Bilateral Exchanges in Social Networks and the Design of Public Institutions

by

Ankur Mani

B.Tech, Electrical Engineering, Indian Institute of Technology, Delhi
MS, Electrical Engineering, Arizona State University, Tempe, AZ

Submitted to the Program in Media Arts and Sciences
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2013

© Massachusetts Institute of Technology 2013. All rights reserved.
Bilateral Exchanges in Social Networks and the Design of Public Institutions

by

Ankur Mani

Submitted to the Program in Media Arts and Sciences on September 26, 2013, in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Abstract

In this thesis, I study bilateral exchanges in social networks. I integrate the market based approach in economics and network based approach in sociology. I also give a joint model of externalities and peer pressure in networks and propose mechanisms to control externalities in networked societies.

The study of exchanges in networks is big and growing. In this thesis, I focus on exclusive exchanges or matching in bipartite networks where the matched couples perform an economic exchange with each other. This thesis makes three contributions to matching theory. First, I relax the standard assumptions of costless transfers between matched couples and introduce a new algorithm to compute stable matchings or core with proofs of correctness and convergence. Second, I introduce a new distributed dynamics and show that it converges to stable matching. Third, I give an axiomatic characterization of fair outcomes in bipartite matching based upon a collective bargaining argument and give a fast algorithm to compute it.

I also introduce a joint model of externalities and peer pressure in networks and propose new design of public institutions that are more efficient in the control of externalities in networked societies. The design principles I propose are well suited for environmental policies, and health policies. The same design principles can also be used for marketing and revenue management for products and services with network externalities. I demonstrate the efficiency of my policies in two experiments. The first experiment showed that my proposed policies give better outcome in promoting higher physical activity among people than traditional policies such as Pigouvian subsidies. The second experiment showed that my proposed policy can help promote energy conservation and I report much higher conservation than any achieved through price shocks or any other energy conservation experiment in my knowledge.

Thesis Supervisor: Alex (Sandy) Pentland
Title: Toshiba Professor of Media Arts and Sciences
Acknowledgments

I am indebted to many people who both directly and indirectly contributed to this thesis. First I would like to thank Alex (Sandy) Pentland for taking me in his tutelage at MIT, providing me with freedom to explore and define my problems while guiding me towards problems of importance to society. I also thank him for being a close collaborator and supporting me during unproductive and productive times equally. His faith in me during the times the research did not seem to move along provided me the confidence to push on and break the barriers. Next, I would like to thank Asuman Ozdaglar. I have learnt a lot from Asu. The most important thing that I learnt from Asu is that the model I work on should reflect the real world and accuracy should not be given up so that it becomes easy to get the results. Her influence is seen in my work on matching and the model is a result of several discussions with Asu. Next I would like to thank Parag Pathak. Parag is one of the most crisp thinkers I know. My interactions with Parag has helped me identify crisp problems and make precise claims about my results.

I thank my collaborators, Iyad Rahwan, Kobi Gal and Claire Michelle Loock, with whom I worked on several experiments that fed into the thesis directly and indirectly. I would also like to thank my other collaborators over the years at MIT with whom I had the pleasure to work on other research. I thank Lav Varshney, Peter Krafft, Wen Dong, Bruno Lepri, Alex Rutherford, Wesley Gifford and Anshul Sheopuri.

I thank Yahoo Key Scientific Challenges and Martin Family Fellowship for Sustainability for the support.

I thank several colleagues who provided valuable comments on my work. I would like to thank Azaraksh Malekian, Ilan Lobel, David Lazer, Yusuke Narita, Benjamin Golub, Itay Fainmasser, participants of the Harvard-MIT networks reading group, MIT CS-Econ reading group, MIT Econ theory lunch, MIT Machine Learning Tea, MIT Systems Dynamics reading group, MIT-U Albany colloquium, and Pentlandians over the years, namely Ben Waber, Taemie Kim, Daniel Olguin Olguin, Wei Pan, Nadav Aharony, Yves-Alexandre DeMontjoye, Manuel Cebrian, Nicole Freedman, Vivek
Singh and several other visitors.

I thank several people at the Media Lab, Edgerton House and other parts of MIT who made life at MIT wonderful and exciting. I would like to thank Haitham Hassanieh, Ana Luisa Santos, Selene Mota, Ehsan Hoque, Inna Lobel, Kathy Miu, Jon Medley, Gleb Akselrod, Cody Gilleland, Sean Smeland, Leeland Ekstrom, Vivek Sharma, Smita Sharma, Shreeharsh Kelkar, Sushmit Goswami, Josh Bendavid, Georgious Papadopoulos and others.

I thank my friends from IIT Delhi and other close friends. I especially thank Hatinder Madan, Dhaval Gothi, Ashish Kapoor, Alok Ranjan, Rohit Dube, Nikhil Garg, Ashutosh Punhani, Sachin Jain, Jai Singh, Chinmay, Kartikey Kumar, Rineesh Bansal, Puru Gupta, Amit Kataria, Roopak Hooda, Vishal Makani, Manish Agrawal, Jayant Kumar, Abhishek Kashyap, Gaurav Diwan, Ritu Shrivastava, Amit Kapoor, Rajesh Kumar, Gunjan Shukla, Aman Singh, Aditya Renduchintala and others. These guys were responsible for the wonderful days during my undergrad and onwards.

I sincerely thank my parents, Amma and Papaji, and my sisters, Priyanka Mani and Swina Mani for their love, faith and continued support throughout my life. I also thank my new family, Mumsie, Pete and Jon Sheren for their love and care they have given me in the past few years. Finally, I thank my loving, caring and supportive wife, Ila. She has been with me through all the ups and downs, through the frustrating days and nights when I was stuck to days when I had a breakthrough. I could not have done it with you. This one is for you BLLC.
Bilateral Exchanges in Social Networks and the Design of Public Institutions

by

Ankur Mani

The following people served as readers for this thesis:

Thesis Reader ........................................................

Asuman Ozdaglar
Professor
Department of Electrical Engineering and Computer Science
Bilateral Exchanges in Social Networks and the Design of Public Institutions

by

Ankur Mani

The following people served as readers for this thesis:

Thesis Reader .........................................................

Parag Pathak
Professor
Department of Economics
# Contents

Cover page .......................... 1

Abstract .................................. 3

Acknowledgments ...................... 5

Contents ................................ 11

List of Figures ......................... 15

List of Tables ......................... 17

1 Introduction .......................... 19

1.1 Introduction .......................... 19

1.2 Related Work ......................... 21

1.3 Outline ................................ 27

2 Generalized Stable Matching in Bipartite Networks 29

2.1 Introduction .......................... 29

2.1.1 Application Scenarios .............. 31

2.1.2 Related Work ......................... 32

2.1.3 Our Contributions ................. 34

2.1.4 Outline ................................ 35

2.2 Model ................................ 35

2.2.1 Network and Payoffs .............. 35
2.12.1 Proof of Lemma 5 ........................................... 65
2.12.2 Proof of Theorem 3 ........................................... 66
2.12.3 Proof of Corollary 3 ........................................... 67
2.13 Appendix: Proofs of Propositions 1 and 2 ...................... 68
  2.13.1 Proof of Proposition 1 ..................................... 68
  2.13.2 Proof of Proposition 2 ..................................... 69
2.14 Appendix: Graph Theoretic Terminology .......................... 70
  2.14.1 Paths ......................................................... 70
  2.14.2 Alternating Paths, Alternating Trees and Near-Perfect Matchings 71
2.15 Appendix: Structural Properties of the Hungarian Forest .......... 72
2.16 Appendix: Proof of Lemma 1 .................................... 74
  2.16.1 Proof of Lemma 4 ........................................... 75
2.17 Appendix: Examples .............................................. 86

3 Fairness in Bipartite Matching With Transfers .......................... 95
  3.1 Introduction ..................................................... 95
  3.2 Model ............................................................. 98
    3.2.1 Balancedness ............................................... 98
    3.2.2 Fairness .................................................... 99
  3.3 Algorithm ....................................................... 100
  3.4 Main Results ..................................................... 104
  3.5 Conclusions and Future Work .................................... 104
  3.6 Proof of Theorem 5 ............................................. 105
  3.7 Proof of Theorem 6 ............................................. 107
  3.8 Proof of Theorem 7 ............................................. 109

4 Design of Public Institutions for Networked Societies .................. 111
  4.1 Introduction ..................................................... 112
  4.2 Main Results ..................................................... 116
    4.2.1 Overview of Mechanism ................................... 116
4.2.2 Applications to Promoting Active Lifestyle and Conserving Electricity

4.3 Localizing Externalities in social networks: Inducing Peer Pressure to Promote Cooperation

4.3.1 Model of Externalities

4.3.2 Externalities with Peer Pressure

4.3.3 Assumptions

4.3.4 Equilibrium with Peer Pressure

4.3.5 Pigouvian Mechanism (Direct Reward)

4.3.6 Social Mechanisms (Rewarding the Peers)

4.3.7 Efficiency and Cost of Social Reward

4.3.8 Illustrative Example: Promoting Energy Efficiency

4.3.9 Properties of the Subgame Perfect Equilibria of the two stage game

4.3.10 Proof of Proposition 17

4.3.11 Proof of Theorem 9

4.3.12 Proofs for Welfare Comparison Results

4.4 Experiments and Applications to Public Policy

4.4.1 Promoting Physical Activity

4.4.2 Promoting Energy Conservation

4.5 Conclusions

5 Conclusions and Future Work

Bibliography
List of Figures

2-1 Running the Hungarian Algorithm to find the stable matching in the standard weighted matching problem ........................................ 87
2-2 Running the Hungarian Algorithm to find the stable matching in the standard weighted matching problem: Iteration 1 ............... 88
2-3 Running the Hungarian Algorithm to find the stable matching in the standard weighted matching problem, only the offers on one of the alternating trees is updated: Iteration 2 .............................. 89
2-4 Running the Hungarian Algorithm to find the stable matching in the standard weighted matching problem, only the offers on one of the alternating trees is updated: Iterations 3 and 4 ...................... 90
2-5 The Generalized stable matching problem ............................... 91
2-6 The Hungarian Algorithm does not work for the generalized stable matching problem ......................................................... 92
2-7 The Generalized Hungarian Algorithm find the A-maximum generalized stable matching for the generalized stable matching problem ..... 93

4-1 (a) Individuals A and B (red) actions cause a global externality on the rest of society (blue). This behavior can be encouraged/discouraged by internalizing the externality in the form of subsidy or tax on A and B directly; (b) Localizing externalities to the peers (yellow) of individuals A and B incentivizes their respective peers to use peer influence to encourage/discourage the behavior causing the positive/negative externality, respectively. ......................................................... 117
4-2 (navy) Raw utility of consumption, maximized at the equilibrium consumption of $X = 79.62$. The peer pressure will lower the equilibrium consumption only slightly to $X = 67.46$ due to high marginal cost of pressure; (green) Externality experienced due to other agents’ consumption; (red) Total utility curve is the difference between the navy and green curves, and is maximized at the (much lower) socially optimal consumption level of $X = 31.19$.

4-3 (navy) Budget required for rewards under Pigouvian mechanism for a target consumption level; (green) Budget required for rewards under Social mechanism for a desired consumption level. It increases much slower as the target consumption level is reduced.

4-4 (navy) Redistribution loss under Pigouvian mechanism for a target consumption level; (green) Redistribution loss under Social mechanism for a desired consumption level. It increases much slower as the target consumption level is reduced; (red) Social loss under Social mechanism for a desired consumption level; (magenta) Total (redistribution + social) loss under Social mechanism for a desired consumption level.

4-5 Average energy consumption. The pattern of consumption correlates with the outside temperature suggesting that people spent more time indoors when it was cold. The consumption of the treatment group is lower than people in the control group.

4-6 Average energy consumption against average weekly temperature. Consumption increases with the decrease in outside temperature suggesting that people spent more time indoors when it was cold. The consumption of the treatment group is lower than people in the control group.
List of Tables

4.1 Regression results: \( \alpha \) is the average consumption, \( \beta \) is the temperature effect and \( \gamma \) is the treatment effect. The p-values are very low suggesting that the regression is statistical significant.
Chapter 1

Introduction

1.1 Introduction

How different are the characteristics of societies that are constrained to local interactions in networks as compared to societies where all interactions happen in organized markets? Among most species and even in modern human societies, exchange, whether of food, information, or labor, naturally tends to occur locally, as encounters happen between nearby individuals in networks. In the absence of coordination across multiple individuals and arbitrage, such exchanges are also often bilateral. Neoclassical economic theory has provided us with a deep understanding of organized markets. We understand equilibrium in these markets as well as the welfare properties and dynamics. This understanding has guided the design of public institutions and the control of externalities for centuries. The market model of society assumes that individual actions are just response to the market norms or prices and are independent of social structure and social interactions. Using the model of organized markets, public policies are designed for an independent population and completely ignore the interactions among individuals in the population. On the other hand, sociological theories of society consider economic actions to be embedded in a social structure. Economic Sociology and in particular network exchange theory in sociology has provided us with very important insights into the properties of local exchanges, the origin of social capital, power in social relationships, fairness in local exchanges and
peer-pressure. We still do not understand however, how these local exchanges govern the large scale properties of networked societies - stability, welfare, dynamics, and fairness - and how can we use peer-pressure to improve social welfare. This lack of understanding has prohibited us from designing effective public institutions and policies for networked societies. Today we have easy availability of ‘big data’ about social and economic interactions. To use this new resource for the betterment of society, we need to understand exchanges in networked societies and build tools for computing the structure of stable and fair networked societies and predict how they may respond to policy changes. This dissertation is an exploration of some questions on networked societies.

There are two kinds of bilateral exchanges in networked societies: (a) ‘Exclusive exchanges’, where each agent can perform an exchange with at most one other agent and (b) ‘Non-exclusive exchanges.’ Exclusive exchanges form the foundation of important social and economic relationships such as marriage and employment and deserve special attention. The fundamental mathematical structure of exclusive exchanges is close to matching that has been well studied in mathematics. While the sociologists were studying matching in exchange networks, economists studied matching with respect to markets with indivisible goods as well. There is a lot of work on matching in economics, sociology, computer science, and operations research, that is related to exclusive exchanges. I focus only on exclusive exchanges and in particular matching in bipartite networks. The work on exchange networks in sociology and matching in economics often assumes that matched agents can costless transfer utilities amongst themselves. This is often justified by assuming that there is a good to be exchanges between the matched agents that has quasilinear utility for both agents. An example of such a good is money. Often such assumption is not valid as in the case of matched couples that do not exchange money for services in the marriage. I study a model of matching that relaxes this assumption. The tools I develop for the study of exclusive exchanges build upon and expand the horizons of the work on matching. My exploration focuses on exclusive exchanges or matching. My results on matching open several questions for further exploration.
When the model of society changes from the market to the network, it is obvious to ask if the design of public institutions change as well. In particular, should the control of externalities in the society be done by mechanisms that act outside the market. Traditionally, taxation and subsidies have been the mechanisms to control externalities in societies. I propose alternate mechanisms for the design of effective public institutions that work with non-market interventions. Although my particular examples concern the design of public policies, these recommendations can also be used for revenue management and marketing of goods with externalities.

The contributions of this dissertation are:

1. A theoretical framework for the study of exclusive bilateral exchanges or matching in networked societies.

2. Computational tools and algorithms with proofs of correctness and convergence for computing the structure of stable and fair exchanges in bipartite networks.

3. Natural dynamics that converge to stable matching in such networks.

4. Proposal for the design of public policy and revenue management policies using peer pressure for networked societies.

5. Experimental evidence that the proposed design of public policies using peer-pressure is more efficient than existing policies such as individual taxation and subsidies and its applications to health and energy policies.

In the next section, I present a concise survey of related work.

1.2 Related Work

My work lies at the boundary of several disciplines including sociology, economics, mathematics and computer science. It is hard to do justice to the wide spectrum of work but I try to present a sample of some of the important and relevant pieces of work. The section is organized in four parts. In the first part, I discuss some
classical literature on exchanges in markets and networks. In the second part, I discuss literature on exclusive exchanges. In the third part, I discuss fairness in exclusive exchanges. In the fourth part, I discuss literature on the control of externalities through social interventions.

**Exchanges in Markets and Networks**

Market exchanges are widely studied in economics. General equilibrium theory in neoclassical economics, studies the society where individuals are price-taking agents interacting within open markets and shows the existence of stable outcomes and convergent dynamics in open markets. In the classical work [116], Walras suggested an auction in which the auctioneer announces prices of the goods and the individuals in the society declare the quantities of each good they would like to buy. The auctioneer then adjusts the prices such that the demand of each good is equal to its supply. In [7], Arrow and Debreu prove the existence of a set of prices for goods in the market of buyers and sellers such that the aggregate demand for each good equal the aggregate supply. The first fundamental theorem of welfare economics suggests that allocations in general equilibrium are pareto-efficient. This line of work however, does not investigate the interactions among individuals in the society. The assumption is that there is an owner of the market that takes all the goods and reallocates them at the right prices. Forward-looking versions of the tatonnement process introduced by Walras including bargaining between individuals in the market have been shown to converge to the Walrasian equilibrium. A good survey is provided in [45].

Some of the early works in exchanges in networks started with the classical work of Granovetter [50]. Granovetter introduced the concept of ‘embeddedness’ and suggested that economic actions are embedded in social networks and not in abstract markets. This led to a new era of ‘economic sociology,’ and a long debate on the goods and bads of the markets and networks. The standard assumption in the markets, “my money is as good as yours,” was questioned in networks [117] and new concepts such as ‘social capital’ were introduced [31][90]. Later in [51], Granovetter commented on how network structures impact economic outcomes. This work is far from complete.
and asks several questions about computation of efficient social structures, predicting the welfare properties of a given social network and others. A good survey of the work on economic sociology is given in [118].

Recently exchanges in networks have been investigated in computer science literature along the lines of bargaining. Some initial work was done in [18]. In [18] the authors study a network consisting of buyers, sellers and price-setting agents. This is a first attempt to extend the Arrow-Debreu model to the networks, where instead of one price-setting agent for the whole market there are several price-setting agents.

**Exclusive Bilateral Exchanges in Networks**

Earlier works on exclusive exchanges in networks in sociological literature provided several theories regarding the equilibrium in exchange networks instead of markets. The work on exchange networks was originally studied in [34],[33]. The authors characterized the distribution of power and value in a social network when individuals in the network perform exclusive bilateral exchanges. They studied a particular kind of equilibrium where all exchanges are locally fair. Recent literature [20],[36],[78] has extended this work in the light of the axiomatic Nash bargaining solution. However, there has been no mathematical framework that analyzes the dynamics of the exchanges in such networks, stability properties of the exchanges networks and the convergence to the equilibrium. Almost all the work on exchanges in networks deals with bargaining and assumes there is a pie at each edge of the network to be split between the individuals at the end of the edges. This assumption is similar to work on matching of workers to firms [62] in economics literature or buyers to sellers [69],[66]. The first work on exclusive exchanges in bipartite networks rather than markets is by Kranton and Minehart [66]. The work discusses a two-stage game, the first stage being network formation and the second, exchanges in networks. The authors characterize the topological conditions of the networks that are socially efficient and show that rational individuals can form socially efficient networks in the absence of production costs. The authors consider a Walrasian auction for the competitive outcome unlike the myopic dynamics discussed in our work.
The matching literature in economics is closely related to our problem of exclusive exchanges. Two particular versions of the problem have been studied in particular. The first is the marriage problem (one-to-one matching), while the second is the college admissions problem (many-to-one matching) [46]. The applications to market design have been studied in the context of national resident matching and school assignment for students in Boston and New York [1],[2]. In these problems, individuals had exogenous preferences for partners and there was no concept of utility transfers or payments to the partners. An important extension of this work has been to study the case when payments can be made between partners. The classical works extend the marriage problem to the markets with indivisibilities [69] and the college admissions problem to the employment problem [62]. More recent work that unifies the two works is [53]. The work closest to exclusive exchanges is [69]. The authors in [69] study allocation of goods, where each seller has exactly one good and each buyer demands only one good. Thus, an exchange is made when a buyer and a seller come to an agreement on the price. The authors show that the society is stable for a set of prices and all stable outcomes are socially efficient. Several algorithms have been proposed to compute this price including [38] in which the authors present the Walrasian auction for indivisible goods. Again, this line of work assumes an organized market and does not investigate local exchanges within networks. A good survey of the matching literature until 1992 can be found in [96].

In my work on exclusive exchanges, I model the bipartite matching case, as in the job matching. I relax the assumption of costless transfer of utilities among matched individuals. I provide algorithms for computing a stable set of exchanges that resists profitable deviations by any subset of agents in the network. I also study natural convergent myopic dynamics that are simple and memory-less.

**Fairness in Exclusive Exchanges**

The question of fairness in matching with transfers among matched pairs is important and well debated. The problem is often motivated by procedural fairness [92] because of the prevalence of central assignments such as the medical resident match-
Procedural fairness guarantees ex-ante fairness even an assignment that was generated by a fair procedure may not be ex-post fair. The common notion of fairness in two-sided matching is median stable matching studied in [109][64]. In [100] the authors extended the work to matching with transfers between matched pairs. These definitions of fairness are motivated by procedural fairness and the geometry of the core. They do not, however, provide a convincing argument for fairness as provided by the axiomatic characterization of the Nash bargaining solution in the case of two agent bargaining. The Nash Bargaining Solution [82] or the Kalai-Smorodinsky solution [56] are the most commonly accepted axiomatic solutions considered to be fair outcomes to a two agent bargaining problem. The most common extension of the Nash bargaining solution to larger coalitions is the Shapley value [101]. However, the Shapley value may not even be inside the core of the matching. A new extension, the coalitional Nash bargaining solution [32], restricts the outcomes to the core but also restricts the set of possible coalitions to deviate from the grand coalition. These problems arise mainly because both axiomatic solutions rely on transferable utilities across any sets of players that is not naturally present in the matching problem. Another variation is the Myerson value [81] that restricts coalitions to the connected components of a graph. However, the exclusive bilateral exchange problem has an even stronger constraint, that each player can perform exchange with at most one of its neighbors.

In the sociology literature on exchanges in networks [34][33], the authors have introduced a notion of ‘balancedness.’ An outcome of the network exchange game is balanced if the payoff profile satisfies the Nash bargaining solution for each exchange in the exchange network. Several experiments conducted by the authors show that people in networks also converge to balanced outcomes. Recently, in computer science literature, [65] Klienberg and Tardos have also explored this notion of fairness and have provided an algorithm to compute the set of balanced outcomes. The problem with the balanced outcomes is that they may not be globally fair and there may be infinitely many balanced outcomes that are only locally fair.

I study the concept of fairness in exclusive exchanges and show that a new concept
of fairness is needed. I introduce the new concept of fairness for exclusive bilateral exchanges that is consistent with the existing concepts and give an algorithm to compute the set of fair exclusive exchanges.

**Control of Externalities in Networked Societies**

Traditional policies to control externalities and promote cooperation include government quotas, and Pigouvian taxation or subsidies that make individuals *internalize* the externality they incur. However, such solutions are based on a model of externality that discounts local exchanges and peer pressure. Recent economic models of peer pressure show that pressure can act as a mechanism for improving the social welfare of the group [24].

There is growing evidence of the power of social influence in general, and peer pressure in particular, in promoting cooperative behavior. In evolutionary biology, mechanisms like direct reciprocity, network reciprocity, and group selection all work by exploiting the locality of interaction [85]. When people interact in small groups in experimental public goods games (PGG), the ability to punish or reward peers (even at a cost) has been shown to promote cooperative behavior [42, 91]. Similarly, economic models of peer pressure show that pressure acts as a mechanism for improving the social welfare of the group [24] and is conducive of successful business partnerships [57]. In microfinance, stronger social ties can increase the likelihood of repayment of joint-liability loans, by facilitating monitoring and enforcement [61]. The effectiveness of peers on loan repayment has been demonstrated even in the absence of joint liability [21]. In relation to natural resource usage, strong community identification is instrumental to preventing the overuse of water resources [114], and so do other community-based incentives [115]. Top-down rules have been blamed for the degraded inshore ground fishery in Maine, in contrast to the Maine lobster fishery, which has been governed by informal community-based institutions and yielded much higher levels of compliance with sustainable resource use [39].

I present a joint model of externalities and peer-pressure and propose an alternative solution to the problem of global cooperation when externalities are global, but
interactions are local [76]. I also present results from two experiments that tested my proposed solutions for the control of externalities.

1.3 Outline

The outline of the rest of the document is as follows. In chapter 2, I present my results on the stability and dynamics of exclusive exchanges in networked societies and algorithms for computing exclusive stable bilateral exchanges. In chapter 3, I present my results on fairness and how they depend on network structure. In chapter 4, I propose a design of public policies that use peer-pressure for the control of externalities in networked societies and provide theoretical proof and experimental evidence that they are more efficient than existing policies. Finally, in chapter 5, I give the conclusions and directions for future research.
Chapter 2

Generalized Stable Matching in Bipartite Networks

2.1 Introduction

The weighted matching problem in bipartite networks and two-sided markets is an important problem with wide applications in assignment games [69], family economics [12] [13], job matching [35],[62], resource allocation [16], matching internet advertisements to users [40] and several others. A weighted matching in a bipartite network is comprised of (i) a matching of agents from one side of the network to the other side and (ii) the splits of the weights of matched edges between the matched agents. The payoff of a matched agent is the split given to the agent. The problem is often studied in two forms, the distributed form and the centralized form. In the distributed form of the problem, the studied concept is stable matching or core \(^1\). A stable matching is a weighted matching with payoffs of the agents satisfying the following conditions: (i) all agents have weakly higher payoffs than their reserved payoffs, and (ii) if any two adjacent agents in the network split the edge between them with one of them deriving strictly higher payoff than in the weighted matching then the other agent derives strictly lower payoff than in the weighted matching. In the centralized form of the problem, the studied concept is the optimal matching. An optimal matching

\(^1\)Shapley and Shubik point out the equivalence of the stable matching and core in [69]
is the matching of agents from one side of the network to the other side, that maximizes the sum of the weights of the matched edges. The integer linear programming formulation of the weighted matching problem and its integer relaxation has provided elegant results about the duality between the optimality and stability. The linear programming reduction has also provided important insights into the structure of the set of stable matchings [98]. The Hungarian algorithm [67] is a classic primal-dual algorithm that finds the maximum weight matching as well as the stable matching. With discretization of possible splits of the edges, the Hungarian algorithm also motivates natural myopic dynamics [16], [35] similar to the deferred acceptance algorithm used in the marriage problem [46]. It has been critical for designing incentive compatible, non-manipulable and optimal mechanisms for assignment problems and motivates the multi-unit auctions in [38].

The weighted matching problem assumes costless transfer of utilities between any two matched agents or costless redistribution of the weights on the matched edges between the end agents. In most real situations this assumption is not true. In the general case when the costless transfer of utilities is not assumed, the utilities of the agents depends both on its match and the split assigned to the agent and is continuous and strictly increasing in the split. We refer to this problem as the Generalized Weighted Matching problem. In this case the problem is a integer non-linear program and there is no strong duality between the optimality and stability. Most tools developed for the original weighted matching problem including the Hungarian algorithm, do not work for the generalized weighted matching problem and they cannot be trivially generalized. Through this chapter, we want to develop tools to study the Generalized Weighted Matching problem. In particular, we introduce an algorithm that converges to the generalized stable matching. The algorithm generalizes the Hungarian algorithm [67] for bipartite matching and hence, we call it the Generalized Hungarian Algorithm. We also introduce the concept of approximately stable matching and a natural dynamics that converge to approximately stable matchings. Our concept of approximate stability is very natural and satisfies important continuity properties. In the limit, we show that approximately stable matchings converge
to stable matchings. Our dynamics are different from the deferred acceptance algorithm. The deferred acceptance algorithm is made for the purpose of allocation by a central market authority. On the other hand, our dynamics simulates distributed asynchronous offers by agents from both sides of the network. We show that under very realized conditions, the dynamics converge almost surely to approximately stable matchings.

2.1.1 Application Scenarios

The lack of tools to study the generalized weighted matching problem has restricted the accurate modeling of many bipartite matching scenarios. We give two examples here, one in family economics and one in matching in markets and networks with indivisibilities like job markets. In family economics, the classical work of Becker [12], [13] provides a unitary model that assumes that families make decisions as a unit. In this model, the external factors such as market changes only result in the costless transfer of utilities among the couples without affecting the net welfare of the household due to Becker’s altruist explanation [13]. Empirical studies have falsified the unitary model and the assumption of costless transfer of utilities [22], [74]. As a response, several other models have been proposed about intra-household allocations including the bargaining models [77], [73] and the collective model [19]. While the bargaining models have some underlying assumptions of fairness, the collective model is the more natural model that only assumes pareto-efficiency in allocations that is empirically verified [28]. While the collective model has been studied for intra-household allocations in committed marriages and the effects of external factors such as market changes, there has been no work on finding stable marriages and the household allocations under the collective model assumption as was done with the unitary model assumption [46]. The Generalized Hungarian Algorithm can find the stable marriages in a heterogeneous society under the assumption of the collective model of household behavior. Our results can also help the investigation of effects of external factors such as labor market and taxation on household decisions such as expenses, fertility, childcare, divorce and the choice of partners.
In the markets with indivisibilities the classical assignment game [69] motivates, job matching [62] and exchanges in buyer-seller networks [66]. The problems assume the quasilinear utility of money. The quasilinearity assumption guarantees that the utilities are transferrable between any pair of matched pairs. Since we know that the budget effects on indirect utilities is non-linear and different for different individuals, the quasilinearity assumption is not justified. In the absence of this assumption, the problem becomes the generalized weighted matching problem. The generalized Hungarian algorithm can find the prices of goods in the core as well as help understand the comparative statics as the availability of goods and the structure of the network changes.

2.1.2 Related Work

The generalized weighted matching problem has been studied as the generalization of the assignment problem in [69]. In the case of the assignment problem, the Walrasian equilibria coincide with the core of the matching or the generalized stable matching. However, the stable matching or core is more general than Walrasian equilibria. In [59], [102], the authors study the existence of the Walrasian equilibria for the generalized assignment game. They show that the generalized assignment game is a balanced game and following Scarf’s result [97] that all balanced games have non-empty core, the assignment game has a non-empty core. The existence of a generalized stable matching has also been shown in [35]. The authors show that when the utilities are continuous, then the infinite game can be approximated by finite games all of which have a non-empty core. In the limit they show that the generalized matching problem also must have a non-empty core. However, none of these studies give a method to find the generalized stable matching. The only construction of the generalized stable matching is in [60] who show the existence of the generalized stable matching through the Kakutani’s fixed point theorem and hence use numerical methods to compute the fixed point for special cases. In [35] and [62], the authors give salary-adjustment processes based on deferred-acceptance algorithm to find the stable matching when
the edges in the network can be split in finitely many ways.\footnote{Under the assumption of separability of the workers in the firms production function, the job matching problem in [62] is equivalent to the generalized weighted matching problem.} The authors in [62] give an important comparative statics result that says that the worst stable matching for one side of the network weakly improves for that side when one agent is added to the other side and the best stable matching for one side of the network weakly deteriorates for that side when one agent is removed from the other side. In [37], the authors study the lattice structure of the set of generalized stable matchings or Walrasian equilibria. They study the minimum price Walrasian equilibria as a mechanism for allocating goods to buyers and show that the minimum price Walrasian equilibrium is buyer optimal and is also non-manipulable by the buyers. Recent work in [4] has shown similar results and provided a construction of such Walrasian equilibrium. Although we do not use these results in this chapter the results in these works are important tools for the study of the generalized stable matching. We provide different characterizations and our own constructive proof of the existence of the generalized stable matching using combinatorial optimization techniques. There is some independent work on generalized weighted matching again in the context of the generalized assignment game that uses graph theoretic techniques like in this chapter. In [25], [26] the authors study the comparative statics of allocations with external market effects. They study graph allocation structures that comprise of an assignment and a directed acyclic graph of goods with edges between goods that belong to the demand set of the same buyer. The edges are directed towards the good that is assigned to the buyer. They study local comparative statics that does not change the graph allocation structures but only the prices and global comparative statics in which larger changes in external market factors or outside options can change the graph allocation structures along with the prices of the goods. They give an important insight into the comparative statics that there can be at most five types of atomic market transitions or changes in the graph allocation structures in response to small enough changes in the external factors. Our algorithm generalizes the Hungarian algorithm and provides important insights into the comparative statics
for special network structures. Our dynamics are the first stochastic and distributed
dynamics studies in the context of matching problems.

2.1.3 Our Contributions

The main contributions in this chapter are the generalization of the Hungarian al-
gorithm to compute the generalized stable matching in bipartite networks and the
natural dynamics that converge to approximately stable matchings. This algorithm
is also the first constructive proof of the existence of the generalized stable match-
ing and builds upon techniques from combinatorial optimization. There are three
important results related to the algorithm.

(1) First, the convergence and correctness of the algorithm depends upon the
update rule for the Generalized Hungarian Algorithm. At every iteration of the
algorithm the matching is updated. An important monotonicity condition that our
algorithm satisfies is that the matching size is weakly increasing. The original Hungar-
ian algorithm satisfies a stronger monotonicity condition that the matching topology
does not change unless the matching size strictly increases. In the general case, the
matching topology may not be preserved but the monotonicity of the matching size
is sufficient for the convergence of the algorithm.

(2) Second, we characterize the generalized stable matching that is optimal for one
side of the network in a special class of graphs. Graphs in this class have a matching
that covers all but one agent such that all agents in the graph can be reached from
the unmatched agent along paths with alternating edges within the matching and
outside the matching. We show that as the reserved payoff of the unmatched agent
increases, the set of payoff vectors for such generalized stable matchings trace a strictly
monotonic path in the Euclidean space.

(3) Third, the algorithm always converges to the generalized stable matching that
gives the highest payoff to one side of the network among the set of all possible
generalized stable matchings. This result is important for designing mechanisms for
allocation of goods in markets and networks.
2.1.4 Outline

The organization of the rest of the chapter is as follows. In section 2.2, we introduce the model. In section 2.3, we show that the assignment game is a special case of our model and show how the Hungarian algorithm cannot be trivially generalized to our model. In sections 2.4 and 2.5, we introduce important concepts from graph theory and develop some important characterization results. In section 2.6, we present our algorithm for computing the generalized stable matching. In section 2.7, we introduce our main results about the correctness and convergence of the algorithm and the optimality of the point of convergence. In section 2.8 we study the approximately stable matching in natural dynamics that converge to approximately stable matchings. In section 2.9, we give our conclusions and directions of future work. The proofs of the results related to the algorithm are given in sections 2.10, 2.11, and 2.12.

