rangle$. The fact that $f(s_k(x_1, x_2)) = \langle +, - \rangle$ implies that $x_2 \leq x_1$; thus if $f(s_{k-1}(x_1 + 1, x_2)) \neq \langle +, - \rangle$ it has to be the case that $f(s_{k-1}(x_1 + 1, x_2)) = \langle +, + \rangle$. Therefore, player 2's utility is:

$$\begin{aligned} u_2(s_k(x_1, x_2)) = q + (1 - c(x_1 + 1, x_2)) \cdot (1 + u_2(s_{k-2}(x_1 + 1, x_2 + 1))) \\ + c(x_1 + 1, x_2) \cdot u_2(s_{k-2}(x_1 + 2, x_2)). \end{aligned}$$

Observe that the following holds:

1. $f(s_{k-1}(x_1 + 1, x_2)) = \langle +, + \rangle \implies u_2(s_{k-1}(x_1 + 1, x_2)) > q + u_2(s_{k-2}(x_1 + 2, x_2))$.
2. $u_2(s_{k-1}(x_1, x_2 + 1)) \geq q + u_2(s_{k-2}(x_1 + 1, x_2 + 1))$.

The first of these statements holds since player 2 is the lower player in the game $G_{k-1}(x_1 + 1, x_2)$, and the second statement holds since by assumption $s_{k-1}(x_1, x_2 + 1)$ is a subgame perfect equilibrium. Thus, we have that:

$$\begin{aligned} u_2(s_k^{\langle +,+ \rangle}(x_1, x_2)) > q + (1 - c(x_1, x_2)) \cdot (1 + u_2(s_{k-2}(x_1 + 1, x_2 + 1))) \\ + c(x_1, x_2) \cdot u_2(s_{k-2}(x_1 + 2, x_2)). \end{aligned}$$

This implies that $u_2(s_k^{\langle +,+ \rangle}(x_1, x_2)) > u_2(s_k(x_1, x_2))$ in contradiction to the assumption that player 2 maximizes its utility by first going for the weaker candidate ($f(s_k(x_1, x_2)) = \langle +, - \rangle$). $\square$

Corollary 6.7.8 *For $x_2 \leq x_1$, if $s_{k'}(x_1, x_2)$ is a subgame perfect equilibrium for every $0 < k' \leq k-1$, x_1 and x_2, then, $f(s_k(x_1, x_2)) = \langle +, - \rangle \implies u_2(s_k(x_1, x_2)) = kq$.*

Proof of Part (3) of Proposition 6.7.5

We show that the utility of the higher player in the game $G_k(x_1, x_2)$ is at least as large as the utility of the lower player. This is based on Claim 6.7.10 and Claim 6.7.11 below.

Proposition 6.7.9 *If $s_{k-1}(x_1, x_2)$ is a subgame perfect equilibrium for every two integers x_1 and x_2 then:*

- *For $x_1 \geq x_2$: $u_1(s_k(x_1, x_2)) \geq u_2(s_k(x_1, x_2))$.*
- *For $x_2 > x_1$: $u_2(s_k(x_1, x_2)) \geq u_1(s_k(x_1, x_2))$.*

Proof:

- For $x_1 \geq x_2$. By Claim 6.7.10 we have that $u_1(s_k(x_1, x_2)) \geq u_2(s_k(x_2, x_1))$. Now, by monotonicity we have that $u_2(s_k(x_2, x_1)) \geq u_2(s_k(x_2, x_2)) \geq u_2(s_k(x_1, x_2))$.
- For $x_2 > x_1$. By Claim 6.7.11 we have that $u_2(s_k(x_1, x_2)) \geq u_1(x_2 - 1, x_1)$. Now, by monotonicity we have that $u_1(x_2 - 1, x_1) \geq u_1(x_1, x_1)$ since because x_1 and x_2 are integers we have that $x_2 - 1 \geq x_1$. Then we have that $u_1(x_1, x_1) \geq u_1(x_1, x_2)$ which completes the proof.

□

The two following claims allow us to show that the utility of the higher player is always greater by relating between the utilities of player 1 and player 2.

Claim 6.7.10 *If for every integers y', z' and $0 < k' \leq k-1$ $s_{k'}(y', z')$ is a subgame perfect equilibrium then: $u_1(s_k(y, z)) \geq u_2(s_k(z, y))$ for every y and z.*

Proof: We prove the claim by induction on k the number of rounds. For the base case $k = 1$, observe that if $y = z$ then clearly $u_1(s_1(y, z)) \geq u_2(s_1(z, y))$; either because the players compete and $c(y, y) \geq (1 - c(y, y))$ or because the players do not compete and $u_1(s_1(y, z)) = 1 > q = u_2(s_1(z, y))$. Else, $y \neq z$, now if $\min\{c(y, z), (1-c(y, z))\} > q$, then in both $s_1(y, z)$ and $s_1(z, y)$ the players compete and since $y \neq z$ we have that $c(y, z) = (1-c(y, z))$, thus $u_1(s_1(y, z)) = u_2(s_1(z, y))$. Else, each of the players goes after a different candidate. If for example $y > z$, then in both games the player with reputation y goes after the stronger candidate and the player with reputation z goes after the weaker candidate. Thus, $u_1(s_k(y, z)) = u_2(s_k(z, y))$.