2.2 Model

In this section we formulate the problem and introduce necessary terminology.

2.2.1 Network and Payoffs

Assume $A$ and $B$ are two finite and mutually exclusive sets of agents and $X = A \cup B$. A bipartite network between $A$ and $B$ is a graph $S = (X, E \subseteq A \times B)$, whose agents are agents in $X$ and whose edges connect agents from $A$ to agents in $B$. Given a bipartite network $S$, we will refer to the set of agents as $X^S$, the sets of agents in $A$ and $B$ as $A^S$ and $B^S$ respectively, and the set of edges as $E^S$ when necessary. When the agent set $X$ is understood, we will refer to the network by the edge set $E$.

Without loss of generality, we will assume that the graph $S$ is connected.

A set of agents $X' \subseteq X$ induces a subgraph $S_{|X'} = (X', E_{|X'})$ of $S$, such that $E_{|X'} = \{(i, j) \in E : i \in X', j \in X'\}$.

The set of neighbors of an agent $i \in X$ is $Nbr^S(i) = \{j \in X : (i, j) \in E\}$. When the context is well understood, we will also refer to $Nbr^S(i)$ as $Nbr^E(i)$ or just $Nbr(i)$.
The set of neighboring agents of \( x \subseteq X \) is \( \text{Nbr}^S(x) = \cup_{i \in x} \text{Nbr}^S(i) \).

A **weight** function \( w : E \rightarrow \mathbb{R}_+^+ \) assigns a weight to each edge. An edge \((i, j)\) with \( i \in A \) and \( j \in B \) has a weight \( w(i, j) \) that can be split between \( i \) and \( j \).

A **split** on the edge \((i, j)\) is the pair \((s_i, s_j)\), with \( s_i + s_j = w(i, j) \).

The agents derive payoffs from the part of the split given to them. For a split \((s_i, s_j)\), the payoff of the agent \( i \) is \( u_i(j, s_i) \) and the payoff of the agent \( j \) is \( u_j(i, s_j) \). The payoff of a agent depends upon both the part of the split given to the agent and the edge on which the split is made. Thus for each edge the payoff of a agent is a unique function of the part of the split given to the person. We assume that the payoff function \( u_j(i, \cdot) \) is *strictly increasing and bicontinuous* over \((0, w(i, j))\) in the split for each pair \((i, j)\). Unless specified otherwise, we will assume that \( \lim_{x \to 0} u_i(j, x) = -\infty \) and \( \lim_{x \to w(i, j)} u_i(j, w(i, j)) = \infty \) for all \( i \in X \) and all \((i, j)\) in \( E \).

When a split is made on \((i, j)\) such that the payoff of agent \( i \) is \( x \), the corresponding payoff of agent \( j \) is \( v_j(i, x) = u_j(i, w(i, j) - u_i^{-1}(j, x)) \). Thus the function \( v_j(i, x) \) is defined for all \( x \in \mathbb{R} \) and is strictly decreasing and bicontinuous for each pair \((i, j)\).

We will refer to these functions as **pareto payoff functions**.

Each agent \( i \in X \) gets a **reserved payoff** \( r_i \), if \( i \) does not split any edge with its neighbors. The reserved payoff vector will be represented as \( r \in \mathbb{R}_+^X \), where the elements of the vector are the reserved payoffs of the agents.

### 2.2.2 Matching

Given a matching \( M \subseteq E \), we will refer to the match for any agent \( i \) as \( M(i) \). A split for the matching \( M \) is a function \( s^M : A \cup B \rightarrow \mathbb{R}_+^+ \) such that \( s^M(i) + s^M(j) = w(i, j) \) for all \((i, j)\) in \( M \), and \( s^M(i) = 0 \), if \( i \) is unmatched. A **weighted matching** is a tuple \((M, s^M)\) of matching and a split for the matching. The **payoff profile** \( U(M, s^M) \) for a weighted matching \((M, s^M)\) is a vector where each element is the payoff of a agent for the given weighted matching. The payoff of a agent \( i \) for the weighted matching

---

3The assumption that the range of the utilities is the whole \( \mathbb{R} \) is not necessary for the algorithm, as we will see in the examples in Appendix 2.17, but simplifies the proofs in the chapter.

4The graph of the function corresponds to the pareto frontier of the utility possibility set for the pair of agents.
(M, s^M) is \( U_i (M, s^M) = u_i (M(i), s^M(i)) \) if \( i \) is matched and \( U_i (M, s^M) = r_i \) if \( i \) is not matched.

A weighted matching is **stable** if all agents have payoffs weakly higher than their reserved payoffs and for all neighboring agents, \((i, j) \in E\), all splits \((s_i, s_j)\) on \((i, j)\) that give a strictly higher payoff to one agent, give strictly worse payoff to the other agent.\(^5\) i.e. \( u_i (j, s_i) > U_i (M, s^M) \) \( \implies \) \( u_j (i, s_j) < U_j (M, s^M) \). Alternatively, a weighted matching is **stable** if all agents have payoffs weakly higher than their reserved payoffs and for all neighboring agents, \((i, j) \in E\), the pareto payoff for the agent, \( i \) given the the payoff of the agent \( j \) for the weighted matching, is weakly less than the payoff of \( i \) for the weighted matching. i.e. \( v_i (j, U_j (M, s^M)) \leq U_i (M, s^M) \).

A special case in this model is the standard assignment game. We will see how the Hungarian algorithm used to find stable matching in the special case fails for the general case.

### 2.3 A Special Case: The Standard Weighted Matching or the Assignment Game

When for all agents \( i \in X \), the payoffs \( u_i (j, s_i) \) is linear in \( s_i \), and does not depend upon the match, \( j \), then the problem reduces to the well studied weighted matching in bipartite networks [98]. For simplicity assume \( u_i (j, s_i) = s_i \) for all \( i \in X \). Then a generalized stable matching is a matching \( M^o \) and a split vector \( s \), such that

\[
\begin{align*}
    s_i + s_j &= w(i, j), \text{ for all } (i, j) \in M^o & (2.1) \\
    s_i &= r_i, \text{ for all } i \text{ unmatched } & (2.2) \\
    s_i + s_j &\geq w(i, j), \text{ for all } (i, j) \in E & (2.3) \\
    s_i &\geq r_i, \text{ for all } i \in X & (2.4)
\end{align*}
\]

\(^5\)For the bipartite matching problem, the core coincides with the set of stable matchings because the absence of profitable bilateral deviations implies absence of profitable multilateral deviations.
A weighted matching is a solution to the above set of inequalities if and only if the splits is a solution to the minimization problem and the matching is the solution to the maximization problem below.

\[
\min \sum_{i \in A \cup B} (s_i - r_i) \quad \quad \max \sum_{(i,j) \in E} (w(i,j) - (r_i + r_j)) \chi(i,j)
\]

such that \(s_i + s_j \geq w(i,j)\), for all \((i,j) \in E\) such that \(\sum_{(i,j) \in E} \chi(i,j) \leq 1\), for all \(i \in A \cup B\)

\[
s_i \geq r_i \quad \quad \chi(i,j) \geq 0, \text{ for all } (i,j) \in E
\]

The corners of the polytope resulting from the constraints in the maximization problem are integral due to the total unimodularity property and therefore the integer constraints have been relaxed. Hence, the maximum is attained by a matching defined by the characteristic function \(\chi^o\) that is the optimal solution to the maximization program. The minimization problem is the dual linear program of the maximization problem. the maximization problem is called the maximum weight matching problem. In a finite graph with finite weights, the optimal solution exists and the optimal value is finite. From the strong duality, the minimization problem has a solution with the minimum value same as the maximum weight over all the matchings. At the optimal solution for the minimization problem, for any edge \((i,j) \in M^o\), \(s_i + s_j = w(i,j)\), where the matching \(M^o\) is defined by the maximizing characteristic function \(\chi^o\), and \(s_i\) and \(s_j\) is the minimizing split. The matching \(M^o\) and the split vector \(s\) is a solution to the system of inequalities 2.1.

The Hungarian algorithm simultaneously finds the maximum weight matching and the minimizing split and hence gives a stable matching. Besides providing a constructive proof of existence of a generalized stable matching in the general case, our algorithm also extends the Hungarian algorithm for bipartite matching for the general case. The extension of the Hungarian algorithm to find the general stable matching is non trivial. We now explain why the Hungarian method will not work for the general case.
2.3.1 Summary of the Hungarian algorithm

We first give a summary of the Hungarian algorithm. An example case is presented in appendix 2.17 that shows the working and limitations of the algorithm. The Hungarian algorithm takes a weighted bipartite graph $S$ with edges $E$ and gives a maximum weight matching $M^\circ$ and a split vector $s$. The matching $M^\circ$ and the split vector $s$ together is a generalized stable matching. For simplicity we assume that there exists a perfect matching in the graph $S$; otherwise dummy agents can be inserted such that the resulting graph has a perfect matching. Without loss of generality, we also assume that all reserved payoffs are 0; a translation of the weights of the edges can reduce the problem to this case when the reserved payoffs are not 0. The Hungarian algorithm is a primal-dual algorithm.

During the algorithm, we maintain a split vector $s$ that is a potential or satisfies the condition $s_i + s_j \geq w(i,j)$, for all $(i,j) \in E$. We refer to the subgraph where the inequality is tight as $E_s = \{(i,j) \in E : s_i + s_j = w(i,j)\}$. We also maintain a matching $M$ in the $E_s$ throughout the algorithm. An orientation of the graph (denoted by $\vec{E}_s$) is obtained by orienting all the edges in $M$ from $B$ to $A$ and all edges outside the matching from $A$ to $B$. Initially, $s = 0$ and $M$ is empty. In each step, either we modify $s$, or modify the orientation to obtain a matching with more edges. We are done if $M$ is a perfect matching.

Changing the matching

In a general step, let $A_\phi \subseteq A$ and $B_\phi \subseteq B$ be the vertices not covered by $M$. Let $Z$ be the set of vertices reachable from $A_\phi$ by a directed path in $\vec{E}_s$. If $B_\phi \cap Z$ is nonempty, then reverse the orientation of a directed path in $\vec{E}_s$ from $A_\phi$ to $B_\phi$. Thus the size of the corresponding matching increases by 1.

Changing the potential

If $B_\phi \cap Z$ is empty, then let $\Delta := \min\{s(i) + s(j) - w(i,j) : i \in Z \cap A, j \in B \setminus Z\}$. $\Delta$ is positive because there are no tight edges between $Z \cap A$ and $B \setminus Z$. Decrease
s by $\Delta$ on the vertices of $Z \cap A$ and increase $s$ by $\Delta$ on the vertices of $Z \cap B$. The resulting $s$ is still a potential. $E_s$ changes, but it still contains $M$. We orient the new edges from $A$ to $B$. By the definition of $\Delta$ the set $Z$ of vertices reachable from $A_\phi$ increases.

We repeat these steps until $M$ is a perfect matching, in which case it gives a maximum weight matching.

Problems with general case

We now show why the Hungarian algorithm does not work for the general case. Please note that a critical thing in the Hungarian algorithm is to maintain a potential throughout the algorithm. When the splits are changed in the iteration, the resulting split is still a potential. A constant decrease in $s$ by $\Delta$ on the vertices of $Z \cap A$ and increase in $s$ by $\Delta$ on the vertices of $Z \cap B$ gives us a new split that is also a potential.

In the general case a potential is a split that satisfies the condition

$$u_i^{-1}(j, o_i) + u_j^{-1}(i, o_j) \geq w(i, j), \text{ for all } (i, j) \in E$$

, where $o_i = u_i(j, s_i), o_j = u_j(i, s_j), \text{ for all } (i, j) \in M^o$. When $u_i$ is nonlinear then a simple update rule as in the linear case will not result into a potential. The edges that were tight in $E_s$ may not satisfy the condition for $s$ to be a potential. It is also not easy to say that an update rule exists that will give a new potential different from the old potential while keeping the matching fixed. We demonstrate this through an example in the appendix. Hence, the Hungarian algorithm cannot be trivially extended to the general case.

To go further, we will need to define some more terminology. It is simpler to work with utilities instead of the splits. So we define some additional terminology and define the problem with utilities instead of splits.
2.4 Offers

An offer profile is a vector $O \in \mathbb{R}^X$ where the element $O_i$ is agent $i$’s offer. We will denote the restriction of an offer profile $O$ to a set of agents $Y$ as $O|_Y$. An offer profile $O$ is feasible, if there exists a weighted matching $(M, s^M)$, with payoff profile $U(M, s^M) = O$. An offer profile $O$ is individually rational if $O \geq r$. An offer profile $O$ is stable if for all neighboring agents, $(i, j) \in E$, agent $i$’s offer is weakly higher than the pareto payoff of $i$ given $j$’s payoff is equal to $j$’s offer, i.e. $O_i \geq v_i(j, O_j)$. Given an offer profile $O$, the equality subgraph $EQ(O)$ is the subset of edges $E$ for which the stability constraint is critical, i.e. $EQ(O) = \{(i, j) \in E : O_j = v_j(i, O_i)\}$. We will refer to the neighbors of a agent $i$ in the equality subgraph $EQ(O)$ as $Nbr^{EQ(O)}(i)$.

We can now reduce the problem in terms of offer profiles. The reduced problem is to find a matching $M^o$ and an offer profile $O^o$ such that

$$O^o_i = r_i, \text{ for all } i \text{ unmatched} \quad (2.5)$$
$$O^o_j = v_j(i, O^o_i), \text{ for all } (i, j) \in M^o \quad (2.6)$$
$$O^o_j \geq v_j(i, O^o_i), \text{ for all } (i, j) \in E \quad (2.7)$$
$$O^o_i \geq r_i, \text{ for all } i \in A \cup B \quad (2.8)$$

The first two sets of constraints are the feasibility constraints, the third set of constraints are the stability constraints and the fourth set of constraints are the individual rationality constraints. Once such an offer profile is identified, the weighted matching can be constructed as follows.

From Feasible Offer Profile to Weighted Matching

We first give a feasibility test for the offer profile in the following proposition.

Proposition 1. An offer profile $O^*$ is feasible for the bipartite network $S$ with reserved payoff profile $r$ if and only if the equality subgraph $EQ(O^*)$ has a matching $M^*$ in which every agent $i \in X$ with offer $O^*_i \neq r_i$ is matched.
The proof of the proposition is simple but is provided in appendix 2.13 for completeness. Given a feasible offer profile, \( O^* \) we can construct a matching \( (M^*, s^{M^*}) \), with payoff profile \( U(M^*, s^{M^*}) = O^* \). The matching \( M^* \) can be constructed from the equality subgraph \( EQ(O^*) \) using the augmenting path algorithm [98]. Since only the agents with offers different from their reserved payoffs need to be matched in \( M^* \), therefore the augmenting path algorithm can be stopped when all such agents are matched. Once, the matching \( M^* \) is constructed, the split is constructed as such:

for all \( (i, j) \in M^* \), set \( s^{M^*}(i) = u^{-1}_i(j, O^*_i) \), and \( s^{M^*}(j) = u^{-1}_j(i, O^*_j) \),

for all unmatched \( i \in X \), \( s^{M^*}(i) = 0 \)

If the offer profile is feasible, individually rational and stable then the constructed weighted matching satisfies the stability constraints and therefore is a generalized stable matching.

Comparing Offer profiles

An offer profile \( O^1 \succeq_A O^2 \) or \( O^1 \) is \textbf{A-preferred} over \( O^2 \), if \( O^1_i \geq O^2_i \) for all \( i \in A \). Given a set of offer profiles, \( \emptyset \), an offer profile \( O^* \in \emptyset \) is the \textbf{A-maximum}(A-minimum) offer profile in \( \emptyset \) if \( O^* \succeq_A O(O^* \preceq_A O) \) for all \( O \in \emptyset \). B-preferred relation and B-maximum(B-minimum) offer profiles are defined similarly. Strictly preferred relations, \( \succ_A \) and \( \succ_B \) are defined similarly. We can compare the offers on a agent in two different stable offer profiles using the critical constraints, when we know the comparison between a neighboring agent’s offers. These comparisons are shown in proposition 2. The proof is simple but we provide it in Appendix A for completeness.

**Proposition 2.** Assume there exists two stable offer profiles \( O^1 \) and \( O^2 \). If \( i \in X \) with \( O^1_i \leq O^2_i \) and \( j \in Nbr^{EQ(O^2)}(i) \) \( EQ(O^2) \), then \( O^1_j \geq O^2_j \). The resulting inequalities are strict when the conditioning inequalities are strict.
2.5 Some Graph Theoretic Terminology and Results

In this section, we introduce some terminology that will be heavily used for the construction of a stable and feasible offer profile. Other standard terminology from graph theory are provided in the appendix 2.14.2. An alternating tree \( T^M \) for a matching \( M \) is an **alternating spanning tree** if it spans all the agents in the network, i.e.-\( X^{T^M} = X \). A bipartite network \( S \) is an **alternating spanning tree admitting network (ASTAN)** if it has a near-perfect matching \( M^o \) and an associated alternating spanning tree \( T^{M^o} \). ASTANs have important structural properties. Without loss of generality, we assume that the unmatched agent is in \( A \). The following lemma provides a test for ASTANs. The proof comes from the Hall’s theorem [55] and is given in appendix 2.16 for completeness.

**Lemma 1.** A bipartite network \( S = (A \cup B, E) \) is an ASTAN if and only if \(|A| - |B| = 1 \) and:

- Every \( B' \subseteq B \) has least \(|B'| + 1 \) neighboring agents in \( A \).
- Every \( A' \subseteq A \) has at least \(|A'| \) neighboring agents in \( B \).

An offer profile \( O \) is a **stable alternating spanning tree generating offer profile (SASTGOP)** if \( O \) is stable and \( EQ(O) \) is an ASTAN. We point that SASTGOPs exist only for ASTANs and also induce an equality subgraph that is an ASTAN.

2.6 The Generalized Hungarian Algorithm

We now present the generalized Hungarian algorithm. The algorithm finds the \( A \)-maximum stable matching given the bipartite network \( S \), the pareto-payoff functions \( v_i(j,.) \) for each pair of neighboring agents \( i, j \), and the reserved payoff vector \( r \). The initialization step is similar to the original Hungarian algorithm. In the general iteration of the algorithm, the offers are updated on the agents in one alternating tree instead of the whole Hungarian forest, as in the original Hungarian algorithm, and the matching is subsequently changed in the resulting equality subgraph. The initialization and the iterations are described as follows.
2.6.1 Initialization

We first define an initial offer profile $O^0$ and the initial matching $M^0$ as follows.

For $t=0$, set

Initializing the Offer Profile

1. For all $j \in B$, set $O^0_j = r_j$ and for all $i \in A$, set $O^0_i = \max_{j \in \text{Nbr}(i)} v_i(j, r_j)$.

2. Set the initial equality subgraph $EQ^0 = EQ(O^0)$.

Initializing the Matching

1. Set $M^0 = M^*(EQ^0)$ some maximum matching in $EQ^0$. The maximum matching can be computed using the augmenting path algorithm.

2. If there is an alternating path from an unmatched agent $i \in A$ with $O^0_i > r_i$ to a matched agent $i^\circ \in A$ with $O^0_{i^\circ} = r_{i^\circ}$, then switch the alternating edges from within the matching $M^0$ to outside the matching $M^0$ and vice-versa along the alternating path from $i$ to $i^\circ$. This leaves the matching size unchanged. Repeat this process until no such alternating paths are present in $EQ^0$.

Initial Hungarian Forest

Set $I^0 = \{i^0_1, i^0_2, ..., i^0_m\}$ as the set of unmatched agents in $A$ with offers strictly higher than agents reserved payoffs. Set $m^0 = |I^0|$, the size of $I^0$, and for all $i^0_k \in I^0$, set $T^0_{i^0_k} = T_{i^0_k}$ as the alternating tree induced by the matching $M^0$ rooted at $i^0_k$.

2.6.2 Iteration

We iteratively change the offer profile to create a sequence of offer profiles. At each time $t \geq 0$, we compute the new offer profile $O^{t+1}$ and change the matching as follows.

While $I^t \neq \emptyset$, pick the first agent, $i^t = i^t_1$ in $I^t$, and the alternating tree $T^t = T_{i^t}$ rooted at $i^t$. 
Changing the Offers

We only change the offers on the agents, $X^{T^*}$, in the chosen alternating tree, $T^*$.

1. For all $i \in A^{T^*}$, set the expanding offer $eo_i = \max\{\max_{j \in Nbr(i) \setminus B^{T^*}} v_i(j, O_j^t), r_i\}$

2. For the subgraph $S_{|X^{T^*}}$, find the desired SASTGOP $O^s$ such that $O^s_i \geq eo_i$ for all $i \in A^{T^*}$ with equality for at least one $i \in A^{T^*}$.

For all $i \in X^{T^*}$ set $O^t_{i+1} = O^s_i$ and for all $i \notin X^{T^*}$ set $O^t_{i+1} = O^t_i$.

3. Set $EQ^{t+1} = EQ(O^{t+1})$.

Changing the Matching

The matching in the new equality subgraph is found by first finding the near perfect matching among the agents $X^{T^*}$ and then refining the matching by removing the augmenting paths as in the original Hungarian algorithm.

1. Pick a near perfect matching $M^*_{|X^{T^*}} \subseteq EQ^{t+1}_{|X^{T^*}}$, within the subgraph of the equality subgraph, induced by the agents in the selected alternating tree, such that $i^t$ is the only unmatched agent in $X^{T^*}$. First set $M^{t+1} = M^t_{|X \setminus X^{T^*}} \cup M^*_{|X^{T^*}}$.

2. Refine the matching by removing the augmenting paths as follows. If there exists an augmenting path from $i^t$ with respect to the matching $M^{t+1}$ within the equality subgraph $EQ^{t+1}$, then switch the edges within the matching and outside the matching along the augmenting path to create a new matching $M^{t+1}$. If there is no augmenting path but there is an alternating path from $i^t$ to some matched agent $i^o \in A$ with $O^t_{i^o} = r_{i^o}$, then switch the alternating edges from within the matching $M^{t+1}$ and outside the matching $M^{t+1}$ along the alternating path from $i^t$ to $i^o$.

Updated Hungarian Forest

If $i^t$ is matched in $M^{t+1}$, or $O^t_{i^t} = r_{i^t}$ then set $m^{t+1} = m^t - 1$, for all $k < m^t$, set $i_{k+1}^{t+1} = i^t_k$, $I^{t+1} = \{i_1^{t+1}, \ldots, i_{m^{t+1}}^{t+1}\}$. Otherwise set $m^{t+1} = m^t$, for all $k \leq m^t$, set
$i_{k}^{t+1} = i_{k}^{t}$, $I^{t+1} = I^{t}$. For all $i_{k}^{t+1} \in I^{t+1}$, $T_{i_{k}^{t+1}}^{t+1}$ is the new alternating tree rooted at $i_{k}^{t+1}$.

### 2.6.3 Example

We show how the algorithm proceeds in an example in appendix 2.17.

### 2.7 Results Related to the Algorithm

In this section we present our results related to the algorithm. The results can be categorized into three groups. The first group of results establishes that the steps of the algorithm are well defined. In particular, we show that offer update is well defined and the required SASTGOP can be computed for any alternating tree. These results also identify several important properties of ASTANs, SASTGOPs. The second group of results establishes that the algorithm converges in finite time to a stable, feasible and individually rational offer profile. The third group of result establishes that the algorithm finds the A-maximum stable matching. We also present an important optimality characterization of SASTGOPs that suggests that the desired SASTGOPs for offer updates in the Generalized Hungarian algorithm can be computed using the same algorithm.

### 2.7.1 Correctness of Offer Updates

We now show that the desired SASTGOP for the offer updates is well defined. We provide a characterization of the set of SASTGOPs for ASTANs. \(^6\) The proof is give in section 2.10.

**Theorem 1.** *If $S$ is an ASTAN then the set of SASTGOPs for $S$ is non-empty and is a path in $R^{|A \cup B|}$. Define $f^{S} : (0, 1) \rightarrow R^{|A \cup B|}$ to be the corresponding continuous*

\(^6\)The characterization of the set of SASTGOPs for ASTANs is of independent interest. It provides a a stronger characterization of the set of SASTGOPs in ASTANs than the characterization of the set of stable feasible and individually rational offer profiles for bipartite graphs. From [37], we know that the set of stable feasible and individually rational offer profiles is a lattice. The theorem says that the set of SASTGOPs for any ASTAN is a strictly monotonic path in the Euclidean space.
path function and $\pi_i$ to be the projection map that maps offer profiles in $\mathbb{R}^{|A \cup B|}$ to the offers on the agent $i \in A \cup B$, then:

- $\pi_i \circ f^S$ is a bicontinuous function for all $i \in A \cup B$.
- $\lim_{x \to 0} |\pi_i \circ f^S(x)| = \lim_{x \to 1} |\pi_i \circ f^S(x)| = \infty$ for all $i \in A \cup B$
- Either $\pi_i \circ f^S$ is strictly increasing for all $i \in A$ and strictly decreasing for all $i \in B$ or $\pi_i \circ f^S$ is strictly decreasing for all $i \in A$ and strictly increasing for all $i \in B$. The either/or condition is because both $f^S$ and $-f^S$ are such path functions.

The immediate consequence of the theorem is that the offer update rule in the algorithm is well defined. To show this, consider the alternating tree $T^t$ at any iteration $t$ of the algorithm. Assume the expanding offers on agent $i \in A^T$ is $eo_i$. We first observe that the subgraph $S_{|X^Tt}$ is a ASTAN and hence theorem 1 applies to the subgraph. Then the desired SASTGOP for $S_{|X^Tt}$ is $O^S_{|X^Tt} = f^S_{|X^Tt}\left(\max_{i \in A^T}\left\{\left(\pi_i \circ f^S_{|X^Tt}\right)^{-1}(eo_i)\right\}\right)$, where $f^S_{|X^Tt}$ is the desired path function for the subgraph $S_{|X^Tt}$ as defined in the theorem. Another important consequence of the theorem is that it establishes the equivalence between the desired SASTGOP for offer updates in the algorithm and the A-minimum of SASTGOPS that are individually rational for agents in $A^T$ when their reserved payoffs are equal to their expanding offers. The following corollary establishes the equivalence. The proof is given in section 2.10.

**Corollary 1.** If $S$ is an ASTAN and $r \in \mathbb{R}^{|A \cup B|}$, then the SASTGOP, $O^S$, such that $O^S_i \geq r_i$ for all $i \in A$, with equality for at least one $i \in A$ is the A-minimum of all SASTGOPS $\odot = \{O \in \mathbb{R}^{|A \cup B|} : O_i \geq r_i, \text{ for all } i \in A\}$.

**Proof Concept for the Theorem**

We will give the proof of the theorem and the corollary in section 2.10. To prove the theorem, we need to show three things. First, we need to show that for any given agent and any given offer on the agent, there exists a SASTGOP that admits the given offer.
on the given agent. Second, we need to show that such a SASTGOP is unique. Third
we need to show that there is some monotonic relation between SASTGOPS or if
there are two different SASTGOPS $O^1$ and $O^2$, then either $O^1 \prec_A O^2$ and $O^1 \succ_B O^2$
or $O^1 \succ_A O^2$ and $O^1 \prec_B O^2$. Here we introduce three lemmas that establish the
three results. The proofs are given in section 2.10.

**Lemma 2** (SASTGOP Uniqueness). Assume $S$ is an ASTAN. Pick any $i \in A$ and
offer $o_i$. Assume there exists a SASTGOP $O$ with $O_i = o_i$. Then $O$ is the unique
SASTGOP with $O_i = o_i$.

**Lemma 3** (SASTGOP Monotonicity). Assume $S$ is an ASTAN. Pick any $i \in A$ and
assume that for all $o_i \leq c_i$ there exists a SASTGOP $O$ with $O_i = o_i$. Pick any two
SASTGOPS $O^1$ and $O^2$ with $O^1_i \leq c_i$ and $O^2_i \leq c_i$. If $O^1_i < O^2_i$, then for all $i' \in A$,
$O^1_i < O^2_i$ and for all $j' \in B$, $O^1_{j'} > O^2_{j'}$.

**Lemma 4** (SASTGOP Existence). Assume $S$ is an ASTAN. Then for all $i \in A$, and
$o_i \in \mathbb{R}$, there exists a SASTGOP $O^*$ with $O^*_i = o_i$.

A simple corollary of lemma 4 is that if the network is an ASTAN, then we can
also pick any agent, $i \in B$, and any offer, $o_i$, on the agent and find a SASTGOP, $O$,
such that $O_i = o_i$. The proofs are given in section 2.10.

**Corollary 2.** Assume $S$ is an ASTAN. Then for all $i \in B$, and $o_i \in \mathbb{R}$, there exists
a SASTGOP $O^*$ with $O^*_i = o_i$.

### 2.7.2 Convergence to A-maximum generalized stable matching

We now state the results about convergence of the algorithm to the A-maximum
generalized stable matching. The first result in this section states that the algorithm
converges in finitely many iterations and the offer profile at the point of convergence is
feasible, stable and individually rational. The second result states that the algorithm
finds a A-maximum stable, feasible and individually rational offer profile.
Convergence

We first give the convergence result.

**Theorem 2.** There exists a finite time $t^*$ for which $O^{t^*}$ is feasible, stable and individually rational.

The proof is given in section 2.11. The proof concept is as follows. To prove the theorem, we need two results. First, we need to show that, for some finite $t^*$, all agents with offers strictly higher than their reserved payoffs are matched by the matching $M^{t^*}$. This implies convergence because then $I^{t^*} = \phi$ and also proves that the offer profile $O^{t^*}$ is feasible since all agents with payoffs different than their reserved payoffs are matched. Second, we need to show that $O^{t^*}$ is stable and individually rational. We state two results in the following propositions which imply the stability and individual rationality of the offer profiles. The proofs are given in section 2.11.

**Proposition 3.** At any iteration $t \geq 0$ of the Generalized Hungarian algorithm, the offer profile $O^t$ is stable and individually rational.

**Proposition 4.** At any iteration $t \geq 0$ of the Generalized Hungarian algorithm, if there exists $j \in B$ with $O^t_j > r_j$, then $j$ is matched in $M^t$.

We then show that for some finite $t^*$, all agents in $A$ with offers strictly higher than their reserved payoffs are matched by the matching $M^{t^*}$.

A-optimality

We now give the optimality results.

**Theorem 3.** The offer profile $O^{t^*}$ given by the generalized Hungarian algorithm is the $A$-maximum and $B$-minimum stable, feasible and individually rational offer profile and the matching $M^{t^*}$ along with the split $s^{M^{t^*}}$ constructed from $O^{t^*}$ and $M^{t^*}$ together are a $A$-maximum and $B$-minimum generalized stable matching $(M^{t^*}, s^{M^{t^*}})$.

The proof of the theorem is given in section 2.12. The summary of the proof is as such. We show that throughout the algorithm, the offer profile is weakly $A$-preferred
to the A-maximum stable, feasible and individually rational offer profile. Since the final offer profile is stable, feasible and individually rational offer profile, then it must be a A-maximum stable, feasible and individually rational offer profile. To show that the algorithm maintains an offer profile that is weakly A-preferred to the A-maximum stable, feasible and individually rational offer profile, we use the following lemma that establishes the A-optimality of the desired SASTGOPS in the algorithm.

**Lemma 5.** If $S$ is an ASTAN then the A-minimum of all SASTGOPS that are individually rational for all agents in $A$ is the A-maximum stable and feasible offer profile for the reserved payoff profile $r$.

The proof of the lemma is given in section 2.12. The above lemma has two important consequences. First, it gives a short proof of the theorem. Second, It suggests that the desired SASTGOP for the offer updates in the generalized Hungarian algorithm can be computed using the generalized Hungarian algorithm. We elaborate the second consequence next.

### 2.7.3 Computing SASTGOP in the Generalized Hungarian Algorithm

We now discuss the computation of desired SASTGOP for offer updates in the Generalized Hungarian algorithm. We first give a corollary to lemma 5 establishing the equivalence of the A-minimum individually rational SASTGOP to the A-maximum stable, feasible and individually rational offer profile for ASTANs.

**Corollary 3.** If $S$ is an ASTAN and it admits an individually rational SASTGOP for a given reserved payoff profile, $r$, then the A-minimum individually rational SASTGOP is equal to the A-maximum stable, feasible and individually rational offer profile for the reserved payoff profile $r$.

The proof of the corollary is given in section 2.12. The desired SASTGOP for offer updates can be computed by the Generalized Hungarian algorithm if it is the A-maximum stable, feasible and individually rational offer profile. We first notice
that in the algorithm, the subgraph $S_{|X^T^t}$ induced by the agents $X^T^t$ in the selected alternating spanning tree $T^t$ at each iteration $t > 0$ is an ASTAN. Following corollary 1, the desired SASTGOP for $S_{|X^T^t}$ is the $\Delta$-minimum of all SASTGOPS that are individually rational for all $i \in A^T^t$ with reserved payoff $eo_i$. The ASTAN $S_{|X^T^t}$ already admits an individually rational SASTGOP $O_{|X^T^t}^t$ for the reserved offer profile $r^t$, where $r_i^t = eo_i$ for all $i \in A^T^t$ and $r_j^t = r_j$ for all $j \in B^T^t$. Therefore, following corollary 3, the desired SASTGOP is the $\Delta$-maximum stable and feasible and individually rational offer profile for the ASTAN $S_{|X^T^t}$ with reserved payoff $r^t$. Hence, the desired SASTGOP can be computed by running the Generalized Hungarian algorithm on $S_{|X^T^t}$ with reserved payoff $r^t$.

### 2.8 Approximately Stable Matchings and Convergent Dynamics

In this section, we introduce the concept of approximately stable matching and a natural dynamics that converges almost surely to approximately stable matching. An $\varepsilon$-stable matching is a matching such that no pair of neighboring agents in the network can simultaneously improve their payoffs by more than $\varepsilon$. If there is some cost to breaking the exchange relationship, then approximately stable matchings are a natural concept. We also show that as $\varepsilon \to 0$, the set of approximately stable matchings converge to the set of stable matchings. The convergence happens for both the matching as well as the splits. We provide a stochastic myopic dynamics, with some similarity to the salary-adjustment process [62] that converges to an $\varepsilon$-stable matching. We also show that for sufficiently small $\varepsilon$, all $\varepsilon$-stable matchings are topologically equivalent (have the same set of matched pairs) to some stable matching.