Next, we assume the correctness for $(k-1)$-round games and prove for k-round games. We distinguish between the following cases:

1. $f(s_k(y, z)) = f(s_k(z, y)) = \langle +, + \rangle$: by using the induction hypothesis we get that that $u_1(s_{k-1}(y + 1, z)) \geq u_2(s_{k-1}(z, y + 1))$ and $u_1(s_{k-1}(y, z + 1)) \geq u_2(s_{k-1}(z+1, y))$. Since $c(y, z) \geq (1-c(y, z))$ this is sufficient for showing that $u_1(s_k(y, z)) \geq u_2(s_k(z, y))$. More generally this shows that $u_1(s_k^{\langle +,+ \rangle}(y, z)) \geq u_2(s_k^{\langle +,+ \rangle}(z, y))$.
2. $f(s_k(y, z)) \neq \langle +, + \rangle$ and $f(s_k(z, y)) \neq \langle +, + \rangle$: by the definition of s_k we have three possible subcases:

- $y < z$: $f(s_k(y,z)) = \langle -,+\rangle$ and $f(s_k(z,y)) = \langle +,-\rangle$.
- $y = z$: $f(s_k(y,y)) = \langle +,-\rangle$.
- $y > z$: $f(s_k(y,z)) = \langle +,-\rangle$ and $f(s_k(z,y)) = \langle -,+\rangle$.

It is not hard to see that for each one of these subcases we can use the induction hypothesis to show that the claim holds.

3. $f(s_k(y,z)) = \langle +,+\rangle$ and $f(s_k(z,y)) \neq \langle +,+\rangle$: if $z > y$, then:

$$u_1(s_k(z,y)) = 1 + u_1(s_{k-1}(z+1,y))$$
$$u_2(s_k(y,z)) = (1 - c(y,z)) \cdot (1 + u_2(s_{k-1}(y,z+1))) + c(y,z) \cdot u_2(s_{k-1}(y+1,z))$$

By using monotonicity and applying the induction hypothesis we get that:

$$u_2(s_{k-1}(y+1,z)) \leq u_2(s_{k-1}(y,z+1)) \leq u_1(s_{k-1}(z+1,y))$$

Thus, the claim holds.

Else, we have that $y > z$. We show that this case is not possible by a *locking argument*. First we observe that this implies that player 1 is the lower player in the game $G_k(z,y)$ and in the first round of the game it prefers to go for the weaker candidate. Since we assume that for any y',z' and $0 < k' \leq k-1$, $s_{k'}(y',z')$ is a subgame perfect equilibrium the requirements of Corollary 6.7.8 hold and thus we have that: $u_1(s_{k-1}(z,y+1)) = (k-1)q$. By applying the induction hypothesis we get that $u_2(s_{k-1}(y+1,z)) \leq u_1(s_{k-1}(z,y+1)) = (k-1)q$. Now, since player 2 is the lower player in the game $G_k(y+1,z)$ we can apply Claim 6.7.2 and conclude that $u_2(s_{k-1}(y+1,z)) = (k-1)q = u_1(s_{k-1}(z,y+1))$. Now, the following chain of inequalities provides a contradiction for the assumption that in the game $G_k(z,y)$ the players do not compete as it shows that the lower player (player 1) actually prefers competing over going for the weaker candidate:

$$u_1(s_k^{\langle +,+\rangle}(z,y)) \geq u_2(s_k^{\langle +,+\rangle}(y,z)) > q + u_2(s_k(y+1,z)) = q + u_1(s_{k-1}(z,y+1))$$
$$= u_1(s_k^{\langle -,+\rangle}(z,y)).$$

The claim is completed since we already treated the case in which $y = z$ as part of the first two cases.

□

We now prove that a claim similar in spirit to the previous one also holds for player 2:

Claim 6.7.11 *If for every integers y', z' and $0 < k' \leq k-1$ $s_{k'}(y', z')$ is a subgame perfect equilibrium, then, $u_2(s_k(y, z+1)) \geq u_1(s_k(z,y))$ and $u_2(s_k(z,y)) \geq u_1(s_k(y,z+1))$ for every two integers y and z.*

Proof: We prove the two inequalities by induction on k simultaneously. We begin with the base case $k = 1$ and distinguish between the following cases:

1. $s_1(y, z+1) = s_1(z, y) = \langle +,+\rangle$: in this case $u_2(s_1(y,z+1)) = (1-c(y,z+1))$, $u_1(s_1(z,y)) = c(z,y)$, $u_2(s_1(z,y)) = (1-c(z,y))$ and $u_1(s_1(y,z+1)) = c(y,z+1)$, thus the claim holds.
2. $s_1(y, z+1) \neq \langle +,+\rangle$ and $s_1(z,y) \neq \langle +,+\rangle$: if $y \leq z$ then $u_2(s_1(y,z+1)) = u_1(s_1(z,y)) = 1$ and $u_1(s_1(y,z+1)) = u_2(s_1(z,y)) = q$, thus the claim holds. Else, $y \geq z+1$, then $u_2(s_1(y,z+1)) = u_1(s_1(z,y)) = q$ and $u_1(s_1(y,z+1)) = u_2(s_1(z,y)) = 1$.
3. The players compete in one of $s_1(y,z+1), s_1(z,y)$ and do not compete in the other: Observe that the following hold:

- If $y \leq z$ then $s_1(y, z+1) = \langle +,+\rangle \implies s_1(z,y) = \langle +,+\rangle$. Observe that $(1 - c(z,y)) \geq c(y, z+1)$. This implies that $u_2(s_1^{\langle +,+\rangle}(z,y)) \geq u_1(s_1^{\langle +,+\rangle}(y, z+1)) > q$. Now since in both cases the lower player is the player with reputation y the claim follows.