#### 2.8.1 Approximate Stability

Given $\varepsilon \geq 0$ a weighted matching $(M, s^M)$ is $\varepsilon$-stable if the following conditions hold:
1. All agents have payoffs weakly higher than their reserved payoffs.

2. The matching is maximal (i.e.- no neighboring pair of agents are unmatched).

3. If \((i,j) \in E\) and both \(i\) and \(j\) are matched, not necessarily to each other, then
\[
v_i \left( j, U_j \left( M, s^M \right) + \varepsilon \right) \leq U_i \left( M, s^M \right) + \varepsilon
\]

4. If \((i,j) \in E\) and \(i\) unmatched but \(j\) matched, then
\[
v_i \left( j, U_j \left( M, s^M \right) + \varepsilon \right) \leq r_i
\]

Given \(\varepsilon \geq 0\) a weighted matching \((M,s^M)\) is strictly \(\varepsilon\)-stable if the following conditions hold:

1. All agents have payoffs weakly higher than their reserved payoffs.

2. The matching is maximal (i.e.- no neighboring pair of agents are unmatched).

3. If \((i,j) \in E\) and both \(i\) and \(j\) are matched, not necessarily to each other, then
\[
v_i \left( j, U_j \left( M, s^M \right) + \varepsilon \right) < U_i \left( M, s^M \right) + \varepsilon
\]

4. If \((i,j) \in E\) and \(i\) unmatched but \(j\) matched, then
\[
v_i \left( j, U_j \left( M, s^M \right) + \varepsilon \right) < r_i
\]

We say that a matching \(M\) admits a \(\varepsilon\)-stable matching is there exists a split \(s^M\) such that \((M, s^M)\) is a \(\varepsilon\)-stable matching.

We now present the continuity properties of the \(\varepsilon\)-stable matchings. An important question is as \(\varepsilon \to 0\), does the set of \(\varepsilon\)-stable weighted matchings converge to the set of stable matchings. We need to show that both the matching and the payoffs converge to the set of stable weighted matchings. First, we define the correspondences that maps \(\varepsilon\) to the set of \(\varepsilon\)-stable matchings.

We define \(\psi\) as a correspondence from \(\mathbb{R}_+\) to the set of matchings.

\[
\psi (\varepsilon) = \{ M : \exists s^M \text{ s.t. } (M, s^M) \text{ is } \varepsilon\text{-stable} \}
\]

For each matching \(M\), we define \(\xi_M\) as the correspondence \(\xi_M : \mathbb{R}_+ \to \mathbb{R}_+^{\lvert A \rvert + \lvert B \rvert}\) from \(\mathbb{R}_+\) to the set of payoff profiles.

\[
\xi_M (\varepsilon) = \{ x \in \mathbb{R}_+^{\lvert A \rvert + \lvert B \rvert} : x = U (M, s^M) \text{ s.t. } (M, s) \text{ is } \varepsilon\text{-stable} \}
\]
2.8. APPROXIMATELY STABLE MATCHINGS AND CONVERGENT DYNAMICS

If the matching $M$ does not admit a $\varepsilon$-stable matching then clearly $\xi_M(\varepsilon)$ is empty. We also note that for any $\varepsilon_1 < \varepsilon_2$, $\psi(\varepsilon_1) \subseteq \psi(\varepsilon_2)$, $\xi_M(\varepsilon_1) \subseteq \xi_M(\varepsilon_2)$ and the set of 0-stable matchings is the same as the set of stable matchings.

We now state the continuity properties of the above correspondences.

**Proposition 5.** The correspondences satisfy the following continuity properties:

1. The correspondence $\psi$ is upper-hemicontinuous at 0.

2. If a matching $M$ admits a 0-stable matching then, the correspondence $\xi_M$ is upper-hemicontinuous at 0.

The continuity properties are important. They imply that by choosing small enough $\varepsilon$ then we can get arbitrarily close to the stable matchings. In fact for sufficiently small $\varepsilon$, we can obtain the matchings that admit a 0-stable matching and the search for stable matchings does not need to consider integer constraints.

**Proof:**

Since the set of stable matchings is non-empty, therefore the set of 0-stable matchings is non-empty. Also $\psi(0) \subseteq \psi(\varepsilon)$ for all $\varepsilon > 0$. Assume some matching $M$ does not admit a 0-stable matching but it admits $\varepsilon$-stable matchings for all $\varepsilon > 0$. This implies that there exists at least one payoff profile $U^0$ that is present in $\xi_M(\varepsilon)$ for each $\varepsilon > 0$. Pick any $(i, j) \in E$ such that both $i$ and $j$ are matched, not necessarily to each other and for any $\varepsilon > 0$, $v_i (j, U^0_j + \varepsilon) \leq U^0_i + \varepsilon$. Now since, $v_i (j, U^0_j) > U^0_i$ therefore from the continuity and monotonicity of the pareto payoff functions, there exists a $\delta > 0$ such that $v_i (j, U^0_j + \delta) > U^0_i + \delta$. This contradicts the assumption that $U^0$ is present in $\xi_M(\varepsilon)$ for each $\varepsilon > 0$. Therefore, $M$ must not admit $\varepsilon$-stable matchings for all $\varepsilon > 0$. Therefore, for sufficiently low $\varepsilon$, $\psi(\varepsilon)$ only consists of matchings that admit 0-stable matchings. This proves the first claim.

Similarly, assume a matching $M$ admits a 0-stable matching. Pick a sequence $\{\varepsilon\}_n$ that converges to 0 and a sequence $\{x\}_n$ such that $x_n \in \xi_M(\varepsilon_n)$. Assume that $\{x\}_n$ converges to $x$. We will show that $x \in \xi_M(0)$. Assume that $x \in \xi_M(0)$. Therefore there exists $(i, j) \in E$ such that both $i$ and $j$ are matched, not necessarily to each other
and \( v_i(j, x_j) > x_i \). Therefore, there exists a \( \delta > 0 \) such that \( v_i(j, x_j + \delta) > x_i + \delta \) and \( x \notin \xi_M(\delta) \). Since \( \xi_M(\delta) \) is a closed set therefore its complement is open and therefore there exists a sufficiently small ball \( B \) around \( x \) such that for all \( y \in B \), \( v_i(j, y_j) > y_i + \delta \). Therefore for all \( \varepsilon_n < \delta \), \( \{x\}_n \notin B \). This implies that \( \{x\}_n \) does not converge to \( x \). By contradiction, the second claim holds.

### 2.8.2 Natural Myopic Dynamics

In this section, we introduce natural myopic dynamics that converge to strictly \( \varepsilon \)-stable weighted matching.

An **independent set** of \( S \) is a set of agents \( Y \subseteq X \) such that for any two agents \( i, j \in Y, (i, j) \notin E \). An example of an independent set is \( A \) and in fact all subsets of \( A \) are independent sets.

The dynamics are as such: Given any \( \varepsilon > 0 \), the dynamics are as follows. At any time \( t \geq 0 \), an independent set \( Y^t \subseteq X \) is chosen randomly and independent of previous choices, to make a proposal. We will refer to the matching at time \( t \) as \( M^t \) and the splits as \( s^t \) and the payoff profile by \( U^t \). All independent sets have a positive probability of being chosen. The agents in \( Y^t \) can either choose one neighbor other than their existing matches and make an offer to split the edge with the neighbor or do nothing. In the second stage, the neighbors of the selected agents either accept an offer or reject the offer. The responders may receive offers from more than one agent but they can only accept one offer. If the responder accepts an offer, then the two agents enter into a contract according to the offer. Agents must break existing contracts to enter into a new one. Any agent that breaks a contract bears a one time penalty of \( \varepsilon \). The set of active contracts at any time comprises the matching at that time. We assume that each agent observes the current payoffs of all its neighbors but does not observe anything else in the network.

We say that the dynamics have converged at time \( T \), if for all times \( t > 0 \) no offers are made. Thus, the dynamics are very similar to the salary adjustment process in [35]. There are two differences between this dynamics and the salary adjustment process. The first difference is that unlike the salary adjustment process, agent from
both sides of the market may make offers. This relaxes the constraints provided by
the institutional structure in the salary adjustment process. The second difference is
that at different times different sets of agents may make proposals. This simulates the
asynchronous nature of offers as it may happen in distributed systems. The salary
adjustment process was designed for a market owner who runs the central
Clearing House for workers and firms. My dynamics simulates the real-world process
of offers between agents distributed in a network. It is natural to assume that the
offers will be asynchronous and agents from both sides of the network may make
offers.

Due to these differences, my dynamics may converge within a larger set. The
salary adjustment process converges at an $\varepsilon$-stable matching that is optimal for one
side of the network. However, my dynamics can converge at $\varepsilon$-stable matchings that
are not extreme. We demonstrate it by the following example.

**Example 1.** Consider $A = \{a_0, a_1\}, B = \{b_0, b_1\}, E = \{(a_0, b_0), (a_1, b_0), (a_1, b_1)\}$,
$w(a_0, b_0) = w(a_1, b_0) = w(a_1, b_1) = 1$. The Pareto payoff function are $v_{a_0}(b_0, x) = v_{a_1}(b_0, x) = v_{a_1}(b_1, x) = x$. Assume that $\varepsilon = 0.01$ and the initial matching is empty.
At time 0, $b_0$ is selected to propose and at time 1, $a_1$ is selected to propose. At time
0, $b_0$ may propose to $a_0$ with a split $s_{a_0} = 0$ and $s_{b_0} = 1$. At time 1, $a_1$ will propose to $b_1$ with a split $s_{b_1} = 0$ and $s_{a_1} = 1$. Both proposals will be accepted and the resulting
matching is $\varepsilon$-stable. We note that the outcome is not optimal for either side $A$ or
side $B$.

We point out that our dynamics generalizes the salary adjustment process. If
$Y^t = A$, for all $t \geq 0$, then the dynamics is equivalent to the salary adjustment
process. In this case, we know that the dynamics converges to an approximately
stable matching following [35]. We state the result for completeness.

**Proposition 6.** If $Y^t = A$, for all $t \geq 0$ and $M^0 = \emptyset$ then in finite time, the dynamics
converges to a strictly $\varepsilon$-stable matching.

Our main result is that for any $\varepsilon > 0$, the dynamics converges to a strictly $\varepsilon$-stable
matching almost surely.
**Theorem 4.** Given any initial state, \((M^0, s^{M^0})\), and any \(\varepsilon > 0\), the dynamics converges to a strictly \(\varepsilon\)-stable matching almost surely.

Before we prove this theorem, we give another result that we will use to prove this theorem. It turns out that even if the initial state of the matching is nonempty, there exists a sequence of independent sets such that if they propose then the dynamics will converge to an approximately stable matching. We give this result now and we will use it in the proof of theorem 4.

**Proposition 7.** Given any initial state, \((M^0, s^{M^0})\), that is not \(\varepsilon\)-stable there exists a finite sequence, \(\{Y\}^t\) such that the dynamics converges to a strictly \(\varepsilon\)-stable matching.

**Proof:**

We will give a proof by construction. Assume that the initial state is not \(\varepsilon\)-stable. At time \(t\), we will refer to the set of matched agents in \(A\) as \(A^t = \{i \in A : M^t(i) \neq \phi\}\). We define \(A^* = A^t\), such that each agent, \(i \in A^*\) has at least one matched neighbor, \(j\), with \(v_i^t(j, U_j^t + \varepsilon) \geq U_i^t + \varepsilon\) or unmatched neighbor with \(v_i^t(j, U_j^t + \varepsilon) \geq U_i^t\).

While \(A^* \neq \phi\) and , set \(Y^t = A^t\). For all subsequent \(t\), set \(Y^t = A^t\).

We will show that given the sequence \(\{Y\}^t\) defined above, the dynamics converges to a stable matching in finite time. First we will show that at some finite time \(A^* = \phi\). Then by similar arguments as in the proof of proposition 6 we will show that the dynamics converges to an \(\varepsilon\)-stable matching. To show that at some finite time \(i \in A^{t+1}\), we note that:

1. Until \(A^* = \phi\), \(A^t\) is weakly shrinking, i.e.- \(A^{t+1} \subseteq A^t\).

2. If \(i \in A^{t+1}\), then \(U_i^{t+1} \geq U_i^t\) and at least one agent \(i \in A^{t+1}\) has \(U_i^{t+1} > U_i^t\).

The first note is because of the following. Since agents outside \(A^t\) do not propose they cannot be matched and hence \(A^t\) cannot grow. The second note is because of the following. At any time \(t\) all nodes in \(A^*\) must propose to one of their neighbors such that if accepted the payoff of the proposer and the acceptor will both improve by at least \(\varepsilon\). Since, at least one of the proposals is accepted, therefore at least one
2.8. APPROXIMATELY STABLE MATCHINGS AND CONVERGENT DYNAMICS

proposer, \( i \in A^* \) has \( U_{i}^{t+1} \geq U_{i}^{t} + \varepsilon \). Since \( i \) is matched, therefore \( i \in A^{t+1} \) and therefore at least one agent \( i \in A^{t+1} \) has \( U_{i}^{t+1} \geq U_{i}^{t} \). Now all agents in \( i \in A^{t+1} \) were also in \( A^{t} \) and since they are matched at time \( t+1 \), so either these agents made new matches thus improving their payoffs by at least \( \varepsilon \) or stayed with the same matches thus maintaining their payoffs.

We also note that the payoffs of the agents are bounded. The payoff of any agent \( i \in A \) is bounded by \( \max_{j \in Nbr(i)} v_{i}(j, r_{j}) \). Therefore in some finite time, \( T^{0}, A^{*t} = \phi \).

After, time \( T \), we note that the payoffs of all agents in \( B \) is weakly increasing and until the matching is \( \varepsilon \)-stable, the payoff of at least one agent in \( B \) increases at all times \( t \geq T^{0} \). This is because as in the salary adjustment process and the deferred acceptance algorithm, the agents in \( B \) once matched are always matched, and they only accept an offer if it improves their payoff by at least \( \varepsilon \). Since, the payoffs of all agents are bounded, therefore by the same argument as above, the dynamics converges to a \( \varepsilon \)-stable matching in finite time. This completes the proof.

We now give the proof of the main theorem.

**Proof:**

From proposition 7, we know that given a state, \( (M^{t}, s^{M^{t}}) \), at any time \( t \), there exits a finite sequence of independent sets \( Y_{t+1}, Y_{t+2}, ..., Y_{T} \), such that if this sequence of independent sets are selected to propose, then the dynamics will converge to a \( \varepsilon \)-stable matching. We will denote \( E_{t} \) as the event that the shortest of such sequences is chosen to propose at time \( t+1 \) until time \( T \). We also know that probability of selecting any independent set is non-zero and thus is lower bounded by some \( \delta = \min \{ P(Y) : Y \text{ is an independent set} \} \). Therefore, the probability, \( P(E_{t}) \), of the event \( E_{t} \) is lower bounded by \( \delta^{T-t} \) since the choice of independent sets at each time is independent of the choice at other times. We denote \( T^{*} \) to be the maximum length of such sequences over all possible states, \( (M, s^{M}) \). It is clear that \( T^{*} \) is finite. Consider the sequence of events \( \{ R \} \), where \( R_{n} \) is the event that from time \( nT^{*} + 1 \) onwards, the sequence \( Y_{nT^{*}+1}, Y_{nT^{*}+1}, ..., Y_{(n+1)T^{*}} \) is a sequence such that, \( (M^{(n+1)T^{*}}, s_{M^{(n+1)T^{*}}}^{(n+1)T^{*}}) \) is \( \varepsilon \)-stable. The probability, \( P(R_{n}) \), of each event, \( R_{n} \) is lower bounded by \( \delta^{T^{*}} \) and the events are independent. Therefore
\[ \sum_{n=1}^{\infty} P(R_n) = \infty \]

Following Borel-Cantelli lemma, infinitely many of \( \{R\}_n \) will occur with probability 1. Therefore the dynamics will converge almost surely to a \( \varepsilon \)-stable matching.

## 2.9 Conclusions and Future Work

In this chapter we extended the stable matching problem in bipartite networks to the general scenario where agents derive value from the part of the split as well as the agent they are matched to. We studied a general case when the payoff is continuous and strictly increasing in the part of the split and introduced the Generalized Hungarian algorithm to find the stable matching that is optimal for one side of the network. We also characterized the set of stable matchings for a special class of bipartite networks. With additional structure on the payoff functions and correlations between payoff functions, more efficient methods can be employed and will be an interesting line of future work. Our algorithm can be an important tool in the study of generalized hungarian algorithm. We also introduced natural dynamics that converge to approximately stable matchings. In the future, we will extend these dynamics to general networks that will also provide algorithms for computing stable matchings in general networks that admit a stable matching. Another important line of investigation is the welfare properties of stable matchings and the duality gap between the optimal matchings and stable matchings. Finally, we wish to apply our results to the study of family economics.
2.10 Proofs of Theorem 1 and associated Lemmas and Corollaries

2.10.1 Proof of Theorem 1

We will first give a construction of the path function and prove its continuity and monotonicity. Pick an arbitrary \( i^o \in A \) and define the function \( g_{i^o} : (0, 1) \rightarrow \mathbb{R} \), where \( g_{i^o}(x) = \tan\left(\pi\left(x - \frac{1}{2}\right)\right) \). We define \( f^S(x) \) as the offer profile \( O \) such that \( O_{i^o} = g_{i^o}(x) \). From lemmas 4 and 2, we know that \( f^S \) is a well defined function. Since the range of \( g_{i^o} \) is whole of \( \mathbb{R} \), therefore the range of \( f^S \) the set of all SASTGOPs.

We now show that \( f^S \) is the desired path function. From lemma 3, we know the function \( \pi_{i^o} \circ f^S \) is strictly increasing for all \( i^o \in A \) and strictly decreasing for all \( i^o \in B \). We will prove the continuity of \( f^S \) by contradiction. Assume that \( f^S \) is discontinuous at some \( x^* \in (0, 1) \). Let \( \lim_{x \to x^-} f^S(x) = O^\inf \) and \( \lim_{x \to x^+} f^S(x) = O^\sup \). If \( f^S \) is discontinuous in the offer on agent \( i \in A \), then \( O^\inf_i < O^\sup_i \) and from the construction of \( f^S \), \( O^\inf_i = O^\sup_i = \pi_{i^o} \circ f^S(x^*) \). From the monotonicity lemma 3, \( \pi_{i^o} \circ f^S(x) < O^\inf_i \) for all \( x < x^* \) and \( \pi_{i^o} \circ f^S(x) > O^\inf_i \) for all \( x > x^* \) and from the uniqueness lemma 2, \( \pi_{i^o} \circ f^S(x^*) \) is unique. Since the range of \( f^S \) covers the set of all SASTGOPs, therefore, there are no SASTGOPs that have offers \( (O^\inf_i, O^\sup_i) \setminus \{\pi_{i^o} \circ f^S(x^*)\} \) on \( i \) contradicting the existence lemma 4. A similar argument shows the contradiction with the corollary 2. Hence, \( f^S \) is a continuous path function and \( \pi_{i^o} \circ f^S \) is a continuous and strictly increasing function for all \( i^o \in A \) and continuous and strictly decreasing function for all \( i^o \in B \). Hence, \( \pi_{i^o} \circ f^S \) is bicontinuous for all \( i^o \in X \).

By the application of the monotonicity lemma 3, and existence lemma 4, \( \lim_{x \to 0} \pi_{i^o} \circ f^S(x) = -\infty \) and \( \lim_{x \to 1} \pi_{i^o} \circ f^S(x) = \infty \) for all \( i^o \in A \) and further by the application of corollary 2, \( \lim_{x \to 0} \pi_{i^o} \circ f^S(x) = \infty \) and \( \lim_{x \to 1} \pi_{i^o} \circ f^S(x) = -\infty \) for all \( i^o \in B \). This completes the proof of the theorem.
2.10.2 Proof of Corollary 1

Pick an ASTAN $S$ and $r \in \mathbb{R}^{|A \cup B|}$. Pick the SASTGOP $O^S$ as defined in the corollary. From theorem 1, we know that $O^S = f^S \left( \max_{i \in A} \left( \pi_i \circ f^S \right)^{-1} (r_i) \right)$. Clearly, $O^S \in \emptyset$. Also, from theorem 1, $f^S \left( \left( \pi_i \circ f^S \right)^{-1} (r_i) \right) \leq_A f^S \left( \left( \pi_i \circ f^S \right)^{-1} (x) \right)$ for all $i \in A$ and $x \geq r_i$. Therefore, $O^S = f^S \left( \max_{i \in A} \left( \pi_i \circ f^S \right)^{-1} (r_i) \right) \leq_A O$ for all $O \in \emptyset$. Hence, $O^S$ is the $A$-minimum SASTGOP in $\emptyset$.

2.10.3 Proof of Lemma 2

Assume there exists two stable alternating spanning tree generating offer profiles $O^1$ and $O^2$ with $O^1_i = O^2_i = o_i$. Then $EQ(O^1) \neq EQ(O^2)$ or else $O^1 = O^2$. Let $M^o(O^1)$ and $M^o(O^2)$ be the associated near perfect matchings and $T^{M^o}(O^1)$ and $T^{M^o}(O^2)$ be the associated alternating spanning trees. Clearly $T^{M^o}(O^1) \neq T^{M^o}(O^2)$ or else $O^1 = O^2$. Without loss of generality assume that $i$ is unmatched in both $M^o(O^1)$ and $M^o(O^2)$. In both alternating spanning trees, the agents at even distances from $i$ belong to $A$ and the agents at odd distances from $i$ belong to $B$. Define

$$AM^1_i = \{i' \in A : O^1_{i'} > O^2_{i'}\} \quad \text{and} \quad BM^1_i = \{j' \in B : O^1_j > O^2_j\}$$

$$AM^2_i = \{i' \in A : O^1_{i'} < O^2_{i'}\} \quad \text{and} \quad BM^2_i = \{j' \in B : O^1_j < O^2_j\}$$

Since both offer profiles are stable, therefore following proposition 2:

1. In the alternating tree $T^{M^o}(O^1)$, the parents and children of agents in $AM^1_i$ and $BM^1_i$ must belong to $BM^2_i$ and $AM^2_i$ respectively.

2. In the alternating tree $T^{M^o}(O^2)$, the parents and children of agents in $AM^2_i$ and $BM^2_i$ must belong to $BM^1_i$ and $AM^1_i$.

Since the agents at odd distance in the alternating trees have exactly one child, and all agents have exactly one parent, therefore:

- We have $|BM^2_i| \geq |AM^1_i|$ and $|AM^1_i| \geq |BM^2_i| + 1$. This gives a contradiction $|BM^2_i| \geq |AM^1_i| \geq |BM^2_i| + 1 > |BM^2_i|$. 

2.10. PROOFS OF THEOREM 1 AND ASSOCIATED LEMMAS AND COROLLARIES

- We have \(|AM_i^2| \geq |BM_i^1|+1\) and \(|BM_i^1| \geq |AM_i^2|\). This gives a contradiction 
  \(|AM_i^2| \geq |BM_i^1|+1 > |BM_i^1| > |AM_i^2|\).

Therefore all the sets \(AM_i^1, AM_i^2, BM_i^1, BM_i^2\) are empty which implies that \(O^1 = O^2\).

2.10.4 Proof of Lemma 3

Clearly, there is no \(i' \in A\), with \(O^1_{i'} = O^2_{i'}\), otherwise by lemma 2 \(o^1_i = o^2_i\). Define

\[ AM'_i = \{ i' \in A : O^1_{i'} > O^2_{i'} \} \quad \text{and} \quad BM'_i = \{ j' \in B : O^1_{j'} < O^2_{j'} \} \]

Since both offer profiles are stable, therefore following proposition 2:

1. In the alternating tree \(T^{M\mathcal{S}}(O^1)\), the parents and children of agents in \(AM'_i\) must belong to \(BM'_i\).

2. In the alternating tree \(T^{M\mathcal{S}}(O^2)\), the parents and children of agents in \(BM'_i\) must belong to \(AM'_i\).

Since the agents at odd distance in the alternating trees have exactly one child, and all agents have exactly one parent, therefore using 1, we have \(|BM'_i| \geq |AM'_i|\) and from 2, we have \(|AM'_i| \geq |BM'_i|+1\). This gives a contradiction \(|BM'_i| \geq |AM'_i| \geq |BM'_i|+1 > |BM'_i|\). Therefore, we have \(AM'_i\) and \(BM'_i\) empty which implies that for all \(i' \in A\), \(O^1_{i'} < O^2_{i'}\) and for all \(j' \in B\), \(O^1_{j'} > O^2_{j'}\).

2.10.5 Proof of Lemma 4

The existence of SASTGOPs is evident from corollary 3. However, in appendix 2.16.1, we give a constructive proof of existence of the SASTGOPs. The construction is similar to the Generalized Hungarian algorithm with some differences. For the construction of SASTGOPs, we relax the individual rationality constraint. We also need to make sure that the offer on the chosen agent is the one provided by the lemma. Also, we assume that the bipartite network is an ASTAN. We first show that the construction is correct for ASTAN with only one agent in \(B\). Then we
assume that the construction is correct for all ASTANs with \(|B| \leq n\) and show that the construction is correct for all ASTANs with \(|B| = n + 1\). The correctness, then follows by induction.

### 2.10.6 Proof of Corollary 2

Since in the equality subgraph induced by any SASTGOP, \(O\), the agents in \(B\) are connected to at least one agent in \(A\) and since the SASTGOP is stable, therefore the offer on any agent \(j \in B\) is \(O_j = \max_{i \in \text{Nbr}(j)} v_i(j, O_i)\). Pick any \(o_j \in R\). For each \(i \in \text{Nbr}(j)\), pick the SASTGOP, \(O^{(i)}\) with \(O^{(i)}_i = v_i(j, o_j)\). Lemma 4 implies that such a SASTGOP exists. Find the neighbor \(i^o\) and the associated SASTGOP, \(O^{(i^o)}\) that is A-maximum of all such SASTGOPS. Lemma 3 implies the existence of such a neighbor and the SASTGOP. The offer on any neighbor \(i \in \text{Nbr}(j) \notin \{i^o\}\) for this SASTGOP is \(O^{(i^o)}_i = o_j\). Thus we have found a SASTGOP \(O^{(i^o)}\) with \(O^{(i^o)}_j = o_j\).

### 2.11 Proof of Theorem 2 and associated propositions

#### 2.11.1 Proof of Proposition 3

At \(t = 0\) we know the offer profile is individually rational because for all \(j \in B\), set \(O^0_j = r_j\) and for all \(i \in A\), \(O^0_i = \max_{j \in \text{Nbr}(i)} v_i(j, O^0_j) \geq r_i\). It is also stable because by construction for all \(i \in A\), \(O^0_i = \max_{j \in \text{Nbr}(i)} v_i(j, O^0_j)\). Assume that for some \(t \geq 0\), the offer profile \(O^t\) is stable and individually rational. If at iteration \(t + 1\), the offers do not change, then offer profile \(O^{t+1}\) is stable and individually rational. If at iteration \(t + 1\), the offers change then we give the following argument. From the construction of \(O^{t+1}\), and by theorem 1 we know that:

- for all \(i' \in A \setminus A^t\), \(O^{t+1}_{i'} = O^t_{i'}\) and for all \(j' \in B \setminus B^t\), \(O^{t+1}_{j'} = O^t_{j'}\)
- for all \(i' \in A^t\), \(r_{i'} < O^{t+1}_{i'} < O^t_{i'}\) and for all \(j' \in B^t\), \(O^{t+1}_{j'} > O^t_{j'}\)
Therefore, \( O^{t+1} \) is individually rational. The edges in \( E \) can be divided into four mutually exclusive subsets: (i) \( E_1 = E \cap (A^T \times B^T) \), (ii) \( E_2 = E \cap ((A \setminus A^T) \times B^T) \), (iii) \( E_3 = E \cap (A^T \times (B \setminus B^T)) \), and (iv) \( E_4 = E \cap ((A \setminus A^T) \times (B \setminus B^T)) \). From the construction, since \( O^{t+1}_{|X^r} \) is a SASTGOP within the subgraph \( S_{|X^r} \), therefore, for all \((i', j') \in E_1, O^{t+1}_{i'} \geq v_{i'} (j', O^{t+1}_{j'}) \). Since \( O^t \) was stable therefore for all \((i', j') \in E_4, O^{t+1}_{i'} = O^t_i \geq v_{i'} (j', O^{t+1}_{j'}) \). From the construction of \( O^{t+1} \) and the stability of \( O^t \) and since the pareto payoff functions are strictly decreasing therefore for all \((i', j') \in E_2, O^{t+1}_{i'} = O^t_i \geq v_{i'} (j', O^{t+1}_{j'}) \). Since for all \( i' \in A_t, O^{t+1}_{i'} = eo_{i'} \), therefore for all \((i, j) \in E_3, O^{t+1}_{i'} \geq eo_{i'} = v_{i'} (j', O^{t+1}_{j'}) = v_{i'} (j', O^{t+1}_{j'}). \)

Therefore we see that for all \((i', j') \in E, O^{t+1}_{i'} \geq v_{i'} (j', O^{t+1}_{j'}) \) and the offer profile \( O^{t+1} \) is stable. By induction at any iteration \( t \geq 0 \), the offer profile \( O^t \) is stable and individually rational.

### 2.11.2 Proof of Proposition 4

At \( t = 0 \), since all agents \( j \in B \) have offers \( r_j \), this holds true. Assume that at some iteration \( t \geq 0 \), if there exists \( j \in B \) with \( O^t_j > r_j \), then \( j \) is matched in \( M^t \). If the offer profile does not change at time \( t + 1 \), then nothing changes and the proposition holds at \( t + 1 \). Otherwise, at time \( t + 1 \), one of the following happens:

1. \( T^t \) connects to a joining agent \( j \in B \setminus B^T \) such that \( j \) has an alternating path to an unmatched agent \( j^o \).

2. \( T^t \) connects to a joining agent \( j \in B \setminus B^T \) such that \( j \) has an alternating path to a matched agent \( i^o \in A \setminus A^T \) with \( O^{t+1}_{i^o} = r_i \).

3. \( O^{t+1}_i = r_i \) for some \( i \in A^T \).

4. \( T^t \) connects to joining agents \( J \subseteq B \setminus B^T \) but the above cases do not happen.

At \( t + 1 \), all agents in \( B^T \) are matched to the agents in \( A^T \) by the selected near-perfect matching \( M^*_t \mid X^r \) at time \( t + 1 \). From the assumption at time \( t \), we know that if there exists \( j \in B \) with \( O^t_j > r_j \), then \( j \) is matched in \( M^t \). Therefore, any agent \( j \in B \setminus B^T \) with \( O^t_j > r_j \) is matched to some agent in \( A \setminus A^T \) in \( M^t \) since \( EQ^t \cap (A \setminus A^T \times B \setminus B^T) = \phi \). For all agents \( j \in B \setminus B^T \), we know that \( O^{t+1}_j = O^t_j \).
In the fourth case \( M^{t+1} = M^*_t \cup M^t_{X \setminus X_T} \) and hence all agents in \( B \) that were matched at time \( t \) are matched at time \( t+1 \). In the first case, before exchanging any edges, we see that as in the third case, all agents in \( B \) that were matched at time \( t \) are matched at time \( t+1 \). By exchanging the edges within the matching \( M^{t+1} \) with edges outside the matching in the augmenting path, the agents in the augmenting path still stay matched and it does not change matching outside the augmenting path, so all agents in \( B \) that were matched at time \( t \) are matched at time \( t+1 \). For the second and third case, by the same argument, all agents in \( B \) that were matched at time \( t \) are matched at time \( t+1 \). Thus we see that all agents in \( B \) that were matched in \( M^t \) are matched in \( M^{t+1} \). Therefore any agent \( j \in B \) with \( O^t_j > r_j \) that was matched in \( M^t \) is matched in \( M^{t+1} \). By induction, the proposition holds.

### 2.11.3 Proof of Theorem 2

From proposition 3, we know that the offer profile \( O^t \) for any time \( t \geq 0 \) is stable and individually rational. We only need to show that at some finite time \( t^* \), \( O^{t*} \) is feasible. From proposition 4, we know that for all \( t \geq 0 \), all agents in \( B \) with offers different from their reserved payoffs are matched. So we only need to show that at some finite time \( t^* \), all agents in \( A \) with positive offers are matched, i.e.- \( m^{t*} = 0 \). For any \( t \), the root of the alternating tree \( T^t \) is \( i^t_1 \). At each time \( t' > t \), until \( i^t_1 \) is matched, the alternating tree \( T^{t'} \) has the root \( i^{t'}_1 = i^t_1 \) and one of the following happens:

1. \( T^{t'-1} \) connects to a joining agent \( j \in B \setminus B^{t'-1} \) such that \( j \) has an alternating path to an unmatched agent \( j^* \) or a matched agent \( i^* \in A \setminus A^{t'-1} \) with \( O^{t'}_{i^*} = r_i \).
2. \( O^{t'}_i = r_i \) for some \( i \in A^{t'-1} \).
3. \( T^{t'-1} \) connects to joining agents \( J \subseteq B \setminus B^{t'-1} \) but the above cases do not happen.

In the third case, \( B^{t'-1} \subset B^{t'} \). Thus the alternating tree rooted at \( i^t_1 \) increases. Since \( B \) is finite, therefore the third case happens only finitely many times. Therefore, at some finite time \( t^* \), either of the first two cases happen. If the first case happens at \( t^* \), then there is an augmenting path from \( i^t_1 \) to \( j^* \) or there is an alternating path
from \(i_1^t\) to some \(i^o\) with \(O_{i_1^o}^o = r_{i_1^o}\). Thus, by construction of matching \(M^t\) at time \(t^o\), \(i_1^t\) is matched in \(M^t\) and \(m^t = m^{t-1}\). If the second case happens at \(t^o\), then there is an alternating path from \(i_1^t\) to some \(i^o\) with \(O_{i^o}^{t^o} = r_i\). Thus, by construction of matching \(M^t\) at time \(t^o\), \(i_1^t\) is matched in \(M^t\) and \(m^t = m^{t-1}\). Hence at time \(t^o\), \(i_1^t\) is matched in \(M^t\) and \(m^t = m^{t-1}\). Since \(m^0\) is finite, and \(m^t\) decreases by 1 in finitely many iterations when \(m^t > 0\), therefore at some finite time \(t^*\), \(m^{t^*} = 0\). Since \(m^{t^*} = 0\), therefore all agents \(i \in A\) with \(O_{i^o}^{t^*} > r_i\) are matched in \(M^{t^*}\) and from proposition 4 that all agents \(j \in B\) with \(O_{j^o}^{t^*} > r_j\) are matched in \(M^{t^*}\). Therefore by proposition 1 \(O^{t^*}\) is a feasible offer profile.