- If $y \geq z+1$ then $s_1(z,y) = \langle +,+\rangle \implies s_1(y, z+1) = \langle +,+\rangle$. Observe that $(1 - c(y, z+1)) \geq c(z,y)$. This implies that $u_2(s_1^{\langle +,+\rangle}(y, z+1)) \geq u_1(s_1^{\langle +,+\rangle}(z,y)) > q$. Now since in both cases the lower player is the player with reputation z or $z+1$ the claim follows.

Thus, we are left with the following two sub-cases:

(a) $s_1(y, z+1) = \langle +,+\rangle$ and $s_1(z,y) = \langle -,+\rangle$: in this case: $u_2(s_1(y, z+1)) = (1 - c(y, z+1))$, $u_1(s_1(z,y)) = q$, $u_2(s_1(z,y)) = 1$ and $u_1(s_1(y, z+1)) = c(y, z+1)$ and the claim holds.

(b) $s_1(y, z+1) = \langle -,+\rangle$ and $s_1(z,y) = \langle +,+\rangle$: in this case: $u_2(s_1(y, z+1)) = 1$, $u_1(s_1(z,y)) = c(z,y)$, $u_2(s_1(z,y)) = (1 - c(z,y))$ and $u_1(s_1(y, z+1)) = q$ and the claim holds.

Next, we assume correctness for $(k-1)$-round games and prove for k-round games. We distinguish between the same cases as we did for the base case:

1. $f(s_k(y, z+1)) = f(s_k(z, y)) = \langle +, + \rangle$**:** the players' utilities are:

$$u_2(s_k(y, z+1)) = (1 - c(y, z+1)) \cdot (1 + u_2(s_{k-1}(y, z+2))) \\ + c(y, z+1) \cdot u_2(s_{k-1}(y+1, z+1))$$

$$u_1(s_k(z, y)) = c(z, y) \cdot (1 + u_1(s_{k-1}(z+1, y))) + (1 - c(z, y)) \cdot u_1(s_{k-1}(z, y+1))$$

$$u_2(s_k(z, y)) = (1 - c(z, y)) \cdot (1 + u_1(s_{k-1}(z, y+1))) + c(z, y) \cdot u_1(s_{k-1}(z+1, y))$$

$$u_1(s_k(y, z+1)) = c(y, z+1) \cdot (1 + u_2(s_{k-1}(y+1, z+1))) \\ + (1 - c(y, z+1)) \cdot u_2(s_{k-1}(y, z+2))$$

   It is not hard to see that by applying the induction hypothesis plus using monotonicity and the facts that $(1 - c(y, z+1)) \geq c(z, y)$ and $(1 - c(z, y)) \geq c(y, z+1)$ the claim holds. By this we have actually shown that a stronger statement holds: $u_2(s_k^{\langle +,+ \rangle}(y, z+1)) \geq u_1(s_k^{\langle +,+ \rangle}(z, y))$ and $u_2(s_k^{\langle +,+ \rangle}(z, y)) \geq u_1(s_k^{\langle +,+ \rangle}(y, z+1))$.

2. $f(s_k(y, z+1)) \neq \langle +, + \rangle$ **and** $f(s_k(z, y)) \neq \langle +, + \rangle$**:** if $y \leq z$ then

$$u_2(s_k(y, z+1)) = 1 + u_2(s_{k-1}(y, z+2)) \ ; \ u_1(s_k(z, y)) = 1 + u_1(s_{k-1}(z+1, y))$$

$$u_2(s_k(z, y)) = q + u_2(s_{k-1}(z+1, y)) \ ; \ u_1(s_k(y, z+1)) = q + u_1(s_{k-1}(y, z+2)).$$

   Thus we can use the induction hypothesis and get that the claim holds. Else, $y \geq z + 1$, and we can again write down the players' utilities and apply the induction hypothesis to get that the claim holds.

3. The players compete in one of $f(s_k(y, z+1)), f(s_k(z, y))$ and do not compete in the other: we will show that the following two lemmas hold:

   Lemma 6.7.12 *For* $y \leq z$*,* $f(s_k(y, z+1)) = \langle +, + \rangle \implies f(s_k(z, y)) = \langle +, + \rangle$.

   Proof: We prove this by using a locking argument very similar to the one we used for Claim 6.7.10. Observe that in both games the lower player is the player with reputation y. Assume towards a contradiction that in $f(s_k(z,y))$ player 2 does not compete. Since player 2 is the lower player, we can use Corollary 6.7.8 to get that $u_2(s_{k-1}(z+1,y)) = (k-1)q$. By applying the induction hypothesis we get that $(k-1)q = u_2(s_{k-1}(z+1,y)) \geq u_1(s_{k-1}(y,z+2))$. Now, since player 1 is the lower player in the game $G_{k-1}(y,z+2)$ we can apply Claim 6.7.2 and conclude that $u_1(s_{k-1}(y,z+2)) = (k-1)q = u_2(s_{k-1}(z+1,y))$. The following chain of inequalities provides a contradiction that in the game $G_k(z,y)$ player 2 prefers to go for the weaker candidate over competing:

$$\begin{aligned} u_2(s_k^{\langle +,+\rangle}(z,y)) \geq u_1(s_k^{\langle +,+\rangle}(y,z+1)) &> q + u_1(s_k(y,z+2)) \\ &= q + u_2(s_{k-1}(z+1,y)). \end{aligned}$$

□

Lemma 6.7.13 *For $y \geq z+1$, $f(s_k(z,y)) = \langle +,+\rangle \implies f(s_k(y,z+1)) = \langle +,+\rangle$.*

Proof: The proof is very similar to the previous lemma. Observe that in both games the lower player is the player with reputation z or $z+1$. Assume towards a contradiction that in $f(s_k(y,z+1))$ player 2 does not compete. Since player 2 is the lower player, we can use Corollary 6.7.8 to get that $u_2(s_{k-1}(y+1,z+1)) = (k-1)q$. By applying the induction hypothesis we get that $(k-1)q = u_2(s_{k-1}(y+1,z+1)) \geq u_1(s_{k-1}(z,y+1))$. Now, since player 1 is the lower player in the game $G_{k-1}(z,y+1)$ we can apply Claim 6.7.2 and conclude that $u_1(s_{k-1}(z,y+1)) = (k-1)q = u_2(s_{k-1}(y+1,z+1))$. The following chain of inequalities provide a contradiction that in the game

$G_k(y, z+1)$ player 2 prefers to go for the weaker candidate over competing:

$$\begin{aligned} u_2(s_k^{\langle +,+\rangle}(y, z+1)) \geq u_1(s_k^{\langle +,+\rangle}(z, y)) &> q + u_1(s_k(z, y+1)) \\ &= q + u_2(s_{k-1}(y+1, z+1)). \end{aligned}$$

$\square$

Thus, we are left with the following two sub-cases:

(a) $f(s_k(y, z+1)) = \langle +, + \rangle$ and $f(s_k(z, y)) = \langle -, + \rangle$: observe that by using the induction hypothesis and monotonicity it is not hard to see that the claim holds since:

$$\begin{aligned} u_2(s_k(y, z+1)) &= u_2(s_k^{\langle +,+\rangle}(y, z+1) > u_2(s_k^{\langle -,+\rangle}(y, z+1) \\ &\geq u_1(s_k^{\langle +,-\rangle}(z, y) = u_1(s_k(z, y)) \\ u_2(s_k(z, y)) &= u_2(s_k^{\langle +,-\rangle}(z, y)) \geq u_2(s_k^{\langle +,+\rangle}(z, y)) \\ &\geq u_1(s_k^{\langle +,+\rangle}(y, z+1)) = u_1(s_k(y, z+1)) \end{aligned}$$

(b) $f(s_k(z, y)) = \langle +, + \rangle$ and $f(s_k(y, z+1)) = \langle -, + \rangle$: observe that by using the induction hypothesis and monotonicity it is not hard to see that the claim holds since:

$$\begin{aligned} u_2(s_k(z, y)) &= u_2(s_k^{\langle +,+\rangle}(z, y)) > u_2(s_k^{\langle +,-\rangle}(z, y)) \\ &\geq u_1(s_k^{\langle -,+\rangle}(y, z+1)) = u_1(s_k(y, z+1)) \\ u_2(s_k(y, z+1)) &= u_2(s_k^{\langle +,-\rangle}(y, z+1) \geq u_2(s_k^{\langle +,+\rangle}(y, z+1) \\ &\geq u_1(s_k^{\langle +,+\rangle}(z, y) = u_1(s_k(z, y)) \end{aligned}$$

$\square$

6.7.2 Additional Properties of the Canonical Equilibrium

We first show that if $u_i(s_k(x_1, x_2)) = kq$ then player i goes for the weaker candidate in $f(s_k(x_1, x_2))$. This immediately implies that if $u_i(s_k(x_1, x_2)) = kq$ then player i in $s_k(x_1, x_2)$ goes for the weaker candidate in every round.

Claim 6.7.14 (generalization of Claim 6.2.4) *The following two statements hold:*

- $u_1(s_k(x_1, x_2)) = kq \implies f(s_k(x_1, x_2)) = \langle -, + \rangle$.
- $u_2(s_k(x_1, x_2)) = kq \implies f(s_k(x_1, x_2)) = \langle +, - \rangle$.