### 2.12 Proof of Theorem 3 and associated Lemmas and Corollaries

#### 2.12.1 Proof of Lemma 5

Assume \(S\) is an ASTAN and \(r\) is the reserved payoff profile. Assume there exists a feasible and stable offer profile \(O^o\) that is strictly preferred by at least one agent in \(A\) over the A-minimum SASTGOP \(O^*\) with \(O_i^* \geq r_i\) for all \(i \in A\). Let us denote \(AM_1 = \{i \in A : O_i^* < O_i^o\}\) and \(BM_1 = \{j \in B : O_j^* > O_j^o\}\). From the stability of \(O^*\), the neighbors of \(AM_1\) in \(EQ(O^o)\) belong to \(BM_1\) and since \(O_i^o > O_i^* \geq r_i\) for all \(i \in AM_1\) therefore from the feasibility of \(O^o\), all agents in \(AM_1\) are matched in any matching \(M^o\) that supports \(O^o\). Therefore \(|AM_1| \leq |BM_1|\). From the stability of \(O^o\), the neighbors of \(BM_1\) in \(EQ(O^*)\) belong to \(AM_1\) and since all agents in \(BM_1\) are the internal agents of the alternating spanning tree \(T\) generated by any near-perfect matching \(M^*\) in the equality subgraph \(EQ(O^*)\) therefore by proposition 1 \(|BM_1| \leq |AM_1| - 1\). The two inequalities contradict unless both \(AM_1\) and \(BM_1\) are empty implying that there is no agent in \(A\) that prefers \(O^o\) over \(O^*\).
2.12.2 Proof of Theorem 3

We first note that as a consequence of lemma 3, the A-maximum stable, feasible and individually rational offer profile is the same as the B-minimum stable, feasible and individually rational offer profile. We will show by induction that for each iteration $t \geq 0$, the offer profile $O^t$ is A-preferred over the A-maximum and B-minimum stable, feasible and individually rational offer profile, $O^*$ and $O^*$ is B-preferred over $O^t$. This will imply that the point of convergence $O^t$ is A-preferred over $O^*$ and $O^*$ is B-preferred over $O^t$. Since $O^t$ is stable, feasible and individually rational, therefore $O^t = O^*$ and consequently the constructed weighted matching $(M^t, s^{M^t})$ is a A-maximum and B-minimum generalized stable matching.

For $t = 0$ it is trivially true because for all $i \in A$, $O^0_i \geq r_i$ and $O^0_i \geq v_i(j, r_j)$ for all $j \in Nbr(i)$. Also, for all $j \in B$, $O^0_i = r_i$ and hence is least B-preferred over all individually rational offer profiles. Assume, that for some $t \geq 0$, the offer profile $O^t$ is A-preferred over $O^*$ and $O^*$ is B-preferred over $O^t$. We will show that the offer profile $O^{t+1}$ is A-preferred over the A-maximum stable, feasible and individually rational offer profile. Since the offers only change for the agents $X^{T^t}$ in the alternating tree $T^t$, it is sufficient to show that all agents in $A^{T^t}$ prefer $O^{t+1}$ over $O^*$ and all agents in $B^{T^t}$ prefer $O^*$ over $O^{t+1}$. For the subgraph $S_{|X^{T^t}}$, which is an ASTAN, consider two reserved payoff profiles $r^1$ and $r^2$ with $r^1_j = r^2_j = r_j$ for all $j \in B^{T^t}$, $r^1_i = eo_i$ for all $i \in A^{T^t}$, the expanding offer at iteration $t$ and $v_i^2 = \max\{\max_{j \in Nbr(i) \setminus B^{T^t}} v_i(j,O^*_j), r_i\}$. Consider the associated A-minimum of SASTGOPs, $O^2$, that are individually rational for all agents in $A^{T^t}$ for the reserved payoff profile $r^2$. The updated offer profile $O^{t+1}_{|X^{T^t}}$ is the A-minimum of SASTGOPs that are individually rational for all agents in $A^{T^t}$ given the reserved payoff profile $r^1$. Since the agents in $B \setminus B^{T^t}$ prefer $O^*$ over $O^t$ therefore $r^1_i \geq r^2_i$. This implies that $O^2 \preceq A^{T^t} O^{t+1}_{|X^{T^t}}$. From lemma 5, we know that $O^2$ is the A-maximum stable and feasible offer profile for the reserved payoff profile $r^2$ within the subgraph $S_{|X^{T^t}}$. Therefore, $O^2 \succeq A^{T^t} O^*_{|X^{T^t}}$. Since $O^*_{|X^{T^t}}$ is a stable and feasible offer profile for the reserved payoff profile $r^2$ within the subgraph $S_{|X^{T^t}}$. Combining the two inequalities we get $O^{t+1}_{|X^{T^t}} \succeq A^{T^t} O^2 \succeq A^{T^t} O^*_{|X^{T^t}}$ or all agents in
2.12. PROOF OF THEOREM 3 AND ASSOCIATED LEMMAS AND COROLLARIES

$A^{T^i}$ prefer $O^{t+1}$ over $O^*$. Since all agents in $B^{T^i}$ are connected to some agent in $A^{T^i}$ in the equality subgraph $EQ(O^{t+1})$ therefore from the stability of $O^{t+1}$ and lemma 3, $O^*_{|X^{T^i}} \succeq B^{T^i} O^{t+1}_{|X^{T^i}}$ or all agents in $B^{T^i}$ prefer $O^*$ over $O^{t+1}$.

2.12.3 Proof of Corollary 3

Assume $S$ is an ASTAN, and it admits an individually rational SASTGOP, $O$ for the reserved payoff profile $r$. Assume $O^*$ is the A-minimum individually rational SASTGOP. From lemma 5, $O^*$ is the A-maximum stable and feasible offer profile for the reserved payoff $r$. We only need to show that $O^*$ is individually rational for all agents in $B$. We note that $O \preceq_A O^*$ and by lemma 3, $O^* \preceq_B O$. Therefore, $O^*$ is an individually rational offer profile.
2.13 Appendix: Proofs of Propositions 1 and 2

2.13.1 Proof of Proposition 1

Proof:

(\rightarrow) Assume $O^*$ is a feasible offer profile and $M^*$ is a matching in $EQ(O^*)$, such that all nodes with offer different from their reserved payoffs are matched in $M^*$. We define a split $s^{M^*}$ as following.

For all $(i, j) \in M^*$, set $s^{M^*}(i) = u_i^{-1}(j, O_i^*)$,
$s^{M^*}(j) = u_j^{-1}(i, O_j^*)$,
for all unmatched $i \in X$, $s^{M^*}(i) = 0$.

The above is a well defined split because for all $(i, j) \in M^*$,

$$s^{M^*}(i) + s^{M^*}(j)$$
$$= u_i^{-1}(j, O_i^*) + s^{M^*}(j)$$
$$= u_i^{-1}(j, v_i(j, O_j^*)) + s^{M^*}(j)$$
$$= u_i^{-1}(j, u_i(j, u_j(i, s^{M^*}(j)))) + s^{M^*}(j)$$
$$= u_i^{-1}(j, u_i(j, w(i, j) - s^{M^*}(j))) + s^{M^*}(j)$$
$$= w(i, j) - s^{M^*}(j) + s^{M^*}(j)$$
$$= w(i, j).$$

Hence, we have a weighted matching $(M^*, s^{M^*})$ with payoff profile $U(M^*, s^{M^*}) = O^*$.

(\leftarrow) We first note that any matching that may have a payoff profile equal to the given offer profile $O^*$ must have all matching edges within $EQ(O^*)$. Let $M$ be the set of all possible matchings within $EQ(O^*)$. Assume that there does not exist any matching in $EQ(O^*)$, such that all nodes with offers different from their
reserved payoffs are matched. Pick any matching $M \in M$. At least one of the nodes with offer different from its reserved payoff is unmatched in $M$. Since the payoff of the unmatched node is its reserved payoff therefore the matching $M^*$ does not have any split with payoff profile equal to the offer profile $O^*$. Since we picked the matching $M$ arbitrarily, therefore for all matchings in $M$, there is no split such that the payoff profile of the resulting weighted matching is equal to the offer profile $O^*$. Therefore there is no weighted matching with payoff profile equal to the offer profile $O^*$. Therefore $O^t$ is not a feasible offer profile.

In this appendix we give the proofs of propositions 2 and 1.

2.13.2 Proof of Proposition 2

Proof:

We only need to prove the first inequality and the second inequality follows similarly. To prove the first inequality, assume that $O^1_j < O^2_j$. Then, $O^1_j < O^2_j = v_j(i, O^2_i) \leq v_j(i, O^1_i)$. The second inequality is due to the strict monotonicity of pareto payoff function. This implies that $O^1$ is not stable contradicting our assumption. Hence, by contradiction, the first statement is true.
2.14 Appendix: Graph Theoretic Terminology

In this section, we define some standard terminology that will be heavily used for the construction of a stable and feasible offer profile.

2.14.1 Paths

A path in the network $S$ is a connected subgraph $P = (X^P, E^P)$, where $X^P \subseteq X$ and $E^P \subseteq E_{|X^P}$, such that two nodes in $X^P$ have exactly one edge in $E^P$ and all other nodes in $X^P$ have exactly two edges in $E^P$. The two nodes with exactly one edge will be referred to as the end nodes. A path $P$ of length $N = |X^P| + 1$ can be seen as a sequence $\{x_n\}_{n \in \{0,\ldots,N\}}$ of nodes in $X$, such that for all $n < N$, $(x_n, x_{n+1}) \in E$. We call this a path from the source $x_0$ to the sink $x_N$. When the source and sink is determined for the path $P$, we will refer a path from the source $i$ to sink $j$ as $P_{i,j}$.

A subpath $P'$ of a path $P$ is a connected subgraph of the path $P$. Alternatively, a subpath $P' \subseteq P$ between nodes $x_m, x_M \in X^P$ is a subsequence $\{x_n^P\}_{n \in \{m,\ldots,M\}}$. A subpath is a path by itself. We will denote the set of paths from $i$ to $j$ in a network $S$ as $P^S_{i,j}$. Given an offer profile $O$, a path $P$ is feasible if $E^P \subseteq EQ(O)$. Given a node $i$ with offer $o_i \geq 0$, a path $P$ with an end node $i$ induces an offer for each node $x_n^P$ in the path as follows:

- $x_0^P = i$, Therefore, $O_{x_0^P} = o_i$

- For $0 < n \leq N$, $O_{x_n^P} = v_{x_n^P} \left( x_{n-1}^P, O_{x_{n-1}^P} \right)$.

For any pair of nodes $i, j$ and a path $P$ from $i$ to $j$, we define the path induced offer function $f_{i,j}^P : R \rightarrow R$ where $f_{i,j}^P(x)$ is the offer that $P$ induces for $j$ given $i$ has the offer $x$. $f_{i,j}^P$ is bicontinuous since the pareto payoff functions are bicontinuous. Also $f_{i,j}^P$ is strictly increasing if both $i, j \in A$ or both $i, j \in B$ and strictly decreasing if either $i \in A$ and $j \in B$ or $i \in B$ and $j \in A$ since pareto payoff functions are strictly decreasing.
2.14.2 Alternating Paths, Alternating Trees and Near-Perfect Matchings

Given a matching $M$, an **alternating path** is a path in which alternating edges belong to the matching and not to the matching $M$. The matching also induces directionality on the alternating paths in the following way. Direct all edges $(i, j) \in M$ from $j \in B$ to $i \in A$ and all edges $(i, j) \notin M$ from $i \in A$ to $j \in B$. An **augmenting path** is an alternating path that starts and ends at an unmatched node. An **alternating tree** for a matching $M$ is a tree $T^M$ whose root is the only unmatched node in the tree and has three properties: (i) every node at odd distance from root has degree 2 in the tree, (ii) all paths in the tree from the root are alternating paths, and (iii) all leaf nodes are at even distance from the root. Every alternating tree has one more node at even distance from root than at odd distance from root. A maximal alternating tree is an alternating tree that includes all nodes reachable on alternating paths from the root. In the rest of the chapter, we will imply maximal alternating trees when we mention alternating trees.

A matching $M^*$ is a **maximum matching** in the network $S$ if $|M^*| = \max\{|M|: M \text{ is a matching in } S\}$. A maximum matching $M^*$ can be obtained using the augmenting path algorithm [98]. By Berge’s Lemma, a matching $M$ is a maximum matching in $S$, if and only if there is no augmenting path in $S$ with respect to the matching $M$[98]. A **near-perfect matching** $M^\circ$ is a matching such that exactly one node is unmatched. A near-perfect matching exists only if $||A|−|B||= 1$. Given a maximum matching $M^*$, an **alternating forest** or a **Hungarian forest** [41] $F^{M^*}$ is a collection of alternating trees rooted at nodes in $A$ induced by the matching. The Hungarian forest $F^{M^*}$ consists of nodes reachable through alternating paths from unmatched nodes in $A$ and all such alternating paths. The Hungarian forest satisfies certain important properties that are given in proposition 8 in Appendix 2.15.
2.15 Appendix: Structural Properties of the Hungarian Forest

Proposition 8. Given a graph $E \subseteq A \times B$, a maximum matching $M \subseteq E$ that generates the Hungarian forest $F = F^M$ containing $m$ alternating trees, the following hold about the Hungarian forest:

- There are at least as many nodes from $B$ outside the Hungarian forest than from $A$, i.e. $|B \setminus B^F| \geq |A \setminus A^F|$.

- There is no edge between a node in $A^F$ and a node in $B \setminus B^F$ in the graph $E$, i.e. $(A^F \times B \setminus B^F) \cap E = \emptyset$. In other words, all the neighbors of $B \setminus B^F$ belong to $A \setminus A^F$ and $E \subseteq (A^F \times B^F) \cup (A \setminus A^F \times B^F) \cup (A \setminus A^F \times B \setminus B^F) = (A \times B^F) \cup (A \setminus A^F \times B \setminus B^F)$.

- All edges in the matching $M$ belong to $A^F \times B^F \cup A \setminus A^F \times B \setminus B^F$.

- Assume $|A| - |B| = n > 0$. Then the number of alternating trees in the Hungarian forest is $n$ more than the number of unmatched nodes in $B \setminus B^F$, i.e. $m = |B \setminus B^F| - |A \setminus A^F| + n$.

Proof: Since all nodes in $A \setminus A^F$ are matched to nodes in $B^F$, the first claim holds.

Pick a node $i' \in A^F$ and an alternating tree $T$ that contains $i'$. If there exists an edge $(i', j') \in E$ between node $i'$ and another node $j' \in B$, then:

- if the edge is in matching $M$, then the alternating path from the root of the alternating tree $T$ to $i'$ includes $j'$ and hence $j' \in B^F$.

- if the edge is not in the matching $M$, then the alternating path from the root of the alternating tree $T$ to $i'$ can be extended to include $j'$. Hence, $j' \in B^F$.

Hence, the second claim follows and the third claim follows from it.

We now prove the fourth claim. The number of unmatched nodes in $B \setminus B^F$ is $|B \setminus B^F| - |A \setminus A^F|$. Since all nodes in $A \setminus A^F$ and $B^F$ are matched and the number
of unmatched nodes in $A$ is $n$ more than the number of unmatched nodes in $B$, therefore the number of unmatched nodes in $A^F$ is $|B \setminus B^F| - |A \setminus A^F| + n$. Since each unmatched node in $A^F$ is the root of a unique alternating tree, and each alternating tree has a unique root that belong to $A^F$ therefore the number of alternating trees in the Hungarian forest is $n$ more than the number of unmatched nodes in $B \setminus B^F$. 
2.16 Appendix: Proof of Lemma 1

**Proof:** (→) Assume $S$ is an ASTAN. Then $S$ has a near perfect matching that induces an alternating spanning tree. Pick any such matching. Pick any subset $B' \subseteq B$. Since all nodes in $B$ are matched in the near-perfect matching, then from the Hall’s theorem [55], $B'$ has edges to at least $|B'|$ nodes in $A$. Since all nodes in $B'$ are interior nodes of an alternating tree, therefore, each node in $B'$ has one unique child it is matched to and one parent it is not matched to. Clearly, there is one parent node different from all the child nodes, or else, there will be a loop in the alternating tree. Hence, $B'$ has edges to at least $|B'|+1$ nodes in $A$ in the alternating tree within the network $S$.

Now, pick any $A' \subset A$ and a near-perfect matching $M$ with an alternating tree $T$. If $A'$ does not contain the root of the alternating tree, then all nodes in $A'$ are matched and hence, the number of neighbors of the set $A'$ must be at least $|A|$. If $A'$ contains the root of the alternating tree, then pick a node $i'$ not in $A'$ and change the matching $M$ by switching the edges within the matching with the edges outside the matching along the alternating path from the root of $T$ to $i'$. This creates a new near-perfect matching and an alternating spanning tree whose root is at $i'$. For this matching, all the nodes in $A'$ are matched and hence $A'$ must have at least $|A'|$ neighbors.

(←) Assume $S$ satisfies the conditions. Then following Berge’s lemma there exists a matching $M^*$ in $S$ that matches all but one node in $A$. Assume $i' \in A$ is the only unmatched node in $A$. Pick the maximal alternating tree $T$, rooted at $i'$. If $T$ contains all nodes in $A$ and consequently $B$, then $T$ is a spanning tree. If not, then $A^T \subset A$ and must have $|A^T|$ neighbors in $B$. Since $|B^T| < |A^T|$, therefore at least one of the neighbors must be outside $B^T$ implying that there is at least one alternating path from $i'$ goes outside alternating tree $T$. This contradicts the maximality of $T$ and hence by contradiction, $S$ is an ASTAN.
2.16.1 Proof of Lemma 4

We first give a summary of the proof. We give a constructive proof of Lemma 4. We use induction to prove this lemma. First we show that the lemma is satisfied for the ASTAN with only 1 node in $B$. Then we assume that the lemma holds for ASTAN with at most $n - 1$ nodes in $B$ and show that it holds for graphs with $n$ nodes in $B$. For the induction we pick an ASTAN with $|B| = n$, we pick an arbitrary node $i \in A$ and an arbitrary offer $o_i \in R$. For the finite graph, we create a sequence of stable offer profiles that converge to a stable alternating spanning tree generating offer profile. The starting offer profile has maximum path induced offers on the nodes in $A$, given the offer $o_i$ on $i$ and the minimum offers on the nodes in $B$ such that the offer profile is stable. We obtain the maximum matching in the equality subgraph induced by the offer profile. The sequence of offer profiles has decreasing offers on nodes in $A$ and increasing offers for nodes in $B$ and at each step, the matching is always the maximum matching in the equality subgraph, following Berge’s lemma [98]. We chose the next offer profile in the sequence as such. We pick an alternating tree for the matching in the equality subgraph and using the assumption for the induction, we set the offers on the nodes in this alternating tree such that at least one edge is added between the alternating tree and the rest of the equality subgraph. The new edge is from a node in $A$ to a node in $B$. We show the following properties for the sequence:

- We show that this offer profile is stable as shown in proposition 9.
- All nodes in $B$ that are once matched stay matched forever, and the nodes in $B$ that are unmatched has the same offer as their initial offers as shown in proposition 10.
- The matching size is weakly increasing and increases by one in finite number of steps after any step, as shown in proposition 12.
- The chosen node $i$ does not belong to the Hungarian forest until the Hungarian forest consists of only one alternating spanning tree covering all nodes in the
graph as shown in proposition 4. This result makes sure that the offer on node $i$ does not ever change.

Thus, we show that in finite number of steps the sequence converges to a stable alternating spanning tree generating offer profile with the offer on $i$ as $o_i$ as shown in proposition 13. Before we give the complete proof of the lemma, we introduce the construction of the sequence and show that the above properties of the sequence.

**Construction of SASTGOP**

Given an ASTAN $S$ with $n$ nodes in $B$, we pick an arbitrary node $i \in A$ and an arbitrary offer $o_i \in \mathbb{R}$. We assume that the lemma holds for all ASTANs with at most $n - 1$ nodes in $B$. We given the construction of a sequence of offer profiles that converges to a SASTGOP with offer $o_i$ on node $i$. The construction is very similar to the Generalized Hungarian Algorithm, however there are some critical differences because of different constraints. The differences between such a SASTGOP and the A-maximum feasible, stable and individually rational offer profile are following. First, the SASTGOP we are looking for need not be individual rational as we are not dealing with any reserved payoffs. Second, we need the equality subgraph induced by the SASTGOP to be an ASTAN. Third, we need a particular offer on the picked node $i \in A$. We will some additional terminology for the construction. Given a maximum matching $M^*(O) \subseteq EQ(O)$ and an alternating tree $T$, an **expanding node** is a node $i' \in A^T$ with an edge $(i', j') \in E \setminus EQ(O)$ with some $j' \in B \setminus B^T$. We will refer $C^T$ to be the set of **expanding nodes** for the tree $T$ and for each $i' \in C^T$, the respective **expanding offer** $eo_{i'} = \max_{j' \in Nbr^E((i') \setminus B^T} v_{i'} (j', O_{i'})$. We will also refer to $B \setminus B^T = Nbr^E (C^T) \setminus B^T$ as the set of **joining nodes** for the tree $T$. We now give the construction.

**Initial Offer Profile:** The initial offer profile $O^0$ is set as follows.

- For all $i' \in A$ set $O^0_{i'} = \max_{P \in \mathbb{P}_{i', j^*}} f_{i, i'} (o_i)$ (maximum path induced offer).
- For all $j \in B$, set $O^0_j = \max_{i' \in Nbr^S(j)} v_{i'} (i', O_{i'}^0).$
From the initialization, the offer profile \(O^0\) is stable and all nodes \(j \in B\) have at least one edge in the equality subgraph \(EQ\left(O^0\right)\).

**Initial Matching:** Pick a maximum matching \(M^0\) in \(EQ\left(O^0\right)\). The maximum matching can be found using the augmenting path algorithm.

**Initial Hungarian Forest:** Set \(m^t\), the number of unmatched nodes in \(A\), \(I^0 = \{i^0_0, i^0_1, ..., i^0_m\}\), the set of unmatched nodes in \(A\), \(T^0_{i^0_k}\), the alternating tree rooted at \(i^0_k \in I^0\), \(i^t = i^t_1\), the root of the chosen alternating tree for iteration and \(T^0 = T^0_{i^0_1}\). The union of all alternating trees is the Hungarian forest \(F^0\) at time 0.

**Test:** If at anytime \(t\), all of the following conditions are met then we have constructed the desired SASTGOP \(O^* = O^t\) and the corresponding alternating spanning tree \(T^t\). The conditions are:

1. \(O^t\) is stable.
2. \(O^t_i = o_i\).
3. \(F^t\) has exactly one alternating tree that spans all nodes in \(X\).

**Changing the Offer Profile:** At any time \(t\), if \(O^t\) is not a SASTGOP, pick \(i^t\) and the alternating tree \(T^t\). For each expanding node in \(C^{T^t}\) with neighbors outside \(T^t\), \(i^t \in A^{T^t}\), set the expanding offer \(eo_{i^t} = \max_{j^t \in Nbr(i^t) \setminus B^{T^t}} v_{i^t}(j^t, O^t_{j^t})\) and set \(O^t_{i^t}\), the SASTGOP in \(S_{X^{T^t}}\) that has \(O^t_{i^t} = eo_{i^t}\). Set \(i^*\), a node such that \(O^t_{i^*} \succeq_A O^t_{i^t}\) for all \(i^t \in A^{T^t}\) or equivalently for all expanding nodes \(i^t\), \(eo_{i^t} \leq O^t_{i^*}\).  

\(\text{\footnotesize{7}}\) By lemma 1 there exists a node \(i^t \in A^{T^t}\) with a neighbor in \(B \setminus B^{T^t}\) so the expanding offers are well defined. Since \(|B^{T^t}| < n\), therefore by the assumption for the problem, within the ASTAN \(S_{X^{T^t}}\), there exists a SASTGOP \(O^{i^*_t}_{X^{T^t}}\) with \(O^t_{i^*} = eo_{i^*}\). Consider any two expanding nodes \(i^t\) and \(i''\), their respective expanding offers \(eo_{i^t}\) and \(eo_{i''}\) and their respective SASTGOPS, \(O^{i^*_t}_{X^{T^t}}\) and \(O^{i''_t}_{X^{T^t}}\). By lemma 3, if \(eo_{i^t} > O^{i''_t}_{X^{T^t}}\), then \(eo_{i''} < O^{i^*_t}_{X^{T^t}}\). By repeated application of lemma 3, we find that there exists an expanding node \(i^*_t\) with expanding offers \(eo_{i^*_t}\) and SASTGOP, \(O^{i^*_t}_{X^{T^t}}\) such that for all expanding nodes \(i^t\), \(eo_{i^t} \leq O^{i^*_t}_{X^{T^t}}\).
Changing the Matching: Pick a near-perfect matching \( M^*_t \subseteq EQ_{\mathcal{X}_t} \) in \( S_{\mathcal{X}_t} \) that leaves \( i^t \) unmatched. \(^8\) First set \( M^{t+1} = M^t_{\mathcal{X}\setminus\mathcal{X}_t} \cup M^*_t \). If there exists an augmenting path from \( i^t \) with respect to the matching \( M^{t+1} \) within the equality subgraph \( EQ^{t+1} \), then switch the edges within the matching and outside the matching along the augmenting path to create a new matching \( M^{t+1} \). \(^9\)

Updated Hungarian Forest: If the matching size changed, then set \( m^{t+1} = m^t - 1 \), for all \( k < m^t \), set \( i^t_k = i^{t+1}_{k+1} \), \( I^{t+1} = \{i^{t+1}_1, ..., i^{t+1}_{m^t+1}\} \). If the matching did not change set \( m^{t+1} = m^t \), for all \( k \leq m^t \), set \( i^t_k = i^t_k \), \( I^{t+1} = I^t \). For all \( i^{t+1}_k \in I^{t+1} \), \( T^{t+1}_{i^t_k} \) is the new alternating tree rooted at \( i^{t+1}_k \). The union of all alternating trees is the Hungarian Forest \( F^{t+1} \) at time \( t+1 \).

Before, we show that the sequence of offer profiles converges to a SASTGOP, we observe some important properties of the sequence.

Characteristics of the Sequence

We now show some important properties of the sequence of offer profiles, the matchings, the Hungarian forests, and the equality subgraphs. In particular, we show that at all times the offer profile is stable and the offer on the node \( i \) never changes. Thus the first two test conditions are met by all offer profiles in the sequence. We also show the monotonicity of the sequence of offer profiles and the topological structure of the sequence of equality graphs and the Hungarian forests.

We first observe that the offer profile \( O^t \) is stable at any time \( t \) as shown in the following proposition.

**Proposition 9.** At any time \( t \), the offer profile \( O^t \) is stable.

**Proof:**

We will prove this proposition by induction. From the construction, we know that the offer profile is stable at \( t = 0 \).

---

\(^8\)Since \( O^t \) is a SASTGOP within the ASTAN \( S_{\mathcal{X}_t} \), therefore there is a near-perfect matching that leaves \( i^t \) unmatched.

\(^9\)Clearly this is a maximum matching in \( EQ^{t+1} \) because there does not exist any other augmenting paths in \( EQ^{t+1} \).
Assume that for some \( t \geq 0 \), the offer profile \( O^t \) is stable. We will show that \( O^{t+1} \) is stable. At iteration \( t+1 \), if the offers change, then:

From the construction of \( O^{t+1} \), and by lemma 3 we know that:

- for all \( i' \in A \setminus A^T_t \), \( O^{t+1}_{i'} = O^t_{i'} \)
- for all \( j' \in B \setminus B^T_t \), \( O^{t+1}_{j'} = O^t_{j'} \)
- for all \( i' \in A^T_t \), \( O^{t+1}_{i'} < O^t_{i'} \)
- for all \( j' \in B^T_t \), \( O^{t+1}_{j'} > O^t_{j'} \)

The edges in \( E \) can be divided into four mutually exclusive subsets:

- \( E_1 = E \cap (A^T_t \times B^T_t) \)
- \( E_2 = E \cap ((A \setminus A^T_t) \times B^T_t) \)
- \( E_3 = E \cap (A^T_t \times (B \setminus B^T_t)) \)
- \( E_4 = E \cap ((A \setminus A^T_t) \times (B \setminus B^T_t)) \)

From the construction, since \( O^{t+1}_{|X^t_t} \) is a SASTGOP within the subgraph \( S_{|X^t_t} \), therefore,

\[
\text{for all } (i', j') \in E_1, O^{t+1}_{i'} \geq v_{i'}(j', O^{t+1}_{j'}) .
\]

Since \( O^t \) was stable therefore

\[
\text{for all } (i', j') \in E_4, O^{t+1}_{i'} = O^t_{i'} \geq v_{i'}(j', O^{t+1}_{j'}) = v_{i'}(j', O^{t+1}_{j'}) .
\]

From the construction of \( O^{t+1} \) and the stability of \( O^t \) and since the pareto payoff functions are strictly decreasing therefore

\[
\text{for all } (i', j') \in E_2, O^{t+1}_{i'} = O^t_{i'} \geq v_{i'}(j', O^{t+1}_{j'}) > v_{i'}(j', O^{t+1}_{j'}) .
\]
Since for all \( i' \in A_t \), \( O^{t+1}_{i'} \geq eo_{i'} \), therefore

for all \( (i, j) \in E3 \), \( O^{t+1}_{i'} \geq eo_{i'} \geq v_{i'} (j', O^{t+1}_{j'}) = v_{i'} (j', O^{t+1}_{j'}) \)

Therefore we see that

for all \( (i', j') \in E \), \( O^{t+1}_{i'} \geq v_{i'} (j', O^{t+1}_{j'}) \)

and hence the offer profile \( O^{t+1} \) is stable. By induction at any iteration \( t \geq 0 \), the offer profile \( O^t \) is stable.

In the following proposition, we show the monotonicity of the offer sequence and show some structural properties of the Hungarian forest and the equality subgraph at any time \( t \).

**Proposition 10.** At any time \( t \geq 0 \):

1. For all \( t' > t \) and all \( i' \in A \), \( O^t_{i'} \leq O^{t'}_{i'} \) and for all \( j' \in B \), \( O^t_{j'} \geq O^{t'}_{j'} \).

2. For any unmatched node \( j \in B \), \( O^t_j = O^0_j \).

3. For any \( j \in B \setminus B^T_t \) either \( O^t_j = O^0_j \) or \( j \) has an alternating path for the matching \( M^t \) in \( EQ^t \) from some \( j^\circ \in B \) such that \( O^t_{j^\circ} = O^0_{j^\circ} \).

4. For any \( j \in B \) either has \( O^t_j = O^0_j \), or \( j \) has a path in \( EQ^t \) to some \( j^\circ \in B \) such that \( O^t_{j^\circ} = O^0_{j^\circ} \).

**Proof:**

Since at any time \( t \), for any expanding node \( i' \in C^T_t \), the expanding offer \( eo_{i'} < O^t_{i'} \), therefore by lemma 3 for all \( i' \in A^T_t \), \( O^t_{i'} < O^{t-1}_{i'} \) and for all \( i' \in A \setminus A^T_t \), \( O^t_{i'} = O^{t-1}_{i'} \). Also, for all \( j' \in B^T_t \), \( O^t_{j'} > O^{t-1}_{j'} \) and for all \( j' \in B \setminus B^T_t \), \( O^t_{j'} = O^{t-1}_{j'} \). Hence the first claim holds.

Clearly, if the node \( j \in B \) is unmatched, it must not have been any alternating tree until time \( t \). Hence \( O^t_j = O^0_j \)
We now prove the third claim. First we note that $B \setminus B^{T^t} = (B^{F^t} \setminus B^{T^t}) \cup (B \setminus B^{F^t})$.

If $j \in B^{F^t} \setminus B^{T^t}$, i.e. $j$ is in the Hungarian forest but not in the alternating tree $T^t$ at time $t$, then $O^t_j = O^0_j$. To verify this, assume that $j \in T^{t_k}_i$ for some $1 < k \leq m^t$. Then all nodes $i' \in A^{T^t_k}$ have $O^t_{i'} = O^0_{i'}$, and therefore by stability of $O^0$ and claim 1, $O^t_j = O^0_j$.

If a node $j \in B \setminus B^{F^t}$, i.e. $j$ is not in the Hungarian forest, then either $O^t_j = O^0_j$ or $j$ has a path to some node $j^o$ for which $O^t_{j^o} = O^0_{j^o}$. For nodes $j \in B^{F^t}$, we will prove the claim by induction. Clearly, the claim holds for $t = 0$. Assume the claim holds for some $t \geq 0$. At time $t+1$, the alternating tree connects to one of the joining nodes $j' \in B \setminus B^{F^t}$. By claim 3 $j'$ has a path to some $j^o$ with $O^t_{j^o} = O^0_{j^o} = O^0_{j^o}$ and hence $j$ has a path to a node $j^o$ for which $O^{t+1}_{j^o} = O^0_{j^o}$. Hence by induction the claim holds.

In the next proposition we characterize the position of the node $i$ in the equality subgraph at any time $t$. The consequence of the proposition is that the offer on node $i$ never changes and the node $i$ appears in the Hungarian forest, if and only if the sequence has converged to a SASTGOP.

**Proposition 11.** At any time $t \geq 0$, the following hold:

1. If $i$ belongs to an alternating tree $T = T^t_k$ for some $k \leq m_t$, then $D^T = \phi$.

**Proof:**
Assume that at time $t$, $i$ belongs to the alternating tree $T = T_{i,k}$ for some $k \leq m_t$ and $D^t \neq \emptyset$. Pick $i' \in C^t$ and $j' \in Nbr (i') \cap D^t$. By proposition 10, $j'$ is reachable by an alternating path $P_{j',j^o}$ from some $j^o$ with $O^t_{j^o} = O^0_{j^o}$ in $EQ^t$. By stability of $O^t_i$, and claim 1 of proposition 10, all neighbors of $j^o$ in $EQ^0$ are also neighbors of $j^o$ in $EQ^t$. Pick a neighbor $i^o$ of $j^o$ in $EQ^0$. From the assumption, no node in $X^{P_{j^o} \cup \{i^o\}}$ belongs to the alternating tree $T$ or else $j$ will be in the alternating tree $T$ as well. Consider the paths $P_{i,j^o} \subseteq EQ^t, P_{j^o,j^o} \subseteq EQ^t$ and the path $P_{i,j^o} = P_{i,i'} \cup \{i',j'\} \cup P_{j^o,j^o} \cup \{j^o,i^o\}$ from $i$ to $i^o$. The offer induced on $i^o$ by the path $P_{i,j^o}$ for the offer $o_i$ on $i$ is

\[
f^{P_{i,j^o}}_{i,i'} (o_i) \\
= v^o_i \left( j^o, f_{i,i'}^{P_{i,j^o}} (v_{j^o} (i', f_{i,i'}^{P_{i,j^o}} (o_i)) \right) \\
> v^o_i \left( j^o, f_{j^o,j^o}^{P_{i,j^o}} (O^t_{j^o}) \right) \text{ since by stability of } O^t, O^t_{j^o} > v^o_{j^o} (i', O^t_{j^o}) = v^o_{j^o} \left( i', f_{i,i'}^{P_{i,j^o}} (o_i) \right) \\
= O^t_{j^o} \\
= O^0_{j^o}
\]

This is a contradiction because by the definition of $O^0, O^0_{j^o}$ is the maximum path induced offer on $i^o$ for the offer $o_i$ on $i$. Hence by contradiction, either $i$ does not belong to the alternating tree $T$ or $D^t = \emptyset$.