Proof: We prove the claim for player 1 but a similar proof also works for player 2. Assume towards a contradiction that $u_1(s_k(x_1, x_2)) = kq$ but $f(s_k(x_1, x_2)) = \langle +, + \rangle$. By Proposition 6.7.5 we have that $s_k(x_1, x_2)$ is a subgame perfect equilibrium, thus, if player 1 prefers to compete it has to be the case that $u_1(s_k^{\langle +,+ \rangle}(x_1, x_2)) > u_1(s_k^{\langle -,+ \rangle}(x_1, x_2))$. Observe that $u_1(s_k^{\langle -,+ \rangle}(x_1, x_2)) \geq kq$ as a player can always guarantee itself a utility of at least kq in equilibrium. This is in contradiction to the assumption that $f(s_k(x_1, x_2)) = \langle +, + \rangle$. $\square$

Next, based on Claim 6.7.7 we can show that if a player competes and wins then in the next round it prefers competing over going for the weaker candidate.

Claim 6.7.15 (generalization of Claim 6.2.5) *If* $f(s_k(x_1, x_2)) = \langle +, + \rangle$ *then:*

- $f(s_{k-1}(x_1 + 1, x_2)) \in \{\langle +, + \rangle, \langle +, - \rangle\}$
- $f(s_{k-1}(x_1, x_2 + 1)) \in \{\langle +, + \rangle, \langle -, + \rangle\}$

Proof: We first show that $f(s_{k-1}(x_1{+}1, x_2)) \in \{\langle +, + \rangle, \langle +, - \rangle\}$. Assume towards contradiction that $f(s_{k-1}(x_1 + 1, x_2)) = \langle -, + \rangle$. First observe that if $c(x_1, x_2) > q$ then $c(x_1 + 1, x_2) > q$ thus player 1 maximizes its utility by competing in the next round as well. It also has to be the case that $x_1 + 1 < x_2$ since otherwise as the higher player in the game $G_k(x_1 + 1, x_2)$ player 1 should go for the stronger candidate. By Corollary 6.7.8 we have that $u_1(s_{k-1}(x_1 + 1, x_2)) = (k - 1)q$. By monotonicity we have that $u_1(s_{k-1}(x_1, x_2 + 1)) \leq u_1(s_{k-1}(x_1 + 1, x_2)) = (k - 1)q$. Thus, we have that $u_1(s_k(x_1, x_2)) \leq c(x_1, x_2) + (k - 1)q \leq kq$. This implies by Claim 6.7.14 that player 1 does not compete in $f(s_k(x_1, x_2))$ in contradiction to the assumption. The proof of the second statement regarding player 2 is very similar and hence omitted. $\square$

BIBLIOGRAPHY

[1] Scott Aaronson, Allan Borodin, Bernard Chazelle, Oded Goldreich, Shafi Goldwasser, Richard Karp, Michael Kearns, Christos Papadimitriou, Madhu Sudan, and Salil Vadhan. Statement on conceptual contributions in theory, 2008. On-line at http://scottaaronson.com/blog/?p=315.

[2] Daniel M. Abrams and Steven H. Strogatz. Modelling the dynamics of language death. *Nature*, 424:900, 2003.

[3] Daron Acemoglu, Asuman Ozdaglar, and Ali ParandehGheibi. Spread of (mis)information in social networks. *Games and Economic Behavior*, 70(2):194–227, November 2010.

[4] Deepak Agarwal, Bee-Chung Chen, Pradheep Elango, Nitin Motgi, Seung-Taek Park, Raghu Ramakrishnan, Scott Roy, and Joe Zachariah. Online models for content optimization. In *Proc. 21st Advances in Neural Information Processing Systems*, pages 17–24, 2008.

[5] John R. Alford, Carolyn L. Funk, and John R. Hibbing. Are political orientations genetically transmitted? *American Political Science Review*, 99(2):153–167, 2005.

[6] Aris Anagnostopoulos, Ravi Kumar, and Mohammad Mahdian. Influence and correlation in social networks. In *Proc. 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 7–15, 2008.

[7] Elliot Anshelevich, Anirban Dasgupta, Jon M. Kleinberg, Éva Tardos, Tom Wexler, and Tim Roughgarden. The price of stability for network design with fair cost allocation. In *Proc. 45th IEEE Symposium on Foundations of Computer Science*, pages 295–304, 2004.

[8] Sinan Aral, Lev Muchnik, and Arun Sundararajan. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proc. Natl. Acad. Sci. USA*, 106(51):21544–21549, December 2009.

[9] Kenneth J. Arrow. Economic welfare and the allocation of resources for invention. In National Bureau of Economic Research, editor, *The Rate and Direction of Inventive Activity: Economic and Social Factors*, pages 609–626. Princeton University Press, 1962.

[10] Robert Axelrod. The dissemination of culture. *Journal of Conflict Resolution*, 41(2):203–226, April 1997.

[11] Moshe Babaioff, Shahar Dobzinski, Sigal Oren, and Aviv Zohar. On bitcoin and red balloons. In *Proceedings of the 13th ACM Conference on Electronic Commerce*, EC '12, 2012.

[12] Lars Backstrom, Eytan Bakshy, Jon Kleinberg, Tom Lenton, and Itamar Rosenn. Center of attention: How facebook users allocate attention across friends. In *Proc. 5th International Conference on Weblogs and Social Media*, 2011.

[13] Lars Backstrom, Jon M. Kleinberg, and Ravi Kumar. Optimizing web traffic via the media scheduling problem. In *Proc. 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 89–98, 2009.

[14] Eytan Bakshy, Itamar Rosenn, Cameron A. Marlow, and Lada A. Adamic. The role of social networks in information diffusion. In *Proc. 21st International World Wide Web Conference*, 2012.