There are two important corollaries of proposition 11. The first corollary provides a simplified test for convergence of the sequence of offer profiles to a SASTGOP. The second corollary guarantees that the offer on node $i$ never changes.

**Corollary 4.** At any time $t$, the node $i$ belongs to the Hungarian forest if and only if the Hungarian forest has only one alternating tree $T^t$ that spans all nodes in $X$.

**Proof:** First assume that the Hungarian forest has only one alternating tree $T^t$ and it contains $i$, then by proposition 11, $D^{T^t} = \emptyset$. If $A^{T^t} \neq A$ then $|Nbr^E (A^{T^t})| = |B^{T^t}| + |D^{T^t}| = |A^{T^t}| - 1 + |D^{T^t}| = |A^{T^t}| - 1 < |A^{T^t}|$ which contradicts lemma 1. Therefore $A^{T^t} = A$ and $|B^{T^t}| = |A^{T^t}| - 1 = |A| - 1 = |B|$. Therefore $B^{T^t} = B$. Therefore $T^t$ spans all nodes in $X$. The other direction is trivial; since the Hungarian forest is an
alternating spanning tree therefore it must contain the node $i$.

**Corollary 5.** $O^t_i = o_i$ for all $t \geq 0$.

**Proof:** Following corollary 4, since the node $i$ is never in the Hungarian forest until there is only one alternating spanning tree in the Hungarian forest, therefore the offer on node $i$ never changes up to the time when all the test conditions are satisfied and the sequence has converged to a SASTGOP.

We now show that the sequence of offer profiles converges to a SASTGOP.

**Convergence**

**Proposition 12.** If at some time $t$, $m^t > 1$, then at some finite time $t^o > t$, the matching increases by 1.

**Proof:** The root of the alternating tree $T^t$ is $i_1^t$. At each time $t' > t$, until $i_1^t$ is matched, the alternating tree $T^{t'-1}$ has the root $i_1^{t'-1} = i_1^t$ and one of the following happens:

1. $T^{t'-1}$ connects to a joining node $j \in D^{T^{t'-1}}$ such that $j$ has an alternating path to an unmatched node $j^o$.
2. $T^{t'-1}$ connects to a joining node $j \in D^{T^{t'-1}}$ such that $j$ has an alternating path to $i$.
3. $T^{t'-1}$ connects to joining nodes $J \subseteq D^{T^{t'-1}}$ such that no nodes in $J$ have an alternating path to either an unmatched node or to $i$.

In the third case, $B^{T^{t'-1}} \subset B^{T^{t'}}$. Thus the alternating tree rooted at $i_1^t$ increases. Since $B$ is finite, therefore the third case happens only finitely many times. Therefore, at some finite time $t^o$, either of the first two cases happen. If the first case happens at $t^o - 1$, then there is an augmenting path from $i_1^t$ to some unmatched node $j^o$. Thus, by construction of matching $M^{t^o}$ at time $t^o$, $i_1^t$ is matched in $M^{t^o}$ and the matching increases by 1. If the second case happens at time $t^o - 1$, then it will contradict proposition 11 unless the first case happens along with the second case at time $t^o - 1$. Hence at time $t^o$, $i_1^t$ is matched in $M^{t^o}$ and the matching increases by 1.
Proposition 13. At some finite time $t^* \geq 0$, the offer profile $O^{t^*}$ is a SASTGOP with $O^{t^*}_i = o_i$.

Proof: By repeated application of proposition 12, we see that at some finite $t^\circ$, there is only one alternating tree in the Hungarian forest. At any time $t' > t^\circ$, unless $T^{t'-1}$ spans all nodes in $X$, the alternating tree $T^{t'-1}$ has the root $i_1^{t^\circ} = i_1^{t'-1}$ and one of the following happens:

1. $T^{t'-1}$ connects to a joining node $j \in D^{t'-1}$ such that $j$ has an alternating path to $i$.

2. $T^{t'-1}$ connects to joining nodes $J \subseteq D^{t'-1}$ such that no nodes in $J$ have an alternating path to $i$.

In the second case, $B^{T^{t'-1}} \subset B^{t'}$. Thus the alternating tree rooted at $i_1^{t^\circ}$ increases. Since $B$ is finite, therefore the second case happens only finitely many times. Therefore, at some finite time $t^* - 1$, the first case happens. By proposition 4 at time $t^*$, the alternating tree $T^{t^*}$ spans all nodes in $X$. Also by proposition 9, the offer profile $O^{t^*}$ is stable and by corollary 5 $O^{t^*}_i = o_i$. Therefore, at some finite time $t^*$, the offer profile $O^{t^*}$ is a SASTGOP and $O^{t^*}_i = o_i$.

By proposition 13, in finite time the sequence $O^t$ converges to a SASTGOP $O^*$ for $|B| = n$ with $O^*_i = o_i$. We now give the proof of the lemma 4.

Proof of Lemma 4

Proof: First we note that since there exists a near-perfect matching with an alternating spanning tree in $S$, therefore by lemma 1, every $A' \subset A$ has at least $|A'|$ neighbors in $B$ and every $B' \subseteq B$ has at least $|B'| + 1$ neighbors in $A$.

We will prove the lemma by induction. The lemma is trivially true for the case when $|B| = 0$. For $|B| = 1$, let $A = \{a_1, a_2\}$ and $B = \{b\}$. The only possible edge set for which this bipartite network is connected is $E = \{(a_1, b), (a_2, b)\}$. The maximum matching $M^\circ = \{(a_1, b)\}$ in this network is a near-perfect matching and it
induces an alternating spanning tree $T^{M^o}$. Without loss of generality, pick $o_{a_1} \in \mathbb{R}$ and set $O'$ as $O'_{o_{a_1}} = o_{a_1}$, $O'_b = v_b(a_1, o_{a_1})$ and $O'_{a_2} = v_{a_2}(b, O'_b)$. Then clearly, $T^{M^o}(O') = T^{M^o}$. Hence the lemma is true when $|B| = 1$.

Now assume that the lemma is true for all $0 \leq |B| < n$. Pick $i \in A$ and $o_i \in \mathbb{R}$. Then following the construction in the previous subsection, we have a sequence that converges to a SASTGOP $O'$ with $O'_{o_i} = o_i$ in finite time. By induction, the lemma is proved.
2.17 Appendix: Examples

In this section, we demonstrate through examples, the workings on the Hungarian algorithm, its limitations and the workings of the Generalized Hungarian Algorithm. First, in figure 2-1, we demonstrate the Hungarian algorithm to find the generalized stable matching in the standard weighted matching problem. In this case the algorithm converges in two iterations and gives the A-maximum generalized stable matching. Next, in figures 2-2, 2-3 and 2-2 we show that the algorithm can be run with changing offers only for one alternating tree at a time as opposed to the whole Hungarian forest. This takes longer to converge but still gives the A-maximum generalized stable matching. We then give an example in figure 2-5 of the generalized stable matching problem and in figure 2-6 show how the Hungarian algorithm fails. For simplicity, we assume that the payoffs of all but one node are equal to the splits. One node derives a non-linear payoff from the split. The payoff function of the odd node looks like a triangle wave going up a ramp with increasing frequency. Finally, in figure 2-7, we show the workings of the Generalized Hungarian algorithm for the same problem and show how the algorithm deals with the limitations of the standard Hungarian algorithm. Note that the payoff function chosen is continuous and strictly increasing but is not differentiable.
Bipartite Network

Legend

Iteration 1 with offers $O_{A1} = 3$, $O_{A2} = O_{A3} = O_{A4} = O_{A5} = 2$, $O_{B1} = O_{B2} = O_{B3} = O_{B4} = O_{B5} = 0$

Hungarian forest includes three alternating trees. $\Delta = \min\{O_i + O_j - w(i,j) : i \in A \text{ nodes in the Hungarian Forest}, j \in B \text{ nodes outside the Hungarian forest}\} = 1$

Iteration 2 with offers $O_{A1} = 2$, $O_{A2} = O_{A3} = O_{A4} = O_{A5} = 1$, $O_{B1} = O_{B3} = O_{B5} = 0$, $O_{B2} = O_{B4} = 1$

Figure 2-1: Running the Hungarian Algorithm to find the stable matching in the standard weighted matching problem
Iteration 1 with offers \( O_{A1} = 3, O_{A2} = O_{A3} = O_{A4} = O_{A5} = 2, O_{B1} = O_{B2} = O_{B3} = O_{B4} = O_{B5} = 0 \)

Hungarian forest includes three alternating trees. \( \Delta = \min\{O_i + O_j - w(i,j) : i \in A \text{ nodes in the first alternating tree}, j \in B \text{ nodes outside the first alternating tree} = 1\} \)

**Figure 2-2:** Running the Hungarian Algorithm to find the stable matching in the standard weighted matching problem: Iteration 1
Iteration 2 with offers $O_{A1} = 2$, $O_{A2} = 1$, $O_{A3} = O_{A4} = O_{A5} = 2$, $O_{B1} = O_{B3} = O_{B5} = 0$, $O_{B2} = 1$

Hungarian forest includes two alternating trees. $\Delta = \min \{O_i + O_j - w(i,j) : i \in A \text{ nodes in the first alternating tree, } j \in B \text{ nodes outside the first alternating tree} \} = 1$

**Figure 2-3:** Running the Hungarian Algorithm to find the stable matching in the standard weighted matching problem, only the offers on one of the alternating trees is updated: Iteration 2
Iteration 3 with offers $O_{A1} = 2, O_{A2} = O_{A3} = 1, O_{A4} = O_{A5} = 2, O_{B1} = O_{B3} = O_{B3} = O_{B5} = 0, O_{B2} = 1$.

Hungarian forest includes one alternating tree. $\Delta = \min\{O_i + O_j - w(i,j) : i \in A \text{ nodes in the first alternating tree}, j \in B \text{ nodes outside the first alternating tree}\} = 1$

Iteration 4 with offers $O_{A1} = 2, O_{A2} = O_{A3} = O_{A4} = O_{A5} = 1, O_{B1} = O_{B3} = O_{B5} = 0, O_{B2} = O_{B4} = 1$

**Figure 2-4:** Running the Hungarian Algorithm to find the stable matching in the standard weighted matching problem, only the offers on one of the alternating trees is updated: Iterations 3 and 4.
The triangle wave with linear drift and increasing frequency represents the payoff function of node A3 when it gets a split on the edge (A3, B3)
Iteration 1 with offers $O_{A1} = O_{A3} = 1$, $O_{A2} = 1.01$, $O_{B1} = O_{B2} = O_{B3} = 0$.

Hungarian forest includes one alternating tree. $\Delta = \min\{O_i + O_j - w(i,j) : i \in A \text{ nodes in the first, } j \in B \text{ nodes outside the first alternating tree} = 1\}$

Reducing the offers on A2 and A3 by $\Delta$ and increasing the offer on B3 by $\Delta$ does not satisfy the stability criterion on the edge (A3, B3).

**Figure 2-6:** The Hungarian Algorithm does not work for the generalized stable matching problem.
Updating the offers on the ASTAN, induced by the nodes A2, A3 and B3, to the desired SASTGOP ($O_{A2} = 0.9981$, $O_{A3} = 0.99$, $O_{B3} = 0.0119$) preserves the stability on the edge (A3, B3)

Iteration 2 with offers $O_{A1} = 1$, $O_{A2} = 0.9981$, $O_{A3} = 0.99$, $O_{B1} = O_{B2} = 0$, $O_{B3} = 0.0119$ preserves the stability on the edge (A3, B3)

From the alternating tree at iteration 2 (top) to the new alternating tree with the new offers. The updated SASTGOP ($O_{A1} = 0.25$, $O_{A2} = 0.23$, $O_{A3} = 0.33$, $O_{B1} = 0$, $O_{B2} = 0.75$, $O_{B3} = 0.78$) does not preserve the topology of the alternating tree but preserves the size of the matching and the new alternating tree spans all the nodes in the ASTAN

Iteration 3 with offers $O_{A1} = 0.25$, $O_{A2} = 0.23$, $O_{A3} = 0.33$, $O_{B1} = 0$, $O_{B2} = 0.75$, $O_{B3} = 0.78$. All nodes are matched

**Figure 2-7:** The Generalized Hungarian Algorithm find the A-maximum generalized stable matching for the generalized stable matching problem
Chapter 3

Fairness in Bipartite Matching With Transfers

3.1 Introduction

In this chapter, we study fairness of matching with wages in bipartite networks. We study the special case of costless transfers because we want to focus on the axiomatic characterization of fairness without the complications arising from the geometry of the core in the case of costly transfers. We want to establish the notion of fairness in bipartite matchings as the Nash bargaining solution does for the two agent bargaining case. In this sense our model is similar to [100].

"What is a fair outcome in bipartite matching?" The question has been addressed with respect to procedural fairness [92] and is important both for centralized assignments such as the national residency matching and distributed matching such and job matching and marriage markets. Laboratory experiments have also demonstrated that when people play distributed matching games, the outcome is often fair as compared to optimal for one side of the network [27]. In the absence of transfers among matched agents, median stable matching [100][109] has been proposed as a procedurally fair matching. In [109], the authors showed the existence of a median stable matching in the absence of transfers. In [100], the authors extended the result of [109] by a continuity argument and showed that a median stable matching exists for
job matching with wages. The authors in [100], also showed the existence of a mean stable matching.

Matching literature does not provided an axiomatic characterization of fair outcomes as Nash bargaining solution [82], or Kalai-Smorodinsky solution [56] does for the two agent bargaining. Extension of Nash bargaining to coalition games have given rise to the concepts of Shapley value [101] and coalition Nash bargaining solution [32] for coalition games and Myerson value for coalition games with coalitional constraints imposed by a graph [81]. However, matching has even stronger constraints on transfers than general coalition games that the transfers can only be made between the matched pairs. In [34], the authors presented a concept of fairness for matching in general networks. They called the concept, balancedness, and based it upon Nash bargaining solution. In a balanced outcome, the extra surplus over any matched pair’s outside options in the network is equally divided between the pair. It turns out that in bipartite matching situations, the balanced outcomes are equivalent to symmetric pairwise bargaining solutions [95] and is the intersection of the kernel and the core. In [65], the authors provided an algorithm for computing the balanced outcomes for general networks. The problem with balanced outcomes is that it is not unique and since it is a local characterization, it need not be globally fair and may not be even procedurally fair. On the other hand, the procedurally fair outcome defined by median stable matching [100] may not even be locally fair or balanced.

We introduce a notion of fairness in bipartite matching with wages and show that there exist stable and fair matching, meaning, no one or no group of agents exploits anyone or another group of agents, and the excess surplus generated by all matched pairs is equally divided between the matched agents.

Stability of matching is sustained by the competition among agents on the same side of the network. Symmetric pairwise bargaining outcomes is sustained by competition among agents on the same side along with bargaining between matched agents with the threat provided by the condition of the rest of the agents. A fair outcome is sustained by both the cooperation and competition among agents on the same side of the network. We use the example of matching between employees and employers.
While competition for jobs among employees brings down the wages in the market, the collective bargaining by the hired employees can increase the wages in the market. The question of fairness involves wages rather than matching because a stable matching can support all possible stable wages. A group of employees can bargain collectively with their employees for the common good. The common good in this case is the minimum wage for each employee in the group above its best option outside of this group of employers. This common good is suitable for collective bargaining as it is free from any problem of externalities within the group. An increase in the common wage, does not decrease the wage of any employee. A group of employers can also collectively bargain with the employees. The common good is the minimum extra surplus obtained by each employer above its option outside of this set of workers. The fair outcome is such that for any subset of employees and their respective employers, the Nash product of the minimum difference between the wage and the option of the employees outside the subset and the minimum difference between the option outside the set and wage given by the employers is maximized over all wages that are stable.

We introduce three results related to the definition of fairness.

- There exists a fair matching and wages that is also stable.

- The fair matching is unique.

- The fair matching coincides with the nucleolus of the bipartite matching.

We also introduce a polynomial time poset labeling algorithm to find the fair matching and wages. This algorithm extends the poset labeling algorithm [65].

The organization of the rest of the chapter is as follows. In section 3.2, we introduce the model and the definition of fair outcomes. In section 3.3, we present the algorithm for computing the fair outcome in any bipartite network. In section 3.4, we present our main results. We give conclusions and directions of future work in section 3.5 and in the rest of the chapter, we give the proofs of our main results with additional intermediate results.
3.2 Model

We define $A$ as the set of workers and $B$ the set of jobs or employers, $X = A \cup B$ is the set of all agents. A weighted bipartite network connects workers to jobs, $S = (X, E, w)$, where $E$ is the set of edges and $w : E \to \mathbb{R}^+$ is a function over the edges and represents the surplus that the worker can create on the job. The set of neighbors of a agent $i \in X$ is $\text{Nbr}^S(i) = \{ j \in X : (i, j) \in E \}$. We will represent a matching of workers to jobs by the matching function $M : X \to X$, where $M(i)$ is the match for $i$. The wages will be represented by the vector $t$. The wage for worker, $i$, will be represented by $t_i$. Each unmatched worker, $i$, gets a wage $t_i = 0$ and his surplus, $u_i = 0$ (without loss of generality, the reserved surplus are 0), and each matched worker, $i$, gets a surplus, $u_i = t_i$. The surplus of any employer from a job, $j$, that is matched to worker $i$, is $u_j = w(i, j) - t_i$ and from unmatched jobs is 0.

A matching is stable if all workers and jobs have a non-negative surplus and no worker can give a higher surplus for an employer while maintaining her wage, i.e.- $\forall (i, j) \in E, u_i + u_j \geq w(i, j)$. This is a standard definition of stable matching in job matching that is also used in [100].

3.2.1 Balancedness

We first discuss the concept of balancedness. This is a well studied concept derived from two agent Nash bargaining solution [34]. A polynomial time algorithm was given by [65] to compute the balanced outcomes. The balanced outcomes are only locally fair and not necessarily unique. It turns out that in bipartite matching situations, the balanced outcomes are equivalent to symmetric pairwise bargaining solutions [95] and coincides with the intersection of the kernel and the core.

The outside option for a worker or a job, $i$, when matched with $j$ is

$$o_i = \max \{ \max_{k \in \text{Nbr}(i)\backslash\{j\}} \{ w(i, k) - u_k \}, 0 \}$$

The extra surplus or slack of a worker or a job from the match is $\sigma_i = u_i - o_i$. 
The slack of any edge, \((i, j)\), between a worker, \(i\) and a job, \(j\), not matched to each other is \(\sigma_{ij} = u_i + u_j - w(i, j)\).

A matching \(M\) with wages \(t\) is balanced \([65]\) if it is stable and

\[
\text{for all } (i, j) \in M, \sigma_i = \sigma_j
\]

### 3.2.2 Fairness

We now introduce a new concept of fairness that extends the two person Nash bargaining solution to collective bargaining case. Given any subset of matched workers \(A^o \subseteq A\) and the set of jobs, \(B^o = M(A^o) \subseteq B\), the option of any worker, \(i \in A^o\) outside \(B^o\) is

\[
o_i(B^o) = \max \left\{ \max_{k \in \text{Nbr}(i) \setminus B^o} \{w(k, i) - u_k\}, 0 \right\}
\]

Similarly, the option of any job, \(j \in B^o\) outside \(A^o\) is

\[
o_j(A^o) = \max \left\{ \max_{k \in \text{Nbr}(j) \setminus A^o} \{w(k, j) - u_k\}, 0 \right\}
\]

The extra surplus of a worker \(i \in A^o\) from the match over the options outside \(B^o\), or slack outside \(B^o\) is \(\sigma_i(B^o) = u_i - o_i(B^o)\).

The extra surplus or slack of a job \(j \in B^o\) from the match over the options outside \(A^o\), or slack outside \(A^o\) is \(\sigma_j(B^o) = u_j - o_j(A^o)\).

A matching \(M\) with wages \(t\) is a fair outcome if it is stable and for all subsets of matched workers \(A^o\) and their matched jobs \(B^o = M(A^o) \subseteq B\),

\[
\text{for all } i \in A^o, \sigma_i(B^o) \geq \min_{j \in B^o} \{\sigma_j(A^o)\} \text{ for all } j \in B^o, \sigma_j(A^o) \geq \min_{i \in A^o} \{\sigma_j(B^o)\}
\]

The above conditions are equivalent to the following two conditions.

\[
\min_{i \in A^o} \sigma_i(B^o) = \min_{j \in B^o} \{\sigma_j(A^o)\}
\]

or \(t|_{A^o \cup B^o}\), the restriction to wages on the set \(A^o \cup B^o\), maximizes the collective
Nash product, \( \min_{i \in A^\circ} \{ \sigma_i (B^\circ) \} \min_{j \in B^\circ} \{ \sigma_j (A^\circ) \} \).

While the workers are competing for jobs and the employers are competing for workers, they can still cooperate and provide a collective bargaining. This is because, an increase in the wage for any workers weakly improves the outside options of other workers and similarly, a decrease in the wage on any job, weekly decreases the outside options for all employers. This increase in the outside options for workers and employers gives them additional advantage in bargaining with their matches and hence a higher surplus.

Assume a coalition of workers are collectively bargaining with the coalition of employers they are matched with. The threat for collective bargaining for workers is the minimum decrease in wage at which one worker quits her job to take up a job outside the coalition of jobs. This is because, at any higher wage, no worker can be attractive to any employer outside of the coalition and the employers in the coalition will not hire any worker they are collectively bargaining with. Then the coalition of jobs will break as there will be one employer who will not be able to compete. Similarly, the employers have a counter threat that is the minimum increase in wage at which one of the employers will fire a worker and hire another worker. The coalition of workers will break at that wage because one worker will not be able to compete.

### 3.3 Algorithm

Since any maximum weight matching admits all possible stable wages, [41] therefore we pick an arbitrary maximum weight matching and find fair wages corresponding to that matching. It is easy to check that the wages that are fair for one maximum weight matching are also fair for all maximum weight matchings.

**Input:** A bipartite network \( S \).

**Output:** A fair and stable matching and wages, \( M^*, t^* \).

**Initialization:** Pick any arbitrary maximum weight matching \( M^* \) and arbitrary stable wages, \( t \). We will construct fair wages, \( t^* \) from \( t \). We will refer to any worker or job as *anchored*, if we have set its fair wage. Clearly, the wages for all unmatched
workers and jobs is set to 0. Using the Edmond-Gallais decomposition, partition the workers and jobs in $X$ into:

1. $P$: the set of workers and jobs that are not matched in all maximum weight matchings. Set the wages for all workers in $P$ to 0 and the wages on any job $j \in P$ equal to the surplus generated by its matched worker, $w(M^*(j), j)$, which is 0 if the job is not matched.

2. $D$: the set of workers and jobs that are neighbors of $P$ in the bipartite network.

3. $C$: $X \setminus (P \cup D)$.

Clearly, the matches of anchored workers and jobs in $M^*$ are also anchored. We will refer to the set of anchored workers and jobs as $\psi \subseteq X$. The wages are a combination of set and original wages, $t^* = t^*_{|\psi} \cup t_{|X \setminus \psi}$. We will refer to maximum slack of the anchored set, $\sigma_{\psi} = \max_{i \in \psi} \{\sigma_i\}$.

**Iteration:**

Until all workers and jobs are anchored, in any general iteration of the algorithm, given the anchored set $\psi$, we find an alternating path (path with edges alternating between matching and non matching edges) or an alternating cycle (even cycle with edges alternating between matching and non matching edges), that has the minimum slack. We only need to consider the alternating paths and cycles that have all degree two workers and jobs not anchored and include the matches of the end agents that are unanchored. The slacks for the alternating paths and cycles are computed as follows:

For any alternating path or cycle, $Z$, we define $\phi(Z)$ as the difference between the weights of matching and non matching edges and $n(Z)$ as the number of non matching edges.

- The slack for an alternating path with both ends unanchored is $\sigma_Z = \frac{\phi(Z)}{n(Z)+2}$.
- The slack for an alternating path with one anchored end, $i$, is $\sigma_Z = \frac{\phi(Z)-u_i}{n(Z)+1}$.
- The slack for an alternating path with both ends, $i, j$, anchored is $\sigma_Z = \frac{\phi(Z)-u_i-u_j}{n(Z)}$. 
• The slack for an alternating cycle is $\sigma_Z = \frac{s(Z)}{m(Z)}$.

The slacks for the alternating paths and cycles are the slacks on the unmatched edges in those structures. Once we have found the structure that has the minimum slack over all such structures, we proceed as follows. If the structure is an alternating path, the wages for all workers in the structure are uniquely identified, given its slack as follows:

• If the end agent is an unanchored worker, its wage is set equal to the slack of the structure.

• If the end agent is an unanchored job, the wage to the worker matched to the job is set equal to the weight of the matching edge minus the slack.

• The wages of the workers and for the jobs inside the path are then computed by spreading from the end agents, keeping the slacks on the unmatched edges equal to the slack of the structure.

If the structure is an alternating cycle, the fair wages for the workers and the jobs is not determined given the slack. The slack determines a range of balanced wages for the workers and the jobs in the cycle and setting wage for any worker will uniquely determine the wages for all other workers and jobs. The maximum of all balanced wages, $t^{max}(Z)$, for the alternating cycle is the balanced wage such that at least one of the workers gets a wage equal to the surplus of his job. The minimum of all balanced wages, $t^{min}(Z)$, is the balanced wage such that at least one of the workers gets a wage 0. All other balanced wages are convex combinations of the maximum and minimum balanced wage. One of these balanced wages will correspond to the fair wage for the workers and jobs in the alternating cycle. To proceed, we collapse the alternating cycle into a matched pair of dummy worker, $i^Z$ and a dummy job, $j^Z$, and restructure the network as described in the collapse method below. We proceed the iterations with this smaller network.

**Collapse:**
3.3. ALGORITHM

- The weight of the matched edge between the dummy agents is equal to the difference between the maximum and the minimum balanced wages for the alternating cycle.

- All edges between any job, \( j \), outside the alternating cycle and the workers in the alternating cycle are replaced with one edge, whose weight is

\[
w(i^Z, j) = \max\{\max_{i \in Z}\{w(i, j) - t_{i}^{\min}(Z)\}, 0\}
\]

- All edges between any worker, \( i \), outside the alternating cycle and jobs in the alternating cycle are replaced with one edge, whose weight is

\[
w(i, j^Z) = \max\{\max_{j \in Z}\{w(i, j) - w(M(j), j) + t_{M(j)}^{\max}(Z)\}, 0\}
\]

When all the agents are anchored the fair wages for the collapsed network is available. Then we recursively expand the collapsed networks to obtain the fair outcome in the original network.

**Expand:** Once the fair wages for the smaller network is available the network is expanded back and fair wages in the collapsed alternating cycle are computed as follows:

- The fair wage for any worker, \( i \), in the alternating cycle is the sum of the fair wage of the dummy agent, \( i^Z \), in the smaller network and \( i \)'s minimum balanced wage, \( t_{i}^{\min}(Z) \) in the alternating cycle, i.e.- \( t_i = t_{i^Z} + t_{i}^{\min}(Z) \).

- The fair wage for the jobs in the alternating cycle is the wage of the worker matched to the job.

We note that removing collapse and expand methods will reduce the algorithm to the poset labeling method in [65]. In [65], the authors showed that slack of the structure found in an earlier iteration is lower than the slack of the structure found in a later iteration. Following the same argument, one can see that in our algorithm as well, the same monotonicity holds.
3.4 Main Results

We give three main results about the collective bargaining outcome and prove them constructively using the algorithm. These results also prove the correctness of the algorithm. The first result tells us that a fair outcome exists for all bipartite networks and the algorithm finds the fair outcome. This also shows the correctness of the algorithm.

**Theorem 5.** *The algorithm finds a fair outcome in any bipartite network.*

In the proof of the above result, we show in proposition 15 that fair outcomes are balanced and since all balanced outcomes are in the core [95][65], therefore fair outcomes are in the core. The second result shows the uniqueness of the fair outcome. We will prove this result using the algorithm as well.

**Theorem 6.** *There is a unique fair outcome in any bipartite network.*

The third result establishes the equivalence of the fair outcomes and the nucleolus. Nucleolus is the lexicographic center of the core [80]. The consequence of this result is that we have found new axiomatic characterization of the nucleolus of bipartite matching.

**Theorem 7.** *The fair outcome coincides with the nucleolus.*

3.5 Conclusions and Future Work

We gave an axiomatic characterization of fair outcomes in bipartite matching under costless transfers based upon the collective bargaining argument. We also presented an algorithm to compute the fair outcome and showed that the fair outcome coincides with the unique nucleolus of the assignment game. In future, we want to study two things. Firstly, we want to study distributed dynamics that convert to fair outcomes. Secondly, we want to extend our results to the general case of costly transfers.
3.6 Proof of Theorem 5

We first give intermediate results that will be used in the proof of the main theorem.

**Proposition 14.** If only alternating paths are found as structures with minimum slack at each iteration, then the algorithm finds the unique balanced outcome.

**Proof:**

When only alternating paths are found at each iteration, the algorithm reduces to the Klienberg-Tardos algorithm [65] for finding balanced outcomes and the balanced outcome is unique following the Aspvall-Shiloach characterization [8]. Therefore the algorithm finds the unique balanced outcome.

**Proposition 15.** Any fair outcome is a balanced outcome.

**Proof:**

Pick a fair outcome \( M, t^* \) in a bipartite network \( S \). Pick any matched worker \( i \) and its matched job \( M(i) \). \( \sigma_i = \sigma_i(M(i)) \) and \( \sigma_{M(i)} = \sigma_{M(i)}(\{i\}) \). Since \( t^* \) is fair, therefore \( \sigma_i(M(i)) = \sigma_{M(i)}(\{i\}) \). Therefore \( \sigma_i = \sigma_{M(i)} \). Since \( i \) and \( j \) were arbitrarily picked, therefore it holds of all workers and their matched jobs. Therefore \( M, t^* \) is a balanced outcome.

**Lemma 6.** If the bipartite network has a unique balanced outcome then it is also the fair outcome.

**Proof:** Assume there is a unique balanced outcome, \( M, t^* \), in the bipartite network \( S \). Pick a subset of matched workers, \( A^o \) and their matching jobs \( B^o = M(A^o) \). Pick any maximal subset \( A' \subseteq A^o \) and their matched jobs \( B' \) such that all agents in \( A' \cup B' \) were anchored earliest among the agents in \( A^o \cup B^o \) and all agents in \( A' \cup B' \) were contiguous agents in the slack minimizing alternating path, in the iteration, they were anchored. Without loss of generality, let us represent the agents as a sequence, \( i_0, j_0, i_1, j_1, ..., i_n, j_n \) according to their position in the alternating path. By construction we know that the slack for all agents in \( i_0, j_0, i_1, j_1, ..., i_n, j_n \) is the same and the best option for the end agents is outside the alternating path and the best option for
any interior agents is the neighbor it is not matched to. Therefore the slack is minimum for \(i_0\) and \(j_n\) among all agents in \(i_0, j_0, i_1, j_1, ..., i_n, j_n\). Since, \(i_0, j_0, i_1, j_1, ..., i_n, j_n\) were the first agents in \(A^o \cup B^o\) to be anchored, they have the minimum slack among all agents in \(A^o \cup B^o\) and the best options for \(i_0\) and \(j_n\) are outside of \(A^o \cup B^o\). This implies that \(\min_{i \in A^o} \sigma_i (B^o) = \sigma_{i_0} = \min_{j \in B^o} \sigma_j (A^o)\). Since, \(A^o\) was chosen arbitrarily, therefore it holds for all subsets of workers, \(A\). Therefore, \(M, t^*\) is a fair outcome in \(S\).

Lemma 7. If an alternating cycle, \(Z\), in the network, \(S\), was collapsed to form the smaller network, \(S^o\), then expanding the network using the fair outcome, \(t^* (S^o)\), in \(S^o\) gives the fair outcome, \(t^* (S)\) in \(S\).

Proof:

Pick any subset of workers \(A^o\) and their matched jobs \(B^o\) that satisfies at least one of the following:

1. One of the workers, \(i^o \in A^o\), with minimum slack outside \(B^o\) is in \(Z\), i.e.-
   \[i^o \in \arg \min_{i \in A^o} \sigma_i (B^o) \in Z\]

2. One of the jobs, \(j^o \in B^o\), with minimum slack outside \(A^o\) is in \(Z\), i.e. –
   \[j^o \in \arg \min_{j \in B^o} \sigma_i (A^o) \in Z\]

3. The worker, \(i' \in A^o\), with the minimum slack outside \(B^o\) has it’s best option outside \(B^o\) with a job, \(j^*\) in \(Z\).

4. The job, \(j' \in B^o\), with the minimum slack outside \(A^o\) has it’s best option outside \(B^o\) with a worker, \(i'\) in \(Z\).

Since, the surplus of all workers and jobs does not change in the expansion process, the above are the only subsets that we need to check for fairness. In the first case, if \(Z \subseteq A^o \cup B^o\), then \(i^Z\) must have been anchored before other agents in the smaller network. This implies that the second case is also satisfied because \(j^Z\) is anchored at the same iteration as \(i^Z\). Therefore the minimum slack over all jobs in \(B^o\) is for some job \(j^o \in Z\). From the balancedness \(i^Z\) has the same slack as \(j^Z\). Therefore,
\( \sigma_i^o (B^o) = \sigma_i^Z ((B^o \setminus Z) \cup \{j^Z\}) = \sigma_j^Z ((A^o \setminus Z) \cup \{i^Z\}) = \sigma_j^o (A^o) \). If all of \( Z \) is not in \( A^o \cup B^o \), then \( \sigma_i^o (B^o) \) is equal to the slack, \( \sigma_Z \), of \( Z \) because it is the minimum slack of all nodes in \( Z \). Therefore, clearly, the alternating cycle \( Z \) was collapsed before any other agent in \( A^o \cup B^o \) was anchored and therefore the minimum slack over all jobs in \( B^o \) is for some job \( j^o \in Z \) and it is equal to the slack, \( \sigma_Z \), of \( Z \). Therefore, \( \sigma_i^o (B^o) = \sigma_i^o = \sigma_Z = \sigma_j^o (A^o) \).

In the third case, \( \sigma_i'(B^o) = u_i' + u_j - w(i', j^*) = u_i' + u_j - w(i', j^Z) - w(M(j^*), j^*) + t_{\text{max}}^o(j^*, j^Z) = u_i' + u_j - w(i', j^Z) \)

which is the slack of \( i' \) outside of the \( (B^o \setminus Z) \cup \{i^Z, j^Z\} \) in the smaller network. This is the minimum slack over all workers in \( (B^o \setminus Z) \cup \{i^Z, j^Z\} \) in the smaller network. Similar, argument follows for the fourth case. Therefore, from the fairness of the outcome in the smaller network, \( \sigma_i'(B^o) = \sigma_i'( (B^o \setminus Z) \cup \{i^Z, j^Z\} ) = \min_{j \in (B^o \setminus Z) \cup \{i^Z, j^Z\}} \{ \sigma_j ((A^o \setminus Z) \cup \{i^Z, j^Z\}) \} = \min_{j \in B^o} \sigma_j (A^o) \).

Therefore, the expansion of the network gives a fair outcome in the expanded network.

We now complete the proof of the main theorem.

**Proof:**

When all the agents are anchored, any alternating paths that were encountered were collapsed. Therefore, following proposition 14, the network has a unique balanced outcome. Following lemma 6, the outcome is also fair. Iteratively expanding the network, maintains the fairness following lemma 7. By induction, when all collapsed alternating cycles are expanded, the outcome is fair for the original network.

### 3.7 Proof of Theorem 6

We first give intermediate results that will be used in the proof of the main result.

**Lemma 8.** If an alternating cycle, \( Z \), in the network, \( S \), was collapsed to form the smaller network, \( S^o \), and \( M, t^* (S^o) \) is the unique fair outcome in \( S^o \) then expanding the network using \( t^* (S^o) \) in \( S^o \) gives the unique fair outcome, \( t^* (S) \) in \( S \).
Proof:

Assume there is only one fair outcome in the smaller network and there is another fair outcome, $M, t^o$ in the expanded network. All fair outcomes are balanced and any balanced outcomes in $Z$ are uniquely determined by the wage of any one worker in $Z$ and it also uniquely determines the surplus for $i^Z$ and $j^Z$ in the smaller network. Define $t(S^o)$ as the restriction of the wages, $t$, on the agents in smaller network including $i^Z$. We will show that it implies that $t(S^o)$ is also fair. This will contradict that the smaller network has only one fair outcome.

Pick any subset of workers $A^o$ and their matched jobs $B^o$ in the smaller network that satisfies at least one of the following:

1. $i^Z \in \arg \min_{i \in A^o} \sigma_i (B^o)$
2. $j^Z \in \arg \min_{j \in B^o} \sigma_j (A^o)$
3. The worker, $i' \in A^o$ with the minimum slack outside $B^o$ has it’s best option outside $B^o$ with $j^Z$.
4. The job, $j' \in B^o$ with the minimum slack outside $A^o$ has it’s best option outside $B^o$ with $i^Z$.

Since, the surplus of workers and jobs does not change in the expansion process, the fairness condition is satisfied for all other subsets of workers and their matched jobs in the smaller network because it is satisfied in the larger network.

The slack, $\sigma_{i^Z} (B^o)$ in the smaller network is equal to $\min_{i \in Z \cap A} \sigma_i ((B^o \setminus \{i^Z\}) \cup (Z \cap B))$ in the larger network. Similarly, the slack, $\sigma_{j^Z} (A^o)$ in the smaller network is equal to $\min_{j \in Z \cap B} \sigma_j ((A^o \setminus \{i^Z\}) \cup (Z \cap A))$ in the larger network. In the first case, since the fairness condition is satisfied for the workers in $(A^o \setminus \{i^Z\}) \cup (Z \cap A)$ and jobs in $(B^o \setminus \{j^Z\}) \cup (Z \cap B)$, therefore

$$\sigma_{i^Z} (B^o) = \min_{i \in (A^o \setminus \{i^Z\}) \cup (Z \cap A)} \sigma_i ((B^o \setminus \{j^Z\}) \cup (Z \cap B))$$

$$= \min_{j \in (B^o \setminus \{j^Z\}) \cup (Z \cap B)} \sigma_j ((A^o \setminus \{i^Z\}) \cup (Z \cap A)) = \sigma_{j^Z} (A^o).$$

Therefore the fairness condition is satisfied for $A^o$ and $B^o$. Similarly, the fairness condition is satisfied for the second case as well.
In the third and fourth cases, clearly, \( i^Z \notin A^o \) and \( j^Z \notin B^o \). Therefore \( \sigma_{i'}(B^o) = \min_{j \in Z \cap B} \sigma_{i' j} \) and \( \sigma_{j'}(A^o) = \min_{i \in Z \cap A} \sigma_{i j'} \). Since, \( t \) is fair for the larger network, therefore, \( \sigma_{i'}(B^o) = \min_{j \in B \setminus B^o} \sigma_{i' j} = \min_{j \in B^o} \sigma_{j}(A^o) \) and \( \sigma_{j'}(A^o) = \min_{i \in A \setminus A^o} \sigma_{i j'} = \min_{i \in A^o} \sigma_{j} \). The second equality follows from the fairness in the larger network. Therefore, the fairness condition is satisfied for \( A^o \) and \( B^o \) in the smaller network.

Therefore in all the cases, the fairness condition is satisfied. This implies that \( M, t \) is a fair outcome in the smaller network which contradicts the assumption that the smaller network had a unique fair outcome.

We now complete the proof of the main theorem.

**Proof:**

When all the agents are anchored, any alternating paths that were encountered were collapsed. Therefore, following proposition 14 and 15, the network has a unique balanced outcome. Following lemma 6 and proposition , the outcome is also the unique fair outcome. Iteratively expanding the network, maintains the fairness following lemma 7 and uniqueness of the fair outcome following lemma 8. By induction, when all collapsed alternating cycles are expanded, the original network has a unique fair outcome.

### 3.8 Proof of Theorem 7

We first give intermediate results that will be used in the proof of the main result.

**Proposition 16.** *If the balanced outcome is unique, it is also the nucleolus.*

**Proof:**

The result follows from the definition of balanced outcomes. Since, it is unique, it is the lexicographic center of the core outcomes and therefore is the nucleolus.

**Lemma 9.** *If an alternating cycle, \( Z \), in the network, \( S \), was collapsed to form the smaller network, \( S^o \), and the fair outcome \( M, t^*(S^o) \) is the nucleolus in \( S^o \) then expanding the network using \( t^*(S^o) \) in \( S^o \) gives the fair outcome which is the nucleolus, \( t^*(S) \) in \( S \).*
Proof:

Assume that the fair outcome is the nucleolus in the smaller network. We want to show that the fair outcome in the larger network is also the nucleolus or lexicographically maximizes the slack on all agents in the larger network. We know that $t^*(S)$ maximizes the minimum slack over all workers and jobs. Since the nucleolus is also balanced therefore the slack, $\sigma_Z$ for the alternating cycle $Z$ in the nucleolus is same as the slack for $Z$ in any balanced outcome and the fair outcome. In the outcome $M$, $t^*(S^o)$, the slack on $iZ$ in the smaller network is equal to the minimum slack, $\min_{i \in Z \cap A} \sigma_i(Z \cap B)$ over all workers in $Z$ in the larger network in the outcome, $M$, $t^*(S)$. Similarly, in the outcome, $M$, $t^*(S^o)$, the slack on $jZ$ in the smaller network is equal to the minimum slack, $\min_{j \in Z \cap B} \sigma_j(Z \cap A)$ over all jobs in $Z$ in the larger network in the outcome, $M$, $t^*(S)$, and from the fairness, they are equal. Therefore, $t^*(S)$, also lexicographically maximizes the slack on all the agents. Therefore, $t^*(S)$ is the nucleolus in the larger network.

We now prove the main result.

Proof: The proof follows by induction similar to the proof of theorem 6 but using proposition 16 and lemma 9. Since the algorithm computes the unique fair outcome which is also the unique nucleolus, the fair outcome coincides with the nucleolus.
Chapter 4

Design of Public Institutions for Networked Societies

The role of public institutions is to provide for public goods, control externalities and establish cooperation in the society. Cooperation in a large society of self-interested individuals is extremely difficult to achieve when the externality of one individual’s action is spread thin and wide on the whole society. This leads to problems such as the ‘tragedy of the commons,’ in which rational action will ultimately make everyone worse off. Traditional policies to promote cooperation include government quotas and Pigouvian taxation or subsidies. The Pigouvian mechanisms make individuals internalize the externality they incur. However, such solutions are based on a model of externality that discounts local exchanges and peer pressure.

In this chapter, I present a joint model of externalities and peer-pressure and propose an alternative solution to the problem of global cooperation when externalities are global, but interactions are local [76]. Rather than making individuals internalize their externalities via Pigouvian mechanisms, the proposed social mechanisms, localize those externalities to one’s peers in a social network. I show how this approach leverages the power of peer-pressure to achieve global cooperation. Surprisingly, these social mechanisms can significantly outperform the Pigouvian mechanism by requiring a lower budget to operate, even when accounting for the social cost of peer pressure. Even when the available budget is very low, the social mechanisms achieve a greater
improvement in the outcome.

I evaluated my proposed social mechanisms against the Pigouvian mechanism in two large scale experiments [3], [75]. The first experiment was designed to test the effect of the mechanisms on the physical activity level of the members of a networked society. The second experiment was designed to test the effect of the mechanisms on the energy consumption of a networked society. In both the experiments, a simple social mechanism was employed that rewarded each individual’s peers for her improved action. The Pigouvian mechanism rewarded each individual directly for her improved action. Both experiments verify the theoretical predictions of the joint model of externalities and peer pressure and suggest that the social mechanism can be more desirable for public policy design than Pigouvian policies.

4.1 Introduction

Cooperation in large societies of self-interested individuals is a crucial, yet extremely difficult goal to achieve [39]. Some of the most important problems in modern society, such as pollution, global warming, rising health care costs and insurance, arise from the inability to achieve cooperation over a large scale. If everyone does their small part in reducing pollution, consuming responsibly, maintaining a healthy lifestyle and contribute to the health insurance, these problems will be solved. The tragedy of the commons occurs when multiple individuals, acting rationally in their own self-interest, will ultimately deplete a common resource, to the detriment of everybody [52]. The cause of the tragedy is that the negative externality from any individual’s non-cooperative action is experienced by the large society, yet the benefit is borne entirely by the individual. For example, when someone buys an SUV with high CO2 emission as opposed to a hybrid vehicle with low CO2 emission, the full cost of pollution (e.g. in terms of health care cost and environmental effects) is borne by the whole society. The cost incurred by the individual is small compared to the benefit, and so she free rides.

One traditional solution to the problem of cooperation in large societies is enforc-
ing quotas that limit the production of negative externalities (e.g. cap on carbon emission). An alternative, market-based approach is Pigouvian taxation [11]. The idea is to bring up the cost of one’s actions to a level that accounts for the externality incurred. For example, a carbon tax may be added to the price of fuel in order to account for the true cost of consumption on the rest of society (and used to plant trees, say). By taking this true cost into account, the individual no longer has incentive to free-ride, and will only purchase the the SUV if it is worth the entire cost to the individual. Governments also often use Pigouvian subsidies (e.g. subsidize the price of solar panels) to encourage activities with positive externalities. In effect, these policies tax everyone in the society and redistribute it as subsidies to enforce cooperation. There are two reasons why subsidies are better than taxes: (a) subsidies give positive feedback and have a better effect [91]; (b) it is harder to institute a public policy in a free society that taxes people if they do not take a cooperative action (such as lead a healthy lifestyle). Similar to Pigouvian taxes, under subsidies individuals internalize the externalities caused by their actions.

The outcomes of such policies may be socially sub-optimal for two reasons. Firstly, the Coasian argument [30] fails due to the presence of significant transaction cost [104] and so a simple redistribution does not achieve a Pareto-efficient outcome. Secondly, these policies assume that the society consists of a population of independent individuals and discounts the fact that individual decisions are influenced by the interactions with the peers in the society.

There is growing evidence of the power of social influence in general, and peer pressure in particular, in promoting cooperative behavior. In evolutionary biology, mechanisms like direct reciprocity, network reciprocity, and group selection all work by exploiting locality of interaction [85]. When people interact in small groups in experimental public goods games (PGG), the ability to punish or reward peers (even at a cost) has been shown to promote cooperative behavior [42, 91]. Similarly, economic models of peer pressure show that pressure acts as a mechanism for improving the social welfare of the group [24] and is conducive of successful business partnerships [57]. In Micro Finance Institutions, stronger social ties can increase the likelihood
of repayment of joint-liability loans, by facilitating monitoring and enforcement [61]. The effectiveness of peers on loan repayment has been demonstrated even in the absence of joint liability [21]. In relation to natural resource usage, strong community identification is instrumental to preventing overuse of water resources [114], and so do other community-based incentives [115]. Top-down rules have been blamed for the degraded inshore ground fishery in Maine, in contrast to the Maine lobster fishery, which has been governed by informal community-based institutions and yielded much higher levels of compliance with sustainable resource use [39]. Recent evidence from microfinance suggests that such peer effects not only harness existing social capital, but can benefit from the creation of new social capital (e.g. by increasing the frequency of meetings among borrowers) [43].

What the above examples highlight is that many mechanisms for promoting cooperative behavior work locally. An individual free-riding on immediate social ties (e.g. causing bad smell in the neighborhood by burning paper for cooking) has much to suffer: complaint, loss of reputation, social exclusion, etc. In contrast, in the ‘tragedy of the commons’ scenario (e.g. CO2 emission from an SUV), most of the externality is felt by individuals who are not peers and have no way to exert peer-pressure. Yet the cost of exerting peer-pressure is higher for any given peer than the externality experienced. Consequently, the peers do not exert sufficient pressure to nudge the individual to cooperate. Indeed, centralized sanctioning institutions exist precisely to promote cooperation in larger human groups, in which peer punishment is not sufficient [103].

Against this background, I propose a new set of social mechanisms for policy makers to address the problem of externalities in a networked society, in which externalities are global, but interactions are local. The mechanisms work by inducing peer-pressure to promote cooperation.

Such social mechanisms have immense possibilities for public policy in particular for health and sustainability. Energy and climate change is one of the most important policy problems in modern society. A reduction in energy demand has immediate impact on the reduction of carbon dioxide emissions. Traditional solutions used by
policy makers have been price based. These solutions mostly involve subsidies for energy efficient products. The majority of public funding for energy efficiency in the US is used for these subsidies with little impact [48]. Taxation on the other hand has been difficult to impose due to political challenges. However, even though subsidies are transfers within the society they are an added burden to the tax payer. The impact of the pricing effect is poorly known and when making such pricing policies, there is no knowledge of the elasticity of demand of energy efficient products.

In the recent years, there has been growing work in the design of non price interventions. Most of these interventions are based upon insights from psychology and large programs have experimentally validated the effect of such programs. The largest of such programs is run by OPOWER. OPOWER mails Home Energy Report letters that compare a household’s energy use to that of similar neighbors and provide energy conservation tips. Earlier work showed that providing social norm information induces people to conserve energy [99],[84]. Similar evidence on the power of social norms was observed in voting [47], retirement savings [15], and charitable giving [44]. As of the end of 2010, OPOWER had contracts to run programs at 47 utilities in 21 states, including six of the largest ten utilities in the U.S. These programs typically yield savings on the order of 2 to 5% across the population in the targeted area of utility consumption, equivalent to the effect of a price increase of 11 to 20% [5]. There are two types of normative feedbacks that a commonly studied, namely descriptive normative feedback and injunctive normative feedback. Descriptive feedback tells the consumers their performance as compared to other consumers of the similar type. Injunctive feedback gives a judgment of their absolute performance to the consumers without any comparison to others. Earlier experiments only used descriptive norms. In these experiments, while the high consumers reduce their energy consumption, the low consumers increased their consumption [99]. This was because the low consumers also wanted to follow the norm and therefore increased their consumption. When the injunctive feedback was added through smiley faces for low consumers and frowny faces for high consumers, a better impact was observed. Recent work has shown that injunctive feedback alone provides comparable effects and undermines the effect of
Such programs use the power of social norms and inactive peer pressure. The information of such social norms induces inactive peer-pressure on consumers who want to stay with the norm a phenomenon commonly known as 'keeping up with the Joneses.' Such peer-pressure is not applied by a conscious choice by the peers on a consumer and requires no interaction among the consumers. I propose the next step towards the non price interventions to reduce energy consumption. My experiments with energy conservation show that the effect of social mechanisms can be much more than Pigouvian mechanisms or normative effects.

The organization of the rest of the chapter is as such. In the next section, I introduce the main results of this chapter. In the following section, I present the theoretical model, introduce the social mechanism and give results related to its performance. In the following section, I discuss two field experiments to study the performance of social mechanisms. Finally, I provide the conclusions in the following section.

4.2 Main Results

I now introduce the main results of this chapter. The first set of results are theoretical and the second set of results verify the theoretical results and suggest applications in public policy.

4.2.1 Overview of Mechanism

The main idea behind the mechanism is illustrated in Figure 4-1. In contrast with the Pigouvian approach, which focuses on the individual causing the externality, social mechanism focuses on their peers on the social network. The idea is to incentivize agent A’s peers to exert (positive or negative) pressure on A. In reality, such peer pressure may take many forms. In the energy consumption scenario, examples of positive pressure include giving useful energy-saving advice to a neighbor, or giving the neighbor a hand at hauling and installing energy efficient appliances.
4.2. MAIN RESULTS

(a) Internalizing of Global Externalities (Pigouvian taxes or subsidies)

(b) Localizing Externalities (peer-based incentives)

Figure 4-1: (a) Individuals A and B (red) actions cause a global externality on the rest of society (blue). This behavior can be encouraged/discouraged by internalizing the externality in the form of subsidy or tax on A and B directly; (b) Localizing externalities to the peers (yellow) of individuals A and B incentivizes their respective peers to use peer influence to encourage/discourage the behavior causing the positive/negative externality, respectively.
of negative pressure include inducing guilt or shame [57], or slashing the tires of the neighbor’s new SUV [91].

Regardless of its form, peer pressure requires costly effort on behalf of agent A’s peers. A peer often has little incentive to exert such pressure, since the resulting effect, in terms of reduction in negative externality experienced by the given peer, is very low. Social mechanism can be summarized by the following question: *If we reward the peers of agent A, can we encourage them to exert more pressure on A to reduce the negative externality? And is this policy efficient compared to Pigouvian policies?*

My main insight is that by targeting the individual’s peers, peer pressure can amplify the desired effect on the target individual. That is, under certain conditions, the resulting reduction in negative externality can be larger, given an identical subsidy budget.

My results hold for positive and negative externalities, positive and negative peer pressure, as well as subsidy and taxation. For simplicity, in the remainder of the chapter, I focus on a single scenario. An individual chooses a level of action (e.g. consumption of electricity) which causes negative externalities (e.g. pollution). An external entity (e.g. municipal government) attempts to encourage the individual to reduce their action level through a direct reward scheme (e.g. subsidized solar panels). At the same time, the individual’s peers are able to exert negative pressure on the individual.

I study a joint strategic model of externalities and peer pressure in social networks where agents take actions that exert externalities on the whole network and may apply costly peer pressure on their peers. I model this as a two stage game in which, in the first stage, the individuals in the network choose the amount of peer pressure to put on each of their peers. In the second stage, the individuals choose their action that puts externality on the whole society. I study the sub-game perfect equilibrium of this game, show existence, conditions for uniqueness of actions across different equilibria, and how the peer-pressure is distributed in the social network in the different equilibria. I find that, in the equilibrium of this game, only the peers
who feel the highest externality apply pressure. Furthermore, the pressure felt by any individual in the network is the same in all the equilibria. This pressure yields some improvement in social surplus, but may not be optimal.

With these characterization results, I explore how, using information about the structure of the social network, optimal social surplus can be achieved using carefully designed social mechanism’s. Furthermore, I show that the social mechanism achieves the optimal outcome at a lower budget and total cost than the Pigouvian mechanism.

Social mechanisms are superior for two reasons: (i) When all the externalities are internalized as in the Pigouvian mechanism then there is no peer pressure on the agent creating the externality, and thus requires additional subsidies; and (ii) When the marginal cost of exerting peer pressure is lower than the marginal externality on the whole society times the marginal response to peer-pressure then the effect of subsidies is amplified in the social mechanism. This amplification increases with the strength of the relationships between the peers, and is inversely proportional to the cost of exerting peer pressure.

I anticipate two applications of such mechanisms: (i) public policy for reducing global externalities such as pollution, and (ii) revenue management for products with network externalities such as collaborative search engines, or social recommendations.

### 4.2.2 Applications to Promoting Active Lifestyle and Conserving Electricity

I applied social mechanisms to two policy problems. The first one was to promote active lifestyle. An experiment was conducted by Aharony et. al. [3] in a graduate dorm at MIT to increase the physical activity level of the residents. The experimental data showed that the cost efficiency of the social mechanism was four times that of the Pigouvian mechanism.

The second problem was to promote electricity conservation. With my collaborators at ETH, I conducted an experiment with a Swiss utility company to test the effect of a social mechanism on the reduction in electricity consumption by inducing
peer pressure. The experimental data shows that the effect of the social mechanism was a reduction in energy consumption by 17.4% that is equivalent to the effect of a short price increase by 96 – 174% or a long term price increase by 45%. The experimental results are remarkable. Neither the pricing mechanisms, nor the normative mechanisms have shown such large effects on energy consumption. This suggests that the social mechanisms have huge implications for energy policy.

4.3 Localizing Externalities in social networks: Inducing Peer Pressure to Promote Cooperation

In this section, I present a joint model of externalities and peer pressure win social networks and propose social mechanisms to control externalities by inducing peer pressure. I also provide theoretical results to show that social mechanisms can outperform Pigouvian mechanisms and identify the testable predictions that can be verified through field experiments.

4.3.1 Model of Externalities

I now present the formal model. Consider a set of agents $N$ in a social network $S = (N, E)$, where $E \subseteq N \times N$. Let $Nbr (i) = \{j : (i, j) \in E\}$ be the set of peers of the actor $i$. I assume that the social network is sparse and agents has at most $K$ peers. Agent $i \in N$ takes an action $x_i \in \mathbb{R}_+$ (e.g. corresponding to units of electricity consumed), and I define $\mathbf{x} \in \mathbb{R}_{+}^{[N]}$ to be the action profile of all agents. Each agent $i$ experiences raw utility from its action defined by the function $u_i : \mathbb{R}_+ \rightarrow \mathbb{R}$. I assume $u_i$ is a twice differentiable and strictly concave raw utility function with a unique maximum and lower bounded first derivative that approaches infinity as the action approaches zero. This is a very natural case (see SI Appendix for more).

In the standard model of externalities [79, Ch. 11], it is assumed that the utility of actor $i$ depends both on the raw utility of its own action as well as the externalities experienced due to the actions of others. The latter can be captured by a function
4.3. LOCALIZING EXTERNALITIES IN SOCIAL NETWORKS: INDUCING PEER PRESSURE TO PROMOTE COOPERATION

$v_i : \mathbb{R}_+ \rightarrow \mathbb{R}$, which is strictly convex and increasing, and captures the externality experienced by $i$ due to the aggregate action of other actors in the population.\(^1\)

Therefore, the total utility of actor $i$, given its own action $x_i$ and the action $x_{-i}$, $\mathcal{N} \setminus \{i\}$ of other agents is as follows:

$$U_i(x_i, x_{-i}) := u_i(x_i) - v_i\left(\sum_{j \neq i} x_j\right).$$

Social surplus is defined as the sum of utilities achieved by all agents in both the models.

$$S(x) := \sum_{i \in \mathcal{N}} U_i(x_i, x_{-i}) \quad (4.1)$$

I choose a linear social surplus function so that the social surplus maximization selects a unique Pareto-efficient outcome, among the infinitely many Pareto-efficient outcomes, that can be compared under different mechanisms.

**Equilibrium in the Standard Externalities Model**

The equilibrium in the standard externalities model is well understood. There is a unique equilibrium, in which each agent takes the action that maximizes the raw utility of her action irrespective of the actions taken by other agents. I will denote, $x^*$, as the action profile at equilibrium, and $x^\circ$, the action profile that maximizes social surplus. Note that $x^\circ$ is also a Pareto-efficient action profile. At optimal action $x^\circ$, for all agents $i$, the raw marginal utility is equal to the marginal externality on the whole society, i.e.- $u'_i(x^\circ_i) = \sum_{j \neq i} u'_j(\sum_{k \neq j} x^\circ_k)$. In the standard model of externalities, I know that at the equilibrium, the agents take action that is higher than the socially optimal action, that is $x^* > x^\circ$ [79] and therefore the social surplus at the equilibrium is sub-optimal. This is the essence of the Tragedy of the Commons.

---

\(^1\) For simplicity, I assume externality is negative, but the analysis should not fundamentally change with positive externalities, negative utility of action and concave externality function.
4.3.2 Externalities with Peer Pressure

Consider, now, that actors have the ability to exert peer pressure on their peers in the social network. I denote the peer-pressure profile by the matrix $\mathbf{p} \in \mathbb{R}^{N \times N}$, where the element $p_{ij}$ is the peer-pressure exerted by the agent $i$ on her peer $j$. Note that if $i$ and $j$ are not peers in the social network, then $p_{ij} = 0$. The $i$th column of $\mathbf{p}^T$ (transpose of $\mathbf{p}$), $\mathbf{p}_{\uparrow i}$, is the vector of peer-pressures exerted by $i$ over her peers. Similarly, the $i$th column of $\mathbf{p}$, $\mathbf{p}_{\downarrow i}$, is the vector of peer-pressures exerted over $i$ by her peers. A similar model was proposed by Calvó-Armengol and Jackson[24]. In their model, there is no network constraint and all agents can exert peer-pressure on all other agents. Also the model deals with binary action space as opposed to the infinite action space in my model.

The utility of an actor then takes the following extended form:

$$U_i(x_i, x_{-i}, \mathbf{p}) = u_i(x_i) - v_i \left( \sum_{j \neq i} x_j \right) - \left( \sum_{j \in \text{Nbr}(i)} p_{ji} \right) (x_i - x_i^o) - \left( \sum_{j \in \text{Nbr}(i)} p_{ij} \right) c$$

(4.2)

Thus, in addition to the raw utility of action and the externality, actor $i$ also experiences potential disutility that is bilinear in the total pressure from the peers and $i$'s own action. An individual’s action and the peer pressure on the individual enter as strategic substitutes in the individual’s utility. The higher $i$’s action is, the more salient the effect of the pressure becomes. Agent $i$ also incurs a cost $c.p_{ij}$ should it wish to exert pressure on neighbor $j$, where $c$ is the marginal cost of exerting such pressure, and $p_{ij}$ is the amount of pressure.

In the vector form, the utility takes the form:

$$U_i(x_i, x_{-i}, \mathbf{p}) = u_i(x_i) - v_i (\mathbf{1}. x_{-i}) - (\mathbf{1}. \mathbf{p}_{\downarrow i}) (x_i - x_i^o) - (\mathbf{1}. \mathbf{p}_{\uparrow i}) c$$

(4.2)

{I will denote $\mathbf{1}$ as a vector of ones with matching dimensions when taking the dot product.}

Although I consider that the marginal cost of exerting peer pressure is identical for all pair of peers, the qualitative nature of my results do not change with different costs. I study the externalities model with peer-pressure as a two-stage game. In
the first stage, actors choose the amount of peer pressure they wish to exert on their peers. In the second stage, actors observe pressure on themselves and then choose their action as a response to the observed pressure. I assume that the raw marginal utility of any agent is convex in her action.\footnote{This assumption is very natural and satisfied by almost all natural utility functions such as the Cobb-Douglas utility function.} Given pressure $p$, I denote $x^*_i(p)$ as the optimal response to the pressure for agent $i$. The optimal response is unique and the marginal raw utility for actor $i$ at the optimal response is equal to the total pressure exerted on $i$, i.e. $u'_i(x^*_i(p)) = \sum_{j \in Nbr(i)} p_{ji}$. As the pressure on an agent increases, the optimal response decreases and the optimal response is convex in the total pressure on the agent. The marginal response by any agent $i$ to pressure $p_{ji}$ from any of her peers $j$ is the reciprocal of the curvature of the raw utility function at the optimal response, i.e.- $\frac{\partial x^*_i(p)}{\partial p_{ji}} = \frac{1}{u''_i(x^*_i(p))}$.

### 4.3.3 Assumptions

As in the Tragedy of the Commons, the scenario I am interested in is when the externality caused rises much slower than the raw utility function and any one agent’s marginal change in action has very small effect on the marginal externality experienced by any given other agent. In this case, the large externality is due to higher aggregate action of all agents. I am interested in the scenario, where the marginal cost of exerting pressure is neither too high that no one cares to exert any pressure, nor too low that pressure is so high that everyone takes the socially optimal action. Therefore, I make the following assumptions about the marginal cost of exerting peer-pressure.

Firstly, I am interested in situations in which the cost of peer-pressure is not too low. Formally, the marginal cost of exerting peer-pressure is higher than the ratio of the marginal externality of any agent, $j$, to the curvature of the raw utility of any of her peers, $i$, when the action profile is socially optimal $x^o$, i.e.- $c > \left| \frac{v'_j(\sum_{k \neq j} x^o_k(p))}{u''_i(x^o_i(p))} \right|$ for all peers $i, j$. In the absence of this condition, optimal pressure is already present, everyone in society is well-behaved (social surplus is optimal), thus eliminating any
need for intervention.

Secondly, I am interested in situations in which, at least for some agents, peer-pressure is worth its cost—if the cost of exerting peer-pressure is too high, then no one will exert peer-pressure on anyone. Formally, for at least one agent \(i\) and one of her peers \(j\), the marginal cost of exerting pressure for \(j\) on \(i\) is lower than the ratio of the marginal externality experienced by \(j\) to the curvature of the raw utility of \(i\) when the action profile is \(x^*\), i.e.
\[
\left| \frac{v_j' \left( \sum_{k \neq j} x_k^* (p) \right)}{u_i'' (x_i^* (p))} \right| < c
\]
This means that \(j\)'s pressure can yield some benefit to her by influencing \(i\)'s behavior.

The above conditions on the cost of peer pressure are satisfied in scenarios such as the Tragedy of the Commons, in which peer pressure can potentially act as a conduit for social welfare improving behavior, but it is too costly to apply sufficient pressure. For example, suppose agent \(i\) is a smoker, paying little regard to others around him, while contributing to the broader problem of lung cancer. From the perspective of his neighbor \(j\), exerting peer pressure by expressing disapproval or asking him to smoke outside can yield a real benefit in terms of reduced discomfort to \(j\). Yet, it is sufficiently costly (read socially awkward) for \(j\) to exert enough pressure to get \(i\) to put out the cigarette, let alone quit. This causes \(j\) to refrain from pushing further.

This is also the case with problems of externalities such as pollution. The change in externality felt by any one agent (who’s putting pressure) due to the reduction in one other agent’s pollution level (e.g. replacing his SUV with a hybrid vehicle) is much smaller than the gain in utility for the agent creating the pollution (and enjoying his SUV). Yet, if a large number of agents reduce their pollution level simultaneously, then the reduction in the externality felt by any one agent (reduced pollution) may be much larger than the loss in the utility of any one agent changing her pollution level (replacing the SUV).

Finally, I assume that the society is large enough that the marginal cost of exerting peer-pressure is less than \(1/2K\) times the ratio of marginal externality felt by the whole society and the curvature of the raw utility function for any agent \(i\), at the socially optimal action profile \(x^o\), i.e.
\[
\left| \frac{\sum_{j \neq i} v_j' \left( \sum_{k \neq j} x_k^o (p) \right)}{u_i'' (x_i^o (p))} \right| = \frac{1}{2K} \left| \frac{u_j' (x_j^o (p))}{u_i'' (x_i^o (p))} \right|
\]
This relates the cost of peer-pressure to the relative locality in interaction. This means that the
4.3. LOCALIZING EXTERNALITIES IN SOCIAL NETWORKS: INDUCING PEER PRESSURE TO PROMOTE COOPERATION

For convenience, I define some notation:

- Let $p_{↓i} := \sum_{j \in Nbr(i)} p_{ji} = 1.p_{↓i}$ denote the total pressure exerted over $i$ by all his neighbors.
- Let $p_{↑i} := \sum_{j \in Nbr(i)} p_{ij} = 1.p_{↑i}$ denote the total pressure exerted by $i$ towards his neighbors.

I study the equilibrium of this game in the next sections.

4.3.4 Equilibrium with Peer Pressure

I now study the subgame perfect equilibrium [88] in the externalities model with peer-pressure. The existence of a subgame perfect equilibrium is shown in section 4.3.9.

Although there may be multiple possible subgame perfect equilibria, the total peer-pressure on any agent $i$ and her action is same in all the subgame perfect equilibria under the following condition: for any pair of peers, $i,j$, the elasticity of the marginal externality felt by $j$ is lower than the elasticity of the curvature of the raw utility of $i$ with respect to the agent $i$’s action, i.e.-

$$\frac{\partial \log v_j'(\sum_{k \neq j} x_k)}{\partial x_i} < \left| \frac{\partial \log |u''_i(x_i)|}{\partial x_i} \right|,$$

for all action profiles that are not strictly dominated. This condition suggests that the marginal externalities grow slower than the decay of marginal curvature of the raw utility function. I will assume that this condition is satisfied for the rest of the chapter. The details of the uniqueness of action profile in all equilibria and prof are given in section 4.3.9. In the subgame perfect equilibrium of this game, I observe that the equilibrium action profile in the game with peer-pressure is lower than the action-profile in the game without peer-pressure. This highlights the potential benefit of peer pressure.

**Theorem 8.** The action profile in any subgame perfect equilibrium in the two-stage game with peer-pressure is strictly lower than the equilibrium action profile in the game without peer-pressure.
The complete proof of the theorem is given in section 4.3.9. Intuitively, under assumption 1, there is positive peer-pressure on all agents in any subgame perfect equilibria and hence the optimal response action for each agent is lower than the optimal response without peer-pressure. I also observe from the application of KKT conditions that in the equilibria, not all agents exert pressure on all their peers (see section 4.3.9 for details). Assume $\mathbf{p}^*$ is the peer-pressure profile in an equilibrium.