[15] Eli Ben-Naim, Paul Krapivsky, and Sidney Redner. Bifurcations and patterns in compromise processes. *Physica D*, 183, 2003.

[16] Donald A. Berry and Bert Fristedt. Bernoulli one-armed bandits - arbitrary discount sequences. *Annals of Statistics*, 7:1086–1105, 1979.

[17] David Bindel, Jon Kleinberg, and Sigal Oren. How bad is forming your own opinion? In *Proc. 52nd IEEE Symposium on Foundations of Computer Science*, 2011.

[18] Larry Blume. The statistical mechanics of strategic interaction. *Games and Economic Behavior*, 5:387–424, 1993.

[19] Jesus Zamora Bonilla. The economics of scientific knowledge. In Uskali Mäki, editor, *Handbook of the Philosophy of Science: The Philosophy of Economics.* Elsevier, to appear.

[20] Allan Borodin, Morten N. Nielsen, and Charles Rackoff. (incremental) priority algorithms. *Algorithmica*, 37(4):295–326, 2003.

[21] Jean-Pierre Bourdieu. The specificity of the scientific field and the social

conditions of the progress of reason. *Social Science of Information*, 14(6):19–47, 1975.

[22] Danah Boyd. Is facebook destroying the american college experience?, March 1st 2013. Apophenia (Danah Boyd's blog), http://www.zephoria.org/thoughts/archives/2013/03/01/facebook-college.html.

[23] Yuri Boykov, Olga Veksler, and Ramin Zabih. Fast approximate energy minimization via graph cuts. In *Proc. 7th International Conference on Computer Vision*, pages 377–384, 1999.

[24] Yann Bramoullé, Habiba Djebbari, and Bernard Fortin. Identification of peer effects through social networks. *Journal of Econometrics*, 159:41–55, 2009.

[25] Andrei Z. Broder, Adam Kirsch, Ravi Kumar, Michael Mitzenmacher, Eli Upfal, and Sergei Vassilvitskii. The hiring problem and lake wobegon strategies. *SIAM Journal on Computing*, 39(4):1233–1255, 2009.

[26] Moira Burke and Robert Kraut. Mopping up: Modeling wikipedia promotion decisions. In *Proc. CSCW'08: ACM Conference on Computer-Supported Cooperative Work*, pages 27–36, 2008.

[27] Claudio Castellano, Santo Fortunato, and Vittorio Loreto. Statistical physics of social dynamics. *Reviews of Modern Physics*, 81:591–646, 2009.

[28] Jiawei Chen, Ulrich Doraszelski, and Joseph E. Harrington. Avoiding market dominance: product compatibility in markets with network effects. *RAND Journal of Economics*, 40(3):455–485, 2009.

[29] Flavio Chierichetti, Jon Kleinberg, and Sigal Oren. On discrete preferences and coordination. In *Proc. 14th ACM Conference on Electronic Commerce*, 2013.

[30] Fan R. K. Chung. *Spectral Graph Theory*. American Mathematical Society, 1997.

[31] Jere M. Cohen. Sources of peer group homogeneity. *Sociology in Education*, 50:227–241, October 1977.

[32] Pranav Dandekar, Ashish Goel, and David T. Lee. Biased assimilation,

homophily, and the dynamics of polarization. *Proc. Natl. Acad. Sci. USA*, 110(15), 2013.

[33] Partha Dasgupta and Eric Maskin. The simple economics of research portfolios. *Economic Journal*, 97(387):581–95, 1987.

[34] Guillaume Deffuant, David Neau, Frederic Amblard, and Gérard Weisbuch. Mixing beliefs among interacting agents. *Advances in Complex Systems*, 3:87–98, 2000.

[35] Morris H. DeGroot. Reaching a consensus. *Journal of the American Statistical Association*, 69:118–121, 1974.

[36] Peter M. DeMarzo, Dimitri Vayanos, and Jeffrey Zweibel. Persuasion bias, social influence, and unidimensional opinions. *Quarterly Journal of Economics*, 118(3):909–968, August 2003.

[37] John P. Dickerson, Ariel D. Procaccia, and Tuomas Sandholm. Dynamic matching via weighted myopia with application to kidney exchange. In *Proc. 26th AAAI Conference on Artificial Intelligence*, 2012.

[38] Glenn Ellison. Learning, local interaction, and coordination. *Econometrica*, 61:1047–1071, 1993.

[39] Uriel Feige, David Peleg, and Guy Kortsarz. The dense k-subgraph problem. *Algorithmica*, 29(3):410–421, 2001.

[40] William Feller. *An Introduction to Probability Theory and Its Applications*, volume 1. John Wiley & Sons, third edition, 1968.

[41] Dimitris Fotakis, Spyros Kontogiannis, Elias Koutsoupias, Marios Mavronicolas, and Paul Spirakis. The structure and complexity of nash equilibria for a selfish routing game. *Theoretical Computer Science*, 410:3305–3326, August 2009.

[42] Noah E. Friedkin and Eugene C. Johnsen. Social influence and opinions. *Journal of Mathematical Sociology*, 15(3-4):193–205, 1990.