(i) Each agent feels pressure only from the peers who have the highest marginal externality at the equilibrium action, i.e.- $p_{ji} > 0$ for the peers $i,j$ only if $v_j' \left( \sum_{k \neq j} x_k^* (\mathbf{p}^*) \right) \geq v_i' \left( \sum_{k \neq i} x_k^* (\mathbf{p}^*) \right)$ for all other peers $l \in Nbr (i)$.

(ii) Each agent puts pressure only on her peers who have the smallest curvature of the raw utility function at the equilibrium action, i.e.- $p_{ji} > 0$ for $i \in Nbr (j)$ only if $|u_i'' (x_i^* (\mathbf{p}^*))| \leq |u_k'' (x_k^* (\mathbf{p}^*))|$ for all $k \in Nbr (j)$.

While the presence of peer-pressure causes lower individual action, the peer-pressure in equilibrium is not sufficient to bring the action down to the optimal level. Next, I investigate how various interventions affect this.

4.3.5 Pigouvian Mechanism (Direct Reward)

In the standard Pigouvian mechanism, agents are rewarded for reduced action. The reward given to any agent $i \in N$ for her action $x_i$ is $r_i (x_i) = u_i' (x_i^*) (x_i^* - x_i)$. The utility function of agent $i$ under the Pigouvian mechanism therefore has the reward as an extra term. The utility function of agent $i$ under the Pigouvian mechanism is

$$U_i(x_i, x_{-i}, \mathbf{p}) = u_i(x_i) - v_i \left( \sum_{j \neq i} x_j \right) - \left( \sum_{j \in Nbr(i)} p_{ji} \right) x_i - \left( \sum_{j \in Nbr(i)} p_{ij} \right) c + r_i (x_i).$$

While this reward incentivizes reduction in individual action, it removes any incentive to exert peer pressure so there is not peer-pressure in the equilibrium.

**Proposition 17.** There is no peer-pressure on any agent in equilibrium under Pigouvian mechanism.
The complete proof is given in section 4.3.10. Intuitively, this is because under Pigouvian mechanism when the marginal reward for any agent is equal to the externality produced by the agent at the socially optimal action profile, the agents completely internalize the externality they produce and in the equilibrium each agent takes the socially optimal action. Following assumption 2, then the marginal cost of exerting peer-pressure is more than the marginal benefit for any agent as a result of the marginal reduction in externality and therefore no agent exerts peer-pressure on her peers. So the Pigouvian mechanism does not use any social capital.

4.3.6 Social Mechanisms (Rewarding the Peers)

As described earlier, the social mechanism rewards individuals for their peers’ low action, in effect subsidizing the cost of peer pressure they incur. There are many possible reward structures for creating social mechanism. I discuss here one structure in which the reward is given to agent $i$ as a result of her peer, agent $j$’s action $x_j$. Formally, $r_{ji} : \mathbb{R}^+ \to \mathbb{R}$ is strictly decreasing in $x_j$ (the more $j$ consumes, the less reward $i$ gets), and $r_{ji}(x_j)$ is the reward given to the actor $i$ for the action $x_j$ by a neighboring actor $j$. Since $i$ can have multiple peers, his utility incorporates aggregate social reward $\sum_{j \in \text{Nbr}(i)} r_{ji}(x_j)$. This reward can be incorporated into the game by the following expanded utility function:

$$U_i(x, p) = u_i(x_i) - v_i \left( \sum_{j \neq i} x_j \right) - x_i \sum_{j \in \text{Nbr}(i)} p_{ji} - c \sum_{j \in \text{Nbr}(i)} p_{ij} + \sum_{j \in \text{Nbr}(i)} r_{ji}(x_j)$$

What might be a suitable approach to allocating such social reward? We would like to identify a reward function which has the following properties:

1. The reward must be simple. I consider reward functions with constant marginal reward (i.e. affine reward functions).

2. A subgame perfect equilibrium of the two-stage game should exist.

3. Equilibrium action should be optimal.
4. Each peer gets rewarded for an agent’s reduced action.

5. Budget for rewards should be minimized over the set of reward functions that satisfy the above conditions.

It turns out that a fairly simple reward function satisfies these conditions. Changing the conditions 1-5 may give different reward functions. Note that this reward has a component that depends upon the consumer and a component that depends upon the neighbor.

**Theorem 9.** The following social reward function, to agent $i$ based on the action of agent $j$, satisfies conditions (1-5):

$$r_{ji}(x_j) = (\alpha_j + \beta_i) (x_j - x_j^*)$$

where $\alpha_j = cu_j''(x_j^*)$ and $\beta_i = v'_i\bigg(\sum_{k \neq i} x_k^*\bigg)$

The marginal reward is such that each agent has an incentive to increase pressure on peers that take a higher action than socially optimal level and reduce pressure on peers that take lower action than social optimal level. The proof of the theorem is given in section 4.3.11.

### 4.3.7 Efficiency and Cost of Social Reward

My most striking result is that (under assumption 3) the social mechanism is significantly more efficient than the Pigouvian mechanism. To establish this, I first compare the budgets needed by the two mechanisms in order to reach optimal action.

**Theorem 10.** The budget for the rewards in the Pigouvian Mechanism is at least twice the budget for the rewards in the social mechanism.

The proof of theorem is provided in section 4.3.12.

While this result is very encouraging, a more comprehensive comparison must account for the total cost of the two mechanisms. Two additional costs needs to be accounted for. First, there is an aggregate social cost of exerting pressure on peers,
incurred by all those agents who choose to do so. This can be thought of as a loss of social capital. While a single definition of the term is still not established [105], applying peer pressure can be seen as utilizing some existing social capital that one possesses [119].

Secondly, there is a redistribution loss incurred by both mechanisms, to account for the cost of distributing rewards (or of taxation). For simplicity, I assume that the redistribution loss or the overhead introduced due to the taxation and distribution of rewards is a fraction, $\delta$ of the total taxation. The social mechanism needs a smaller budget for the rewards and thus need lower taxation on the society. Since taxation has overheads and reduces social surplus, lower taxation is better. However, social capital is used in the social mechanism and it is important to find out the total loss of the social capital. I show in the next theorem that the loss incurred in the social capital due to peer-pressure in the equilibrium under social mechanism is smaller than the increase in the redistribution loss in the equilibrium under Pigouvian mechanism over the the equilibrium under social mechanism.

**Theorem 11.** The loss in social capital in the equilibrium under the social mechanism is lower than the redistribution loss in the equilibrium under the Pigouvian mechanism for all redistribution factors $\delta \geq \frac{1}{2K}$.

The proof of the theorem is given in section 4.3.12. Based on this result, I can characterize conditions under which the cost of the social mechanism is lower than the Pigouvian mechanism.

**Corollary 6.** The total loss in the equilibrium under the social mechanism including the loss in the social capital and the redistribution loss is lower than the the redistribution loss in the equilibrium under the Pigouvian mechanism for distribution factors $\delta \geq \frac{1}{K}$.

The proof of the theorem is given in section 4.3.12. This result shows that the social mechanism is superior, even when accounting for the cost to society, and not just considering the cost to the party operating the mechanism. Often, the budget available is lower than the required budget for rewards. Under these circumstances,
the marginal reward is kept low and the whole externality created by the agents may not be completely internalized under the Pigouvian mechanism or completely localized under the social mechanism. A corollary is that when the marginal reward rate is very low, the action profile in equilibrium under the social mechanism is lower than the action profile in equilibrium under the Pigouvian mechanism. This is important for two reasons. Firstly, since the raw utility functions and the externality functions are often difficult to estimate due to limited availability of accurate data, and since the budget is often insufficient, for most practical purposes, the marginal reward rate is low. Secondly, it can be tested experimentally. My results provide three testable predictions, namely:

1. Following theorem 1, the equilibrium action profile in the game with peer-pressure has a higher social surplus than the equilibrium action profile in the game without peer-pressure.

2. As a corollary of theorem 4, for low marginal rewards, the equilibrium action profile under the social mechanism has a higher social surplus than the equilibrium action profile under the Pigouvian mechanism.

3. When the marginal cost of exerting peer-pressure decreases (i.e.- peers are closer friends), the social surplus in the equilibrium under social mechanism increases.

I now demonstrate the results through an illustrative example.

4.3.8 Illustrative Example: Promoting Energy Efficiency

Consider a homogeneous set of agents $N = \{1, ..., 100\}$, each agent has 10 peers in the social network. Each agent consumes electricity priced at 4 per unit. The raw utility function of all agents is $u_i(x_i) = 12x_i^{0.8} - 4x_i$ for all $i \in N$, and the externality function of all agents is $v_i(y) = 0.0001(y)^{1.5}$ for all $i \in N$. The marginal cost of exerting pressure is $c = 1$ per unit. Assume that the redistribution loss is $\delta = 0.1$ per unit of reward. Figure 4-2 shows how the raw utility of consumption, externality and
total utility of each agent changes with the increase in consumption, assuming each agent has the same consumption.

Figure 4-2: (navy) Raw utility of consumption, maximized at the equilibrium consumption of $X = 79.62$. The peer pressure will lower the equilibrium consumption only slightly to $X = 67.46$ due to high marginal cost of pressure; (green) Externality experienced due to other agents’ consumption; (red) Total utility curve is the difference between the navy and green curves, and is maximized at the (much lower) socially optimal consumption level of $X = 31.19$.

In the absence of peer pressure, we have a symmetric equilibrium with $x_i^* = 79.61$ for all $i \in N$. The socially optimal consumption is $x_i^o = 31.19$ for all $i \in N$. So the electricity consumption is more than two and a half times the socially optimal level. The ratio of the marginal externality on any agent to the curvature of the raw utility of any other agent at the equilibrium consumption $x^*$ is $1.33 > c$ and at socially optimum consumption $x^o$ is $0.27 < c$. Therefore the cost of exerting the socially optimal level of pressure on the peers is much higher than the resulting reduction in
externality experienced by any agent. The symmetric equilibrium in the model with peer-pressure is better than the model without peer-pressure. The peer-pressure on any agent in equilibrium is 0.13, significantly lower than the optimal peer-pressure 0.83. Therefore, the consumption in equilibrium is 67.46 which is more than twice the socially optimal consumption. Figure 4-2 illustrates the situation, highlighting that even if peer-pressure is possible, it is too costly to apply sufficiently enough to yield socially optimal consumption.

The total reward budget required to reduce the consumption to the socially optimal level using the Pigouvian mechanism will be 3995.40. On the other hand, the total reward budget using the social mechanism will be 1095.26. Figure 4-3 shows that the required budget for any target consumption level is lower under the social mechanism than under the Pigouvian mechanism.

The social cost in equilibrium under the social mechanism will be 82.50. Moreover, the combined total social cost and the redistribution cost under the social mechanism is 192.03. This is half of the total redistribution cost of 399.54 under the Pigouvian mechanism. Figure 4-4 shows that the redistribution and social loss for any target consumption level is lower under the social mechanism than under the Pigouvian mechanism.

4.3.9 Properties of the Subgame Perfect Equilibria of the two stage game

I now discuss the sub-game perfect equilibrium in the externalities model with peer-pressure.

Existence of the Subgame Perfect Equilibria

I first show how the peer pressure profile in the first stage effects the equilibrium in the second stage game. In the second stage of the game for the given pressure profile \( p \), each actor \( i \) chooses her action \( x_i \) that maximizes her utility \( U_i(x_i, x_{-i}, p) \) given the peer-pressure exerted on her. Note that an actor’s marginal utility in her own action
in the subgame starting at the second stage is independent of the actions of the other actors in the second stage. Only the first three terms in the utility function depend upon the action profile in the second stage and only the first and the third term depends upon an actor’s own action $x_i$. Therefore, in essence, the sub-game starting at the second stage is similar to the game in the standard externalities model. In the second stage, actors observe pressure on themselves and then choose their action as a response to the observed pressure. Given pressure $p_i$, I denote $x_i^*(p_i)$ or with some abuse of notation, $x_i^*(p_{i:j})$ as the optimal response to the pressure for agent $i$, which is also his equilibrium action in the second stage game. Since the marginal raw utility of action for all actors approaches infinity as action approaches zero, therefore this action for all actors is greater than zero. The optimal response is unique and the
Figure 4-4: (navy) Redistribution loss under Pigouvian mechanism for a target consumption level; (green) Redistribution loss under Social mechanism for a desired consumption level. It increases much slower as the target consumption level is reduced; (red) Social loss under Social mechanism for a desired consumption level; (magenta) Total (redistribution + social) loss under Social mechanism for a desired consumption level.

marginal raw utility for actor $i$ at the optimal response is equal to the total pressure exerted on $i$, i.e. $u'_i(x^*_i(p_{ji})) = p_{ji}$. The marginal response by any agent $i$ to pressure $p_{ji}$ from any of her peers $j$ is the reciprocal of the curvature of the raw utility function at the optimal response, i.e.- $\frac{\partial x^*_i(p_{ji})}{\partial p_{ji}} = \frac{1}{u''(x^*_i(p_{ji}))}$ \(^3\). Since, the raw utility of action is concave, the optimal response in the second stage is decreasing in the felt peer-pressure, i.e.- $\frac{\partial x^*_i(p_{ji})}{\partial p_{ji}} = \frac{1}{u''(x^*_i(p_{ji}))} < 0$. Also, since the raw marginal utility of action is convex in the agent’s action, therefore, the optimal response in the second stage is convex in the felt peer-pressure, $\frac{\partial^2 x^*_i(p_{ji})}{\partial p^2_{ji}} > 0$.

\(^3\)By using the inverse derivative law: $(f^{-1})' = \frac{1}{f'(f^{-1})}$
In the first stage of the game, the threat of higher action in the second stage, is an incentive for the actors to exert pressure on their peers. Given that the equilibrium action is taken in the second stage of the game, the actors in the first stage of the game choose peer-pressure on their neighbors that maximizes their utility function

$$U_i(p) = u_i(x_i^*(p_{ij})) - v_i \left( \sum_{j \neq i} x_j^*(p_{ij}) \right) - p_\downarrow x_i^*(p_{ij}) - p_\uparrow c \quad (4.4)$$

Peer Pressures are strategic substitutes

**Proposition 18.** The peer-pressure that any actor $i$ exerts on any of her peers $j$ is strict strategic substitute of any peer-pressure not exerted on $i$. Therefore, if any agent increases peer-pressure on one of her neighbors other than $i$, then the best-response peer-pressure exerted by $i$ on all other agents decreases.

**Proof:**

For any actor $i$ and any two of her peers $j, k$,

$$\frac{\partial^2 U_i}{\partial p_{ij} \partial p_{ik}} = -v_i'' \left( \sum_{m \neq i} x_m^*(p_{im}) \right) \frac{\partial x_j^*(p_{ij})}{\partial p_{ij}} \frac{\partial x_k^*(p_{ik})}{\partial p_{ik}} < 0,$$

since the externality is concave in the action and the best response action of the actors is decreasing in the felt peer-pressure. Therefore, the peer-pressures $p_{ij}$ and $p_{ik}$ are strict strategic substitutes of each other.

For any actor $i$ and $k$ with a common peer $j$,

$$\frac{\partial^2 U_i}{\partial p_{ij} \partial p_{kj}} = -v_i'' \left( \sum_{m \neq i} x_m^*(p_{im}) \right) \frac{\partial x_j^*(p_{ij})}{\partial p_{ij}} \frac{\partial x_k^*(p_{ik})}{\partial p_{kj}} - v_i' \left( \sum_{m \neq i} x_m^*(p_{im}) \right) \frac{\partial^2 x_j^*(p_{ij})}{\partial p_{ij} \partial p_{kj}} < 0,$$

since the externality is concave in the action and the best response action of the actors is decreasing and convex in the felt peer-pressure. Therefore, the peer-pressures $p_{ij}$ and $p_{kj}$ are strict strategic substitutes of each other.

For any distinct pair of neighboring actors $i, j$ and $k, l$,

$$\frac{\partial^2 U_i}{\partial p_{ij} \partial p_{kl}} = -v_i'' \left( \sum_{m \neq i} x_m^*(p_{im}) \right) \frac{\partial x_j^*(p_{ij})}{\partial p_{ij}} \frac{\partial x_l^*(p_{il})}{\partial p_{kl}} < 0,$$

since the externality is concave in the action and the best response action of the actors is decreasing in the felt peer-pressure. Therefore, the peer-pressures $p_{ij}$ and $p_{kl}$ are strict strategic substitutes of each other.

Eliminating Some Strictly Dominated Strategies
Since, all peer pressures are positive and the optimal response of the agents decrease with the felt pressure, the optimal response of any agent in the second stage game is always below the equilibrium action in the standard externalities model, \( x_i^* \). Therefore, all actions \( x_i > x_i^{max} = x_i^* \) and are strictly dominated. Let us denote \( p_{ji}^{max} = \arg \max_{p_{ji} \in \mathbb{R}_+} U_j \left( x_i^* (p_{ji}), x_{-i}^*, p_{ji}, 0 \right) \), the pressure that \( j \) puts on \( i \) that maximizes \( j \)'s utility given she does not put any peer-pressure on any other peer and no one else puts any peer-pressure on anyone else. Following Proposition 1, any peer pressure strategy \( p_{ij} \) in which \( p_{ji} > p_{ji}^{max} \) is strictly dominated. Since the maximum pressure on \( i \) from any of her peers is finite, therefore the maximum total pressure \( p_{ii}^{max} \) on agent \( i \) is finite and since \( i \)'s marginal raw utility of action approaches infinity as her action approaches zero, therefore all actions \( x_i < x_i^* (p_{ii}^{max}) = x_i^{min} \) in the second stage game are strictly dominated.

Therefore, for any agent \( i \), there exists critical actions, \( x_i^{min} \) and \( x_i^{max} \), such that all actions not belonging to \([x_i^{min}, x_i^{max}]\) are iteratively strictly dominated. Similarly, there exists a maximum peer pressure, \( p_{ij}^{max} \) such that for any pair of peers \( i, j \), all \( p_{ji} > p_{ij}^{max} \) are strictly dominated. There are more iterated strictly dominated strategies but I will restrict attention to these.

Therefore, I can restrict the set of actions for any actor \( i \) to \( X_i = [x_i^{min}, x_i^{max}] \) and peer-pressure strategy set for any actor \( i \) to \( P_i = \prod_{j \in Nbr(i)} [0, p_{ij}^{max}] \) for all \( p_{-i} \)(peer pressure profile of all agents other than \( i \)) without any loss of generality. The set of peer-pressure strategy profiles \( P = \prod_{i \in N} P_i \) is also compact and convex.

Existence of Subgame Perfect Equilibrium

**Proposition 19.** The set of subgame perfect equilibria in the two stage game is non-empty.

**Proof:** I first show that for all \( i \in N \) and for all \( p_{-i}, v_i \left( \sum_{j \neq i} x_j^* (p_{-i}, p_{i}^{\uparrow}) \right) \) is strictly convex in \( p_{i}^{\uparrow} \). To show this, I consider two arbitrary peer-pressure vectors \( p_{i}^{\prime} \) and \( p_{i}^{\prime\prime} \).

\[
v_i \left( \sum_{j \neq i} x_j^* (p_{-i}, \lambda p_{i}^{\prime} + (1 - \lambda) p_{i}^{\prime\prime}) \right)
\]
4.3. LOCALIZING EXTERNALITIES IN SOCIAL NETWORKS: INDUCING PEER PRESSURE TO PROMOTE COOPERATION

\[
\begin{align*}
&= v_i \left( \sum_{j \in N \setminus (Nbr(i) \cup \{i\})} x_j^* (p_{-i}, \lambda p_{ij}^\prime + (1 - \lambda) p_{ij}^{\prime\prime}) + \sum_{j \in Nbr(i)} x_j^* (p_{-i}, \lambda p_{ij}^\prime + (1 - \lambda) p_{ij}^{\prime\prime}) \right) \\
&< v_i \left( \sum_{j \in N \setminus (Nbr(i) \cup \{i\})} x_j^* (p_{-i}, \lambda p_{ij}^\prime + (1 - \lambda) p_{ij}^{\prime\prime}) + \sum_{j \in Nbr(i)} x_j^* (p_{-i}, \lambda p_{ij}^\prime + (1 - \lambda) x_j^* (p_{-i}, p_{ij}^{\prime\prime})) \right)
\end{align*}
\]

Since, for all \( j \in Nbr(i) \), \( x_j^* (p_{ij}) \) is strictly convex in \( p_{ij} \) and as a consequence in \( p_{ij} \)

\[
\begin{align*}
&= v_i \left( \sum_{j \in N \setminus (Nbr(i) \cup \{i\})} x_j^* (p_{-i}) + \sum_{j \in Nbr(i)} \lambda x_j^* (p_{-i}, p_{ij}^\prime) + (1 - \lambda) x_j^* (p_{-i}, p_{ij}^{\prime\prime}) \right) \\
&< \lambda v_i \left( \sum_{j \in N \setminus (Nbr(i) \cup \{i\})} x_j^* (p_{-i}) + \sum_{j \in Nbr(i)} x_j^* (p_{-i}, p_{ij}^\prime) \right)
\end{align*}
\]

Therefore the utility, \( U_i \), of actor \( i \) is strictly concave in \( p_{ij} \) for all \( p_{-i} \). Also, since \( v_i \) approaches \(-\infty\) as its parameter approaches \( \infty \), therefore there is a unique and finite \( p_{ij} \) for each \( p_{-i} \) that maximizes \( i \)'s utility.

Since, the peer-pressure strategy set, \( P_i \) for any actor \( i \) is compact and convex for any \( p_{-i} \) and the utility function of \( i \) is continuous and strictly concave, therefore by the theorem of the maximum \( [14] \), the best-response peer-pressure strategy, \( p_{ij}^* (p) \), of actor \( i \) is continuous in \( p \) and the best-response peer-pressure function \( p^*: P \rightarrow P \) is a continuous function from a compact and convex set to itself. Therefore, by Brower’s fixed point theorem, the best-response peer-pressure function \( p^* \) has a fixed point. Therefore, the set of equilibria is non-empty.

\[4\] Since, for all \( j \in Nbr(i) \), \( x_j^* (p_{ij}) \) is strictly convex in \( p_{ij} \) and as a consequence in \( p_{ij} \)

\[5\] Since, for all \( j \notin Nbr(i) \), \( p_{ij} \) is independent of \( p_{ij} \), therefore \( x_j^* (p) = x_j^* (p_{ij}) \) is independent of \( p_{ij} \)

\[6\] Since \( v_i \) is strictly convex
Who puts pressure on whom?

I make two observations for the equilibrium pressure profile \( p^* \) that follow from applying the Karush-Kuhn-Tucker (KKT) conditions [68].

**Proposition 20.** The set of peers an actor puts pressure on are the ones who have the smallest curvature of the raw utility function at the equilibrium action, i.e.- \( p_{ji} > 0 \) for \( j \in N, i \in Nbr(i) \) only if \( u_i''(x_i^*(p_{ji}^*)) \leq u_k''(x_k^*(p_{jk}^*)) \) for all \( k \in Nbr(i) \).

**Proof:** For the actor \( j \), if \( p_{ji}^* > 0 \), then by applying KKT condition to the equilibrium, I have:

\[
c = -v_j' \left( \sum_{k \neq j} x_k^*(p^*) \right) \frac{1}{u_i''(x_i^*(p_{ji}^*))}
\]

Therefore for any peer \( i \) on whom \( j \) exerts pressure, the curvature of \( i \)'s raw utility function \( |u_i''(x_i^*(p_{ji}^*))| = \frac{c}{v_j'(x_j^*(p^*))} \). and if \( p_{jk}^* = 0 \) for some \( k \in Nbr(j) \).

\[
c > -v_j' \left( \sum_{k \neq j} x_k^*(p^*) \right) \frac{1}{u_k''(x_k^*(p_{jk}^*))}
\]

Therefore for any peer \( k \) on whom \( j \) does not exert pressure, the curvature of \( k \)'s raw utility function \( |u_k''(x_k^*(p_{jk}^*))| > \frac{c}{v_j'(x_j^*(p^*))} \). Therefore the proposition holds.

In words, agent \( j \) will exert pressure on every neighbor for which the marginal cost of exerting pressure \( c \) equals the marginal reduction in externalities experienced from that neighbor. Moreover, note that all these neighbors will have identical marginal reduction in externality. On the other hand, \( j \) will not exert any pressure on the other neighbors, for which the marginal reduction in externality is lower than \( c \).

**Proposition 21.** An actor feels pressure from the peers who have the highest marginal externality at the equilibrium action, i.e.- \( p_{ji}^* > 0 \) for \( i \in N, j \in Nbr(i) \) only if \( v_j' \left( 1.x^*(p^*) - x_j^*(p_{ji}^*) \right) \geq v_k' \left( 1.x^*(p^*) - x_k^*(p_{jk}^*) \right) \) for all \( k \in Nbr(i) \).

**Proof:** For the actor \( i \), if \( p_{ji}^* > 0 \), then

\[
c \cdot |u_i''(x_i^*(p_{ji}^*))| = v_j' \left( \sum_{l \neq j} x_l^*(p^*) \right)
\]
and for all \( k \in Nbr (i) \) if \( p_{ki}^* = 0 \).

\[
c \left| u''_i \left( x_i^* (p_i^*) \right) \right| > v_k' \left( \sum_{i \neq k} x_i^* (p^*) \right)
\]

Therefore the proposition holds.

The above observations provide a system of equations that can be solved to find an equilibrium pressure profile \( p^* \) and the resulting action profile \( x^* \). Note that the equilibrium pressure profile need not be unique. However, as we see in the next subsection, the total pressure on each agent and consequently the action profile is same in all equilibria.

**Uniqueness of action profile in all subgame perfect equilibria**

**Assumption A1:** the absolute elasticity of the marginal externality felt by any agent \( j \in N \) is lower than the absolute elasticity of the curvature of the raw utility of any other agent \( i \in N \) with respect to the agent \( i \)'s action, i.e.

\[
\left| \frac{\partial \log |u''_i (x_i)|}{\partial x_i} \right| < \frac{\partial \log v'_j (x_i + 1, x_{-i,j})}{\partial x_i}
\]

for all action profiles \( x \in \prod_{k \in N} [x_{k, \text{min}}, x_{k, \text{max}}] \).

This assumption suggests that the marginal externalities grow slower than the decay of curvature of the raw utility function. We will see that this condition guarantees that the peer-pressure on any agent and consequently her own action does not vary across different equilibria. I first give a result that will help understand the implication of assumption A1.

**Proposition 22.** Define \( h (x, y) = -af (x) + bg (y) \) and where \( f \) and \( g \) are continuous, differentiable and positive functions where \( f \) is strictly decreasing and \( g \) is strictly increasing defined over closed intervals. Assume \( \frac{\partial \log g (y)}{\partial y} < -\frac{\partial \log f (x)}{\partial x} \). If \( h (x_1, y_1) = h (x_2, y_2) = 0 \), with \( y_2 > y_1 \) and \( x_2 < x_1 \). Then \( x_1 - x_2 < y_2 - y_1 \).

**Proof:**

Let \( l \) be the path from \( (x_1, y_1) \) to \( (x_2, y_2) \) along the indifference curve of \( h \), i.e. - \( h (x, y) = 0 \) for all \( (x, y) \) along \( l \).
\[ h(x_2, y_2) - h(x_2, y_2) = \int_{x_1, y_1}^{x_2, y_2} \nabla h(x, y).dl = 0 \]

\[ \Rightarrow \int_{x_1, y_1}^{x_2, y_2} (-af'(x) dx + bg'(y) dy).dl = 0 \]

\[ \Rightarrow \int_{x_1, y_1}^{x_2, y_2} \left(-\frac{af'(x)}{af(x)} dx + \frac{bg'(y)}{bg(y)} dy\right).dl = 0 \]

\{Since \( af(x) = bg(y) \) for all \((x, y)\) along \(l\)\}

\[ \Rightarrow \int_{x_1, y_1}^{x_2, y_2} \left(-\frac{\partial \log f(x)}{\partial x} dx + \frac{\partial \log g(y)}{\partial y} dy\right).dl = 0 \]

\[ \Rightarrow \int_{x_1, y_1}^{x_2, y_2} -dx.dl < \int_{x_1, y_1}^{x_2, y_2} dy.dl \]

\{Since \( -\frac{\partial \log f(x)}{\partial x} > \frac{\partial \log g(y)}{\partial y} > 0 \) for all \((x, y)\) along \(l\)\}

\[ \Rightarrow x_1 - x_2 < y_2 - y_1 \]

Using the above result, I show the uniqueness of action profile in the two-stage game with peer pressure.

**Proposition 23.** The action profile in the second stage of the game is same in all subgame perfect equilibria of the two-stage game under assumption A1.

**Proof:** I will prove this by contradiction. Assume, there are two subgame perfect equilibria \((\hat{p}, \hat{x}), (\hat{p}, \check{x})\) with two different action profiles \(\hat{x}\) and \(\check{x}\) respectively. Without loss of generality, assume that \(1' \hat{x} \geq 1' \check{x}\). Let \(i^* = \arg \max_{i \in N} \hat{x}_i - \check{x}_i\). Clearly, \(\hat{x}_{i^*} - \check{x}_{i^*} > 0\). This implies that \(\check{p}_{i^*} > \hat{p}_{i^*} \geq 0\). Pick a peer, \(j^* \in Nbr(i^*)\) such that \(\check{p}_{j^*} > 0\). I know that such a neighbor exists because \(\check{p}_{i^*} > 0\).
Therefore by KKT condition,

\[ cu''_i (\hat{x}_i^*) + v'_j \left( \sum_{k \neq j} \hat{x}_k \right) = 0 \]

If

\[ cu''_i (\hat{x}_i^*) + v'_j \left( \sum_{k \neq j} \hat{x}_k \right) < 0 \]

then by the previous proposition

\[ (1'.\hat{x} - \hat{x}_j^*) - (1'.\hat{x} - \hat{x}_j^*) > \hat{x}_i^* - \hat{x}_i^* \]

If

\[ cu''_i (\hat{x}_i^*) + v'_j \left( \sum_{k \neq j} \hat{x}_k \right) > 0 \]

then define \( x_i^* \) to be the action of \( i \) such that

\[ cu''_i (x_i^*) + v'_j \left( \sum_{k \neq j} \hat{x}_k \right) = 0 \]

Since the raw marginal utility of action is convex, therefore \( \hat{x}_i^* < x_i^* \) and therefore using the previous proposition

\[ (1'.\hat{x} - \hat{x}_j^*) - (1'.\hat{x} - \hat{x}_j^*) > x_i^* - \hat{x}_i^* > \hat{x}_i^* - \hat{x}_i^* \]

This implies that

\[ \hat{x}_j^* - \hat{x}_j^* > \hat{x}_i^* - \hat{x}_i^* + 1'.\hat{x} - 1'.\hat{x} > \hat{x}_i^* - \hat{x}_i^* = \max_{i \in N} \hat{x}_i - \hat{x}_i \geq \hat{x}_j^* - \hat{x}_j^* \]

which is a contradiction. Therefore, there cannot be two subgame perfect equilibria with different action profiles.

**Comparing the Equilibrium Action in the Two Games**

Proof of Theorem 8
Proof:

The action profile in the equilibrium in the game with peer-pressure is same as the action profile in the equilibrium of the second stage in the two-stage game when the peer-pressure profile in the first stage is 0. Since the equilibrium action profile in the second stage decreases with any increase in the peer-pressure profile in the first stage and the peer-pressure profile in the subgame perfect equilibrium of the two-stage game is positive, therefore the result holds.

4.3.10 Proof of Proposition 17

Proof:

When there is no peer-pressure on any agent, the equilibrium action profile is $x^\circ$. For any pair of peers $i,j$, $c > -\frac{v'_j(x^\circ)}{u'_i(x_i)}$, therefore by KKT conditions the best response peer-pressure for agent $j$, $p_{ij} = 0$. Therefore there is no peer-pressure on any agent in the equilibrium under Pigouvian mechanism.

4.3.11 Proof of Theorem 9

Proof: In the first stage, the actors choose the pressure such that action profile in the second stage maximizes their reward minus the externality on them minus the cost they incur to exert pressure. The optimum amount of pressure is such that the action profile in the second stage is socially optimal. The threat of over action reducing the reward for the actors in the second stage sustains the optimum pressure exerted by the actors. Thus, at the optimum pressure profile $p^\circ$, we should have in the second stage:

$$
\left(-v'_i \left( \sum_{k \neq i} x^*_k(p) \right) \frac{\partial x^*_j(p)}{\partial p_{ij}} \right. \\
\left. - c + r'_{ji} \left( x^*_j(p) \right) \frac{\partial x^*_j(p)}{\partial p_{ij}} \right) \bigg|_{p^\circ} = 0
$$

which gives:

$$
\left( \left( r'_{ji}(x^*_j(p)) - v'_i \left( \sum_{k \neq i} x^*_k(p) \right) \right) \frac{\partial x^*_j(p)}{\partial p_{ij}} \right) \bigg|_{p^\circ} = c
$$
Therefore:

\[ r'_{ji}(x^*_j) = cu''_j(x^*_j) + v'_i \left( \sum_{k \neq i} x^o_k \right) \]

This characterizes the reward function such that the pressure profile in the first stage is optimum. This is not a full characterization of the reward function as it only suggests constraints on the reward function such that the optimal pressure is chosen in the first stage. There are infinitely many reward functions that satisfy these constraints. In short the constraint suggests that at action levels below the optimal, the marginal reward of the neighbors additional action is higher than the sum of the marginal reduction in cost by reducing pressure on the neighbor and the marginal externality introduced by the neighbors action. Thus, it is advantageous for the actors to reduce pressure on their neighbors. On the other hand, at action levels above optimum, the marginal reward for the neighbors additional action is lower than the sum of the marginal reduction in cost by reducing pressure on the neighbor and the marginal externality introduced by the neighbors action. Thus, it is advantageous for the actors to increase pressure on their neighbors. To fully characterize the reward function, I use conditions 1 and 2.