[43] Michael R. Garey and David S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W.H. Freeman, 1979.

[44] Bezalel Gavish and Suresh Sridhar. Computing the 2-median on tree networks in $o(n \lg n)$ time. *Networks*, 26:305–317, 1995.

[45] J.C. Gittins and D.M. Jones. A dynamic allocation index for the sequential design of experiments. In J. Gani, editor, *Progress in Statistics*, pages 241–266. North-Holland, Amsterdam, NL, 1974.

[46] Michel X. Goemans, Li Li, Vahab S. Mirrokni, and Marina Thottan. Market sharing games applied to content distribution in ad hoc networks. *IEEE Journal on Selected Areas in Communications*, 24(5):1020–1033, 2006.

[47] A. J. Goldman. Optimal center location in simple networks. *Transportation Science*, 5:212–221, 1971.

[48] Oded Goldreich. Innovations in Computer Science (ICS), 2009. On-line at http://www.wisdom.weizmann.ac.il/∼oded/ics.html.

[49] Benjamin Golub and Matthew O. Jackson. Naive learning in social networks: Convergence, influence and the wisdom of crowds. *American Economic Journal: Microeconomics*, 2(1):112–149, 2010.

[50] Gene H. Golub and Charles F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, third edition, 1996.

[51] Georg Gottlob, Gianluigi Greco, and Francesco Scarcello. Pure nash equilibria: hard and easy games. In *Journal of Artificial Intelligence Research*, pages 215–230. ACM Press, 2003.

[52] Tim Gowers, Gil Kalai, Michael Nielsen, and Terence Tao. The polymath blog. http://polymathprojects.org/about/.

[53] Mark Granovetter. *Getting a Job: A Study of Contacts and Careers*. University of Chicago Press, 1974.

[54] Guillaume Haeringer and Myrna Wooders. Decentralized job matching. *International Journal of Game Theory*, 40:1–28, 2011.

[55] Kais Hazma. The smallest uniform upper bound on the distance between the mean and the median of the binomial and Poisson distributions. *Statistics and Probability Letters*, 23:21–25, 1995.

[56] Rainer Hegselmann and Ulrich Krause. Opinion dynamics and bounded confidence: Models, analysis and simulation. *Journal of Artificial Societies and Social Simulation*, 5(3), 2002.

[57] Nicole Immorlica, Adam Tauman Kalai, Brendan Lucier, Ankur Moitra, Andrew Postlewaite, and Moshe Tennenholtz. Dueling algorithms. In *Proc. 43rd ACM Symposium on Theory of Computing*, pages 215–224, 2011.

[58] Nicole Immorlica, Robert D. Kleinberg, and Mohammad Mahdian. Secretary problems with competing employers. In *Proc. 2ndWorkshop on Internet and Network Economics*, pages 389–400, 2006.

[59] Matthew O. Jackson. *Social and Economic Networks*. Princeton University Press, 2008.

[60] Denise B. Kandel. Homophily, selection, and socialization in adolescent friendships. *American Journal of Sociology*, 84(2):427–436, September 1978.

[61] David Kempe, Jon Kleinberg, Sigal Oren, and Aleksandrs Slivkins. Selection and influence in cultural dynamics. In *Proc. 14th ACM Conference on Electronic Commerce*, 2013.

[62] Philip Kitcher. The division of cognitive labor. *Journal of Philosophy*, 86(1):5–21, January 1990.

[63] Philip Kitcher. *The Advancement of Science*. Oxford University Press, 1993.

[64] Jon Kleinberg and Sigal Oren. Mechanisms for (mis)allocating scientific credit. In *Proceedings of the 43rd annual ACM symposium on Theory of computing*, STOC '11, pages 529–538, 2011.

[65] Jon M. Kleinberg and Éva Tardos. Approximation algorithms for classification problems with pairwise relationships: metric labeling and markov random fields. *Journal of the ACM*, 49(5):616–639, 2002.

[66] David Krackhardt. A plunge into networks. *Science*, 326:47–48, 2 October 2009.

[67] Timothy LaFond and Jennifer Neville. Randomization tests for distinguishing social influence and homophily effects. In *Proc. 19th International World Wide Web Conference*, pages 601–610, 2010.

[68] Jure Leskovec, Dan Huttenlocher, and Jon Kleinberg. Governance in social media: A case study of the Wikipedia promotion process. In *Proc. 4th International Conference on Weblogs and Social Media*, 2010.

[69] Miller McPherson, Lynn Smith-Lovin, and James M. Cook. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27:415–444, 2001.

[70] Robert K. Merton. The Matthew effect in science. *Science*, 159(3810):56–63, January 1968.

[71] Robert K. Merton. *The Sociology of Science: Theoretical and Empirical Investigations*. University of Chicago Press, 1973.

[72] Dov Monderer and Lloyd S. Shapley. Potential games. *Games and Economic Behavior*, 14:124–143, 1996.

[73] Stephen Morris. Contagion. *Review of Economic Studies*, 67:57–78, 2000.

[74] Dale T. Mortensen and Christopher A. Pissarides. New developments in models of search in the labor market. In O. Ashenfelter and D. Card, editors, *Handbook of Labor Economics*, volume 3, pages 2567–2627. Elsevier, 1999.

[75] Sendhil Mullainathan, Joshua Schwartzstein, and Andrei Shleifer. Coarse thinking and persuasion. *Quarterly Journal of Economics*, 123(2):577–619, 2008.

[76] Sean A. Munson and Paul Resnick. Presenting diverse political opinions: how and how much. In *Proc. 28th ACM Conference on Human Factors in Computing Systems*, pages 1457–1466, 2010.

[77] Sean A. Munson, Daniel Xiaodan Zhou, and Paul Resnick. Sidelines: An algorithm for increasing diversity in news and opinion aggregators. In *Proc. 3rd International Conference on Weblogs and Social Media*, 2009.

[78] David G. Myers. The internet as social amplifier. In John Brockman, editor, *How is the internet changing the way you think?* 2010.

[79] Noam Nisan, Tim Roughgarden, Eva Tardos, and Vijay V. Vazirani. *Algorithmic Game Theory*. Cambridge University Press, New York, NY, USA, 2007.

[80] Christos Papadimitriou. Algorithms, games, and the internet. In *Proceedings of the thirty-third annual ACM symposium on Theory of computing*, STOC '01, pages 749–753, New York, NY, USA, 2001. ACM.

[81] Eli Pariser. *The Filter Bubble: What the Internet Is Hiding from You*. Penguin, 2011.

[82] Marco Patriarca and Teemu Leppanen. Modeling language competition. *Physica A*, 338(1–2):296–299, July 2004.

[83] Charles S. Peirce. Note on the theory of economy of research. In Arthur W. Burks, editor, *Collected Papers of Charles Sander Peirce*, volume 7, pages 76–83. Harvard University Press, 1958.

[84] Robin Pemantle. A survey of random processes with reinforcement. *Probability Surveys*, 4:1–79, 2007.

[85] Keith Poole and Howard Rosenthal. Patterns of congressional voting. *American Journal of Political Science*, 35(1):228–278, 1991.

[86] Richard Rogerson, Robert Shimer, and Randall Wright. Search-theoretic models of the labor market: A survey. *Journal of Economic Literature*, 43(4):959–988, 2005.

[87] Margaret W. Rossiter. The matthew matilda effect in science.

[88] Alvin E. Roth and Marilda Sotomayor. *Two-sided matching: A study in game-theoretic modeling and analysis*. Cambridge University Press, 1990.

[89] Tim Roughgarden. *Selfish Routing and the Price of Anarchy*. MIT Press, 2005.

[90] Tim Roughgarden. Intrinsic robustness of the price of anarchy. In *Proc. 41st ACM Symposium on Theory of Computing*, pages 513–522, 2009.

[91] Esther-Mirjam Sent. Econmics of science: Survey and suggestions. *Journal of Economic Methodology*, 6(1):95–124, March 1999.

[92] Cosma R. Shalizi and Andrew C. Thomas. Homophily and contagion are generically confounded in observational social network studies. *Sociological Methods and Research*, 40:211–239, 2011.

[93] Robert Shimer and Lones Smith. Assortative matching and search. *Econometrica*, 68(2):343–369, March 2000.

[94] Stergios Skaperdas. Contest success functions. *Economic Theory*, 7:283–290, 1996.

[95] David Spector. Rational debate and one-dimensional conflict. *Quarterly Journal of Economics*, 115(1):181–200, 2000.

[96] Dietrich Stauffer, Xavier Castello, Victor M. Eguiluz, and Maxi San Miguel. Microscopic Abrams-Strogatz model of language competition. *Physica A*, 374(2):835–842, February 2007.

[97] Klaus Stein and Claudia Hess. Does it matter who contributes: A study on featured articles in the German Wikipedia. In *Proc. 18th ACM Conference on Hypertext and Hypermedia*, pages 171–174, 2007.

[98] Michael Strevens. The role of the priority rule in science. *Journal of Philosophy*, 100(2):55–79, 2003.

[99] Michael Strevens. The role of the Matthew Effect in science. *Studies in History and Philosophy of Science*, 37:159–170, 2006.

[100] Eric Sun, Itamar Rosenn, Cameron Marlow, and Thomas M. Lento. Gesundheit! Modeling contagion through Facebook News Feed. In *Proc. 3rd International Conference on Weblogs and Social Media*, 2009.

[101] Cass R. Sunstein. *Republic.com 2.0.* Princeton University Press, 2009.

[102] Éva Tardos and Tom Wexler. Network formation games and the potential function method. In Noam Nisan, Tim Roughgarden, Éva Tardos, and Vijay Vazirani, editors, *Algorithmic Game Theory*, pages 487–516. Cambridge University Press, 2007.

[103] Gordon Tullock. Efficient rent seeking. In G. Tullock J. Buchanan, R. Tollison, editor, *Towards a theory of the rent-seeking society.* Texas A&M Press, 1980.

[104] Adrian Vetta. Nash equilibria in competitive societies, with applications to facility location, traffic routing and auctions. In *Proc. 43rd IEEE Symposium on Foundations of Computer Science*, pages 416–425, 2002.

[105] James Zou, Sujit Gujar, and David C. Parkes. Tolerable manipulability in dynamic assignment without money. In *Proc. 24th AAAI Conference on Artificial Intelligence*, 2010.