The following reward function satisfies the conditions:

\[ r_{ji}(x_j) = - (\alpha_j + \beta_i) \left( x^*_j - x_j \right) \text{ where } \alpha_j = cu''_j(x^*_j) \text{ and } \beta_i = v'_i \left( \sum_{k \neq i} x^o_k \right) \]

### 4.3.12 Proofs for Welfare Comparison Results

**Proof of Theorem 10**

**Proof:**

The budget for the total reward distributed to the peers of agent \( j \) at the optimal

\[ \frac{\partial x^*_j(p)}{\partial p_{ij}} = \frac{1}{u'(x^*_j(p))} \]

using the inverse derivative law: \( (f^{-1})' = \frac{1}{f'(f^{-1})} \).
action profile is

\[ B^*_j = \sum_{i \in \text{Nbr}(j)} r_{ji}(x^o_j) = \sum_{i \in \text{Nbr}(j)} (\alpha_j + \beta_i) (x^*_j - x^o_j) \]

Under the Pigouvian subsidies, the total subsidy given to the agent \( j \) at the optimal action is:

\[ B^p_j = \left( \sum_{i \neq j} \beta_i \right) (x^*_j - x^o_j) \]

The ratio of the two budgets for the consumer \( j \) at the optimal action is:

\[ \frac{B^*_j}{B^p_j} = \frac{-\sum_{i \in \text{Nbr}(j)} (\beta_i + \alpha_j)}{\sum_{i \in \mathbb{N}} \beta_i} < \frac{-\sum_{i \in \text{Nbr}(j)} (\alpha_j)}{\sum_{i \in \mathbb{N}} \beta_i} = -\frac{|\text{Nbr}(j)|}{u'_j(x^o_j)} \]

\[ < -\frac{|\text{Nbr}(j)|}{2K} \leq \frac{1}{2}, \text{ since } c < -\frac{u'_j(x^o_j)}{2K u''_j(x^o_j)} \]

Therefore, the total budget in the social mechanism is lower than the total budget in the Pigouvian mechanism. \( \sum_{j \in \mathbb{N}} B^*_j < \sum_{j \in \mathbb{N}} \frac{|\text{Nbr}(j)| B^p_j}{2K} < \frac{1}{2} \sum_{j \in \mathbb{N}} B^*_j. \)

**Proof of Theorem 11**

**Proof:**

The redistribution loss in the equilibrium under the Pigouvian mechanism due to the rewards given to any agent \( i \) is

\[ \delta u'_i(x^o_i) (x^*_i - x^o_i). \]

The loss in social capital in the social mechanism is the cost of applying the peer-pressure. There is no loss of social capital due to agents’ action because each agent
4.3. LOCALIZING EXTERNALITIES IN SOCIAL NETWORKS: INDUCING PEER PRESSURE TO
PROMOTE COOPERATION

's action is socially optimal. The total loss in the social capital to the peers of agent
i is

\[ c(p^\circ_{i,i}) = cu_i'(x^\circ_i) \]

\[ = c(u_i'(x^\circ_i) - u_i'(x^*_i)), \{\text{Since } u_i'(x^*_i) = 0\} \]

\[ = c \left( \int_{x^\circ_i}^{x^*_i} -u''_i(x) \, dx \right) \]

\[ < c \left( \int_{x^\circ_i}^{x^*_i} -u''_i(x) \, dx \right), \]

\{Since the raw marginal utility is strictly convex, therefore\}

\[ u''_i(x^\circ_i) < u''_i(x) \text{ for all } i \text{ and for all } x > x^\circ_i \}

\[ = -cu''_i(x^\circ_i) (x^*_i - x^\circ_i) \]

\[ < \frac{1}{2K} u'_i(x^\circ_i) (x^*_i - x^\circ_i), \{\text{Since } c < -\frac{1}{2K} u''_i(x^\circ_i) \text{ for all } i\} \]

\[ \leq \delta u'_i(x^\circ_i) (x^*_i - x^\circ_i). \]

Proof of Corollary 6

Proof:

The redistribution loss in the equilibrium under social mechanism due to the
rewards distributed to the peers of any agent \( i \) is

\[ -\delta K \left( cu''_i(x^\circ_i) + v'_i(x^\circ_i \cdot x^\circ_i) \right) (x^*_i - x^\circ_i) \]
Therefore, the total loss in the equilibrium under social mechanism from rewards and social loss to the agents of $i$ is
\[
cu'_i (x_i^o) - \delta K \left( cu''_i (x_i^o) + v'_i (x_{-i}^o) \right) (x^*_i - x_i^o)
\]
\[
< cu'_i (x_i^o) - \delta K cu''_i (x_i^o) (x^*_i - x_i^o)
\]
\[
< cu'_i (x_i^o) + \frac{1}{2} \delta u'_i (x_i^o) (x^*_i - x_i^o)
\]
\[
< \left( \frac{1}{2K} + \frac{\delta}{2} \right) u'_i (x_i^o) (x^*_i - x_i^o)
\]
\[
= \left( \frac{1}{2K \delta} + \frac{1}{2} \right) \delta u'_i (x_i^o) (x^*_i - x_i^o)
\]
\[
\leq \delta u'_i (x_i^o) (x^*_i - x_i^o)
\]
, since $\delta \geq \frac{1}{K}$.

4.4 Experiments and Applications to Public Policy

I now discuss two experiments designed to verify the testable predictions of my results. The experiments are also a suggestive template for policy interventions to promote active lifestyle and energy conservation.

4.4.1 Promoting Physical Activity

This experiment was done by my colleagues [3] in my consultation. In the "Friends and Family" study at MIT [3] the social mechanism was tested against the Pigouvian mechanism. The details of the experiment are provided in [3]. My model and results
are consistent with data observed in the experiment and the three testable predictions are verified. In the MIT "Friends and Family" study, the aim was to encourage people to increase their daily physical activity, as measured by the accelerometer sensors in their mobile phones. The relationships between individuals were captured through surveys and physical proximity measured by bluetooth data from the phones. Each individual was assigned two peers. Some individuals had close friends as their peers and some had strangers as peers. Three treatments were experimented with. In the self-monitoring (control) condition, individuals were rewarded for their own increased physical activity. This condition simulates the Pigouvian mechanism. Two treatment conditions involved assigning two "buddies" to each participant. In the "peer-view" treatment condition, the participant was shown the buddies’ activity levels, but was still rewarded for his/her own activity level. This condition simulates the game with peer-pressure under the Pigouvian mechanism. In the "peer-reward" condition, it was the buddies who received a reward proportional to the participant’s activity. This condition simulates the social mechanism. The peer-view condition mediates normative forces [99], while the peer-reward condition was designed to promote peer pressure.

Before the mechanisms were deployed, the activity levels of the individuals were monitored for a period of 23 days. The average normalized activity levels (measured by accelerometers) of the three groups were as follows: (a) 1.162 for the control group, (b) 1.266 for the peer-see group, and (c) 1.216 for the peer-reward group. The marginal reward was computed such that the budget for the reward in all the three conditions was roughly the same. The marginal rewards were (a) $ 83.3 (=1/.012)$ in the control condition, (b) $ 39.5 (=1/.0253)$ in the peer-see condition (roughly half of the marginal reward in the control condition) and (c) $ 12 (=-1/(2x0.0416))$, in the peer-reward condition (roughly one third the marginal reward in the peer-see condition) since the reward was given to two peers for each agent. After the mechanisms were deployed and removing the initial 19 days left for stabilizing the behaviors, the activity levels were monitored for the subsequent 20 days. The change in the average normalized activity levels (measured by accelerometers) of the three
groups were as follows: (a) 0.037 (3.2%) for the control group, (b) 0.070 (5.5%) for the peer-see group, and (c) 0.126 (10.4%) for the peer-reward group.

Results reveal that peer pressure is indeed present when the information about an individual’s actions is known to the peers. The "peer-see" condition showed a significant increase in activity level compared to the control (0.070 vs 0.037, p<0.01), even when the peers did not receive any reward, suggesting an effect of peer pressure. This verifies the first testable prediction.

The data reveals that at low marginal rewards, the peer-based mechanism yields higher returns. Marginal reward is the reward distributed per unit increase in physical activity. Despite lower marginal reward, the peer-reward condition yielded significantly larger increase in activity level than both the control and peer-see conditions (0.126 vs 0.07 and 0.037 respectively). This verifies the second testable prediction.

Furthermore, when both assigned peers of an individual were close friends, the peer-reward condition exhibited twice the improvement in the normalized activity levels as compared to the same condition with weaker social ties with the assigned peers. The change in the average normalized activity level for the peer-reward mechanism (a) in the triads consisting of close friends was 0.269 for an average reward of $3.00 and (b) in the triads consisting of strangers was 0.137 for an average reward of $2.95. This shows that when social capital is high, leading to lower cost of exerting pressure, the peer-reward mechanism has even better returns. This verifies the third testable prediction.

4.4.2 Promoting Energy Conservation

I conducted an experiment with my collaborators in Switzerland to promote energy conservation using the social mechanism [75]. Recent studies have shown that policies involving social influence can provide better outcomes than traditional institutional solutions [99]. Experiments showed that adding a descriptive normative message to energy bills, detailing average neighborhood usage, can lead to energy savings [99]. Promotion of social norms, however, also led to a boomerang effect, causing energy savers to increase their consumption. This could be eliminated, however, by providing
an injunctive message conveying social approval to energy savers. My theoretical results suggest that under very low budget, the subsidies are wasteful and a better outcome can be achieved by sharing information about consumption among the peers because it induces peer pressure. A recent experiment in energy conservation provides evidence [120].

I had two goals from this experiment. First, I wanted to verify the theoretical results in this chapter especially the prediction that under low budget for rewards, social mechanisms will perform better than the traditional mechanisms. I also want to propose moving beyond the promotion social norms, and to utilize the power of peer pressure [24] in social networks, beyond the implicit pressure induced by sharing information, to regulate consumption. This requires studying energy consumption in the context of direct peer interaction among people who share real social relationships. Earlier in the chapter, I presented a joint model of peer-pressure and externalities and introduced social mechanisms that promote reduction in externalities, indirectly through peer-pressure. I presented theoretical results about the benefits of social mechanisms over Pigouvian Mechanisms. In this section, I discuss my 20-week long experiment in Graubünden targeting 56,500 customers of an energy company. The web portal (http://munx.ch) went online on August 28th 2012.

I had 401 voluntary consumers that participated in the experiment. The population consists of 75% males and 25% females. 51% live in one-family houses and most households use oil for heating and are roughly between 100 and 200 square meters. Registered users can either form groups of up to five users or be alone. The baseline consumption for a sample of consumers was recorded from historical data. Users are rewarded 10 points for enter their meter readings once a week and points for filling surveys, that are used to purchase energy efficient products. Users can also make buddies with whom they share their consumption information and can exchange emails. Each buddy gets 5 points each week the consumer reduces her consumption over previous week. This introduces a local externality. I find that consumers who made buddies reduced consumption over previous week, more often and had much lower consumption during the experiment than the people who did not make bud-
dies. Historical data showed that statistically, there was no difference in the annual consumptions between the group of consumers who made buddies and the group of consumers who did not make buddies. The consumption decreased with temperature across the population suggesting that people were indoors more on colder days. The difference between the consumption within groups and outside groups was more on colder days. The experimental data verifies the theoretical prediction and also shows that social mechanisms could be powerful tools for energy policy makers. I now discuss the experiment.

Background

The web portal Munx.ch served as a basis for a large-scale energy efficiency campaign for residential energy customers of a Swiss utility company. Unlike in the US, there are no national policies in Switzerland yet that require utility companies to achieve a specific percentage reduction in energy use through energy efficiency measures. However, Switzerland is planning to implement policies that require utility companies to support their residential energy customers in saving on average 1.5% of their energy consumption compared to the previous year. Already now, Swiss utility companies are offering services like energy audits to help residential customers save energy. The partnering utility company in Switzerland considers Munx.ch an essential part of their energy efficiency endeavors and made the portal available to all its private customers (56500). The average electricity consumption of the residential customers of the Swiss utility is 6039 kWh per year at a cost of 0.10 Swiss Francs per kilowatt hour. The campaign aimed at raising awareness among consumers and engaging consumers in energy conservation. The registration and participation was voluntarily. To promote the web portal and motivate people to register, the utility company informed their customers via their customer magazine, local trade fairs, web advertisements and newspaper articles. The web portal was launched on August 28th 2012. Since then, 1055 customers had registered in the first twenty weeks. In addition to achieving immediate energy savings, the portal was developed to experimentally investigate socio-psychological concepts (e.g., social norms) that may help to promote
eco-friendly behavior among its users when implemented in some form of Green IS (e.g., feedback systems that built upon smart metering). To render related studies possible, the system allows for randomly assigning users to different treatment groups (i.e. experimental conditions) and for recording electricity consumption data for each user. It is thus possible to compare the effects of different interventions on energy demand.

**Functionalities of Munx.ch**

At its core, Munx.ch allows its users to collect and analyze data about their electricity consumption. Users are incentivized for their providing information (e.g. providing meter readings or filling in questionnaires) with bonus points (1 point = 0.10 Swiss Francs) that can be redeemed for energy efficient products (e.g. LED light bulbs) in an online shop connected to the web portal. The incentives were not tied to energy savings. Since no widespread smart metering infrastructure was available when the portal was launched in 2012, Munx.ch currently depends on manual data collection by its users. The users can enter the meter readings at any time. The website stores consumption data separately for day and night consumption at the household level, if information on both day and night consumption is available. To motivate the users to periodically enter their data at least once a week, the system can be configured to send automatic reminders via e-mail, and people get bonus points for doing so. Users can retrieve the historical data in the form of a table that depicts all of the meter readings entered in the past 12 months. The table further includes previous consumption levels and the associated energy efficiency level. In addition, the user gets a graphical depiction of his or her consumption history over time, which allows for comparing his or her actual consumption with prior consumption levels and to track the progress [23]. In addition to feedback on historical consumption, the users are provided with feedback on consumption efficiency in two ways. First, the weekly electricity consumption of each household is compared with the corresponding average in the user’s neighborhood. The data on the neighborhood’s consumption comes from the utility company (alongside with data on street, zip code, area). This type of
comparative feedback can be termed as descriptive normative feedback, since people learn which consumption levels are typical or ‘normal’ for their neighbors and how they compare to them [99], [29]. Second, the web portal calculates the efficiency level of each household based on the household’s consumption and the household characteristics. Each user creates a household profile upon registering with the website. This profile includes the type of building, the total size of the house or apartment, the number of people living in the household, the number of appliances, and the type of heating. Then, based on the collected raw data, the web portal calculates the weekly averages and the efficiency levels, which range from ‘A’ (high efficiency) to ‘G’ (low efficiency). This type of feedback can be regarded as injunctive normative feedback, because the efficiency level represents an indicator for social approval or disapproval [99], [29].

In addition to multiple types of consumption feedback, Munx.ch offers its users a variety of information and services to help them reduce their electricity consumption. For instance, the users may consult the web portal to learn more about the efficiency improvements that they can make by employing the newest generation of household appliances. Additionally, the users may receive simple behavioral advice (e.g., unplug electrical items before leaving for vacation). These savings tips are tailored to the characteristics of the energy consumers; tenants are not advised to insulate their houses, for example. An example of the more sophisticated services offered by Munx.ch is the ‘Appliance Oracle’. The Oracle provides the users with estimates of their weekly electricity demand based on a list of the appliances in their households and their daily average operating time. The consumption data depend on the average values, which can be modified by the user if he or she knows the exact values. Finally, Munx.ch offers a quiz on energy-related knowledge (e.g., "How much electricity does a traditional light bulb convert into light?"). All aforementioned services aim at increasing the knowledge and ability of energy consumers since these factors may facilitate or impede the enactment of energy conservation behaviors and influence individual motivations [87]. When these factors strongly inhibit or facilitate pro-environmental actions, psychological motivations are relatively unimportant [106], [107]. The tasks and the quiz aim to convey knowledge in an entertaining way.
since fun has a positive effect on the learning process. Fun reinforces intrinsic motivation [17]. In addition, quizzes have been found to motivate people to log on regularly to a website [49].

**Intervention**

In order to implement the social mechanism, I introduced a 'buddy-system'. Users can invite up to five 'buddies' via e-mail to form teams. The buddies can but do not need to be customers of the partnering Swiss utility. If a user invites a friend to be his buddy, he or she will receive an invitation including a link to Munx.ch. If he/she is already registered, he/she can see the buddy request on the Buddy-Subsite of Munx.ch. In case he/she is not yet a user of the web portal, he/she needs to register first to respond to the buddy request (reciprocity). As soon as and only if he/she responds to the user’s buddy request, they form a team. The user gets five points if the buddy manages to reduce his/her energy consumption, by any amount, compared to the previous week. She gets no points if the buddy maintains the consumption level (0% savings) or increases consumption (negative savings). The user can praise the buddy for achieved savings or motivate him to increase saving efforts by sending e-mails. The communication via e-mail is not tracked by the system. Since the user can invite up to five buddies, he/she gets five points for each buddy who manages to reduce his/her energy consumption. Hence, the user can gain 25 bonus points at most, if all five buddies were able to reduce consumption compared to the previous week. Due to reciprocity, the buddies of a user also receive five bonus points each if the user reduced consumption. However, users did not receive bonus points if they reduced consumption themselves.

**Sample and data collection**

The web portal is used by 1055 participants (74% are male, 49% between 31 and 40 years old). The buddy system was available to all users of Munx.ch since the launch on August 28th in 2012. The experiment is conducted with 401 users and comprehends two experimental conditions. The treatment group comprises users
who have at least one buddy. That requires that the users have sent out invitations to up to five friends and at least one friend has confirmed the buddy request. The control group comprises users who have sent out invitations but none of the friends has confirmed the buddy request. Hence the user does not have any buddies and is not part of any team. Naturally, users in the experiment were not randomly assigned to the experimental conditions. The experiment relied on meter readings that measure continuous behavior and are transferred manually to the web portal. In order to motivate the users to regularly transfer the meter readings, the users get 10 bonus points. Users also had the possibility to enter meter readings in retrospect, since some of them did not want to transfer a meter reading to the web portal right after having it noted down. Instead, they collected multiple readings and then transferred them all at once to the web portal by specifying the date of each meter reading. To make data transfer as simple as possible, multiple strategies are applied. First, detailed graphical and textual instructions about the location of the electricity meter and how to read it are incorporated. The utility company has used this information for many years to support its customers, who read their electricity meters for billing purposes. Second, algorithms were implemented that assessed the validity of the transferred electricity meter readings. For example, if a customer enters a negative value or a reading that is lower than the previous reading, then he or she will receive an error notification, and the value will not be saved. In the context of a former study with an Austrian utility company, it was observed that the users were honest when entering the meter readings [71]. Based on the meter readings that were entered by users of Munx.ch, weekly consumption levels and percentages of change in electricity consumption were calculated. The dependent variables for the experiment groups were tracked over 20 weeks.

Data Characteristics

Unlike the US, where heating is the primary use of electricity in homes [111], only 5% of the households in Switzerland use electricity for heat [10], so the electricity is mainly used for refrigerators, heating water, lighting, household appliances and
electronics in most of the households.

In the first twenty weeks of the experiment, 401 customers out of 1055 participants of the web portal, signed up and entered metered readings at least once. Out of them 132 customers made buddies and joined the treatment group. 208 customers entered metered readings more than once and out of these 52 were in the treatment group. I analyze data for these 208 customers. I was also able to obtain historical consumption data for the year of 2011 for 62 customers by matching their identity online to their addresses with the utility company.

Since participants volunteered to join in the experiment, an obvious question is whether we observed selection bias in the population. However, any self-selection among environmentally conscious people can be reasonably expected to apply to both the control and the treatment group. Therefore there should not be any selection bias of this type in the experiment.

Having said that, there still may be some selection bias because people in the treatment group had an intrinsic characteristic (compared to the control group) that enabled them to make buddies. In order to identify if this led to any selection bias, we studied the historical consumption of the control group and the treatment group. The average daily consumption over the entire year for the whole population was 14 kWh/day and the standard deviation was 8.19 kWh/day. This average daily consumption is consistent with the average household electricity consumption in Switzerland [108]. The treatment group had an average consumption of 0.27 kWh/day lower than the overall average but the p-value was very large (0.91) suggesting that the difference is not statistically significant. It is seen with the statistics that there is no imbalance between the control group and the treatment group in the historical consumption. Therefore, we can reasonable argue that any difference in the consumption patterns during the experiment is due to the treatment. We also studied the consumption patterns of the people before making buddies and after making buddies. As we discuss in the results section that there was no difference between the consumption patterns between the control group and the people in the treatment group before making buddies. This suggests that the observed difference between the control group and the
treatment group was purely the treatment effect.

Empirical Strategy

I want to study two effects in this experiment. Firstly, I want to verify that the social mechanism has a positive effect. Since the incentive was for the customers to pressure their buddies to reduce consumption over the past week, I would like to see if the treatment group reduced the consumption over the past week more often than the control group. Formally, the hypothesis is as follows.

Hypothesis 1. The treatment group will report reduced consumption over previous week more often than the control group.

Secondly, I want to study the effect of the treatment on average consumption. This is important because it helps us compare the treatment effect with other interventions such as pricing and normative interventions. The empirical model I use is:

\[ y_{i,t} = \alpha + \beta t + \gamma x_i + \nu_{i,t} \]

where \( y_{i,t} \) is the average hourly consumption over the entire week of the \( i \)th consumer when the average weekly temperature is \( t \), \( \alpha \) is the average baseline consumption of the population, \( \beta \) is the temperature effect, \( \gamma \) is the treatment effect, \( x_i \) is the treatment indicator and \( \nu_{i,t} \) is the estimation error.

Results

The treatment group reported reduced consumption over the previous week 30.27% of the time while the control group reported reduced consumption over the previous week only 25.23% of the time. Before making buddies, the customers reduced consumption over the previous week only 25.56% times which is similar to the control group. This suggests that the improved in the consumption behavior was clearly a treatment effect. This clearly verifies the hypothesis. The treatment led to reduction in consumption over previous weeks, 20% times more often.
Figure 4-5: Average energy consumption. The pattern of consumption correlates with the outside temperature suggesting that people spent more time indoors when it was cold. The consumption of the treatment group is lower than people in the control group.

Figure 4.4.2 shows the average consumption of the treatment, control and overall population in different weeks. It is clear that the treatment group consistently performed better than the control group. The consumption pattern is negatively correlated with the average weekly temperature of Graubünden. Even though the heating is not electric in 95% of the households, a rise in the consumption with decreasing temperature suggests that consumers spend more time indoors on colder days and thus use more electricity.

Figure 4.4.2 shows the scatter plot of the average consumption of the treatment and control groups against the average weekly temperature. I see that the con-
Figure 4-6: Average energy consumption against average weekly temperature. Consumption increases with the decrease in outside temperature suggesting that people spent more time indoors when it was cold. The consumption of the treatment group is lower than people in the control group.

Consumption is decreasing in temperature for both groups but the treatment group has consistently lower consumption than the control group.

Effect of treatment

I now discuss the total effect of the treatment. Table 4.4.2 shows the regression results when regressing the average hourly consumption over the entire week against the temperature and treatment indicator as in the empirical model. The sample size is large enough that p-values are very low suggesting that the results are statistically significant. I find that the treatment leads to 17.4% reduction in consumption over
4.4. EXPERIMENTS AND APPLICATIONS TO PUBLIC POLICY

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>StdError</th>
<th>tStat</th>
<th>p−value</th>
</tr>
</thead>
<tbody>
<tr>
<td>α</td>
<td>1.035</td>
<td>0.039</td>
<td>26.727</td>
<td>2.250e−136</td>
</tr>
<tr>
<td>β</td>
<td>−0.018</td>
<td>0.004</td>
<td>−4.3545</td>
<td>1.395e−05</td>
</tr>
<tr>
<td>γ</td>
<td>−0.180</td>
<td>0.053</td>
<td>−3.414</td>
<td>0.00065</td>
</tr>
</tbody>
</table>

Table 4.1: Regression results: $\alpha$ is the average consumption, $\beta$ is the temperature effect and $\gamma$ is the treatment effect. The p-values are very low suggesting that the regression is statistical significant.

...the control group. This effect is huge as compared to the other studies to promote energy consumption.

It is important to understand these changes in the light of user behavior. Since the reduction in consumption was short term, it is reasonable to assume that it was due to changes in behavior rather than investment in household infrastructure and energy efficient appliances. The treatment effect is 4.32 kWh/day. This is significant reduction in the consumption and cannot be attributed to one responsible change in behavior. This could be due to all around changes. Reducing the use of 5 gallons of water each day reduces the consumption by 1 kWh/day. Using dishwashers once in three days, when it is full, instead of daily, saves on an average 0.7 kWh/day. Increasing the temperature of your refrigerator from 2°C to 7°C saves on an average 0.5 kWh/day. Putting appliances on standby in a four person household instead of always on saves on an average 1 kWh/day. Reducing the use of unnecessary 40 watt light bulbs for 30 hours or reducing the use of 3 light bulbs on an average every day would save on an average 1.2 kWh/day. This overall change in behavior could achieve the observed effect of reduction in consumption. Although, the infrastructural changes can obtain the effect by the purchase of energy efficient appliances and led lights but it is often not reasonable to assume these changes happening in short time scales. We now compare the treatment effect as compared to other studies.

**Effect compared to pricing and normative effects**

I do not have information about the elasticity of the electricity demand with respect to price in Switzerland. However, in [94] showed that the 60 day elasticity of electricity
demand with respect to price in California is -0.10 to -0.18. If the results from [94] could be generalized with caution, then the effect of the treatment (17.4% reduction in demand) is equivalent to a 97% to 174% increase in price in the short term. The authors in [93] estimate that the long-run elasticity of residential electricity demand in California is 0.39. Therefore the treatment effect is equivalent to 45% long term increase in the price. This is extremely significant and shows the power of the social mechanism and peer pressure. The US Energy Information Administration estimates that a recently-proposed carbon cap-and-trade program would increase electricity prices by 2.5% in 2020 and 20% in 2030 [112]. This further accentuates the importance of the social mechanisms and its advantages over pricing mechanisms for climate change policies to regulate carbon dioxide emissions.

On the other hand, the normative effects observed in the OPOWER’s several experiments were 2.1% [5] on an average. Even though the cost of OPOWER’s program is very low, the effect is also significantly small as compared to the social mechanism.

Cost effectiveness of the mechanism

Cost effectiveness of a mechanism is important because it is how energy conservation programs are evaluated. This cost is different from the social welfare effects that are hard to measure. The cost of the experiment, other than administrative costs, that is equivalent to other studies is the cost of rewards distributed to the consumers. We avoid the cost of the website because it is considered a fixed cost that does not increase with participation. Since the rewards were distributed only when a buddy reduces consumption over the previous week, the cost incurred is:

\[
Total \ cost = \sum_{i \in \ Treatment} D_i \times r \times B_i
\]

where \(D_i\) is the number of times consumer \(i\) reduces consumption (from one week to the next), \(r\) is the reward, and \(B_i\) is the number of buddies that \(i\) managed to recruit.

This is because the treatment effect is the daily effect. The total cost of rewards
incurred in the experiment was 96.50 Swiss Francs. The cost effectiveness is:

\[
\frac{\text{Total cost}}{E \times T \times N}
\]

where \( E \) is the average treatment effect per day, \( T \) is the duration of treatment in days, and \( N \) is the size of the treatment group. The cost effectiveness for the experiment is 0.31 Swiss rappen or centime per kWh saved. This is significantly lower than the cost effectiveness of OPOWER’s programs (3.31) US cents per kWh saved.

We mention cost effectiveness with caution because it does not completely capture the social welfare effect. Although, it is hard to measure the overall welfare effect, recent work has shown that there may be a ‘moral licensing’ effect and consumers who reduce consumption of one resource, increased consumption of other resource [110].

## 4.5 Conclusions

I presented a joint model of externalities and peer interactions in social networks. I then introduced tools for policy-makers to promote cooperation in situations where individual free-riding leads to socially inefficient outcomes. The social mechanism encourages peers to utilize their ability to exert pressure locally in order to promote cooperation globally. I discussed two field experiments that verify the theoretical predictions. The field experiments also demonstrated the power of social mechanisms for health and energy policy.

A key advantage of my approach is that the precise means by which pressure is exerted is left to the individuals in the network, and thus works at the normative level [58]. To quote Nobel Prize winner Elinor Ostrom: “International donors and nongovernmental organizations, as well as national governments and charities, have often acted, under the banner of environmental conservation, in a way that has unwittingly destroyed the very social capital – shared relationship, norms, knowledge and understanding – that has been used by resource users to sustain the productivity
of natural capital over the ages” [89]. Building on Ostrom’s sentiment, by adjusting the incentives at the local level, my peer-based incentive mechanism leverages existing social capital and local regulatory mechanisms, rather than bypassing (or even suppressing) them.

I introduced one kind of social mechanism that localizes externality in social networks. I discuss two other possible mechanisms below.

1. Since we observed that only few agents exert peer-pressure on their peers, one way to create a social mechanism is to reward only one chosen peer. There is experimental evidence that, in small groups, designating a solitary punisher can promote cooperative behavior [9, 86]. This reduces the budget requirement for the social mechanism by eliminating second-order free riding [103], in which one peer leaves it to another peer to exert the costly peer-pressure. Obviously, this mechanism denies equal opportunity to all peers to earn rewards and requires accurate information about peer-interaction.

2. Another way to create a social mechanism is to reward agents not just for their immediate peers’ actions, but also their peers’ peers or the actions of agents that are within \( n \) hops away. Such mechanisms achieve lower localization of externality, but keep a check on second-order free riding. They also avoid the formation of coalitions among immediate peers (if utilities are transferable among the peers) that may undermine the effect of the social mechanism.

In a recent review of data-centric uses of social networks to accelerate behavior change and improve collective performance, Valente identified four major classes of network intervention [113]: (i) targeting influential individuals to use them as conduits for desirable behavior; (ii) segmentation and targeting of groups with particular network characteristics; (iii) altering (or designing) the network structure; and (iv) induction of particular peer-to-peer interaction. The theoretical underpinnings of the first three intervention classes have received much attention: influential individual identification [6, 63, 70], characteristic segmentation of groups [83], and network
structure design [54]. This work fills an important gap by putting the induction of peer-to-peer influence on solid theoretical foundation.

My work opens up several new theoretical questions and provides opportunity for investigation through large scale experiments. Firstly, what are the limits on the cooperation that can be achieved through peer-pressure? The ability to exert peer-pressure relies on existing exchanges between people, and therefore has limits, beyond which the peer-relationship may break. Secondly, how do the mechanisms change when the marginal cost of exerting peer-pressure is not constant? The loss in utility to one person for reducing externality can be compensated by changing the exchanges in the relationship, causing the person exerting pressure to bear some loss in the exchange. This loss is the basis of the cost of exerting peer-pressure, and therefore the marginal cost of exerting peer-pressure may not be constant. Finally, how do various social mechanisms differ in the dimensions of budget requirements, social cost, risk of non-cooperative action due to the lack of peer interaction, noisy information about the social network, or the risk of undermining cooperation through collusion?

The experiments I conducted were first of their kind and the success of the experiments suggest that social mechanisms must be explored from the public policy perspective. The results open up opportunities for investigation through large-scale experiments for in the domains of energy conservation, recycling, and promoting healthy behavior.
Chapter 5

Conclusions and Future Work

In this dissertation we explored the area of bilateral exchanges in networks and in particular extended the stable matching problem in bipartite networks to the general scenario where agents may not be able to costless transfer utilities among themselves. We studied a general case when the payoff is continuous and strictly increasing in the part of the split and introduced the Generalized Hungarian algorithm to find the stable matching that is optimal for one side of the network. We also characterized the set of stable matchings for a special class of bipartite networks. With additional structure on the payoff functions and correlations between payoff functions, more efficient methods can be employed and is an interesting line of future work. Our algorithm can be an important tool in the study of generalized matching problem. We also introduced natural dynamics that converge to approximately stable matchings. We also gave an axiomatic characterization of fair outcomes in bipartite matching under the special case of costless transfers based upon the collective bargaining argument. We also presented an algorithm to compute the fair outcome and showed that the fair outcome coincides with the unique nucleolus of the assignment game.

We also presented a joint model of externalities and peer interactions in social networks. We then introduced tools for policy-makers to promote cooperation in situations where individual free-riding leads to socially inefficient outcomes. The social mechanism encourages peers to utilize their ability to exert pressure locally in order to promote cooperation globally. We discussed two field experiments that
verify the theoretical predictions. The field experiments also demonstrated the power of social mechanisms for health and energy policy.

In a recent review of data-centric uses of social networks to accelerate behavior change and improve collective performance, Valente identified four major classes of network intervention [113]: (i) targeting influential individuals to use them as conduits for desirable behavior; (ii) segmentation and targeting of groups with particular network characteristics; (iii) altering (or designing) the network structure; and (iv) induction of particular peer-to-peer interaction. The theoretical underpinnings of the first three intervention classes have received much attention: influential individual identification [6, 63, 70], characteristic segmentation of groups [83], and network structure design [54]. This work fills an important gap by putting the induction of peer-to-peer influence on solid theoretical foundation.

Our work opens up several new theoretical questions and provides opportunity for investigation through large scale experiments. The work on matching can be extended to general networks. General networks need not have stable matching. In future work, we will characterize the set of networks that admit stable matchings and provide a decomposition of the set of edges in such networks. We will also work on algorithms for computing stable matchings in general networks that admit a stable matching. Another important line of investigation is the welfare properties of stable matchings and the duality gap between the optimal matchings and stable matchings. Finally, we wish to apply our results to the study of family economics. In future, we want to study distributed dynamics that convert to fair outcomes. Secondly, we want to extend our results to the general case of costly transfers. The lessons from the work on exclusive exchanges will also help in the study of non-exclusive exchanges and intermediation and arbitrage in networks.

The work on the localization of externalizes in networks raises several new questions and opportunity for future work. Firstly, what are the limits on the cooperation that can be achieved through peer-pressure? The ability to exert peer-pressure relies on existing exchanges between people, and therefore has limits, beyond which the peer-relationship may break. Secondly, how do the mechanisms change when the
marginal cost of exerting peer-pressure is not constant? The loss in utility to one person for reducing externality can be compensated by changing the exchanges in the relationship, causing the person exerting pressure to bear some loss in the exchange. This loss is the basis of the cost of exerting peer-pressure, and therefore the marginal cost of exerting peer-pressure may not be constant. Finally, how do various social mechanisms differ in the dimensions of budget requirements, social cost, risk of non-cooperative action due to the lack of peer interaction, noisy information about the social network, or the risk of undermining cooperation through collusion? The experiments we conducted were first of their kind and the success of the experiments suggest that social mechanisms must be explored from the public policy perspective. The results open up opportunities for investigation through large-scale experiments for in the domains of energy conservation, recycling, and promoting healthy behavior.
Bibliography


