Essays on the Computation of Economic Equilibria and Its Applications

by

Ye Du

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Computer Science and Engineering)
in The University of Michigan
2009

Doctoral Committee:

Associate Professor Yaoyun Shi, Chair
Professor Romesh Saigal
Professor Michael P. Wellman
Assistant Professor Rahul Sami

ACKNOWLEDGEMENTS

This dissertation would not have been possible without the support and help of many people. First of all, I would like to thank my advisor Yaoyun Shi, who was responsible to bring me to the University of Michigan five years ago. Yaoyun is the most open minded advisor I can imagine. From the very beginning of my PhD journey, he encouraged me to pursue any research topics that I felt really fascinated about. With his supports, I chose algorithmic game theory as my main research area. Although it is significantly different from quantum computation and information, Yaoyun was always willing to listen to my research and provided many encouragements when I got stuck.

I first met Xiaotie Deng at Banff, Canada, 2007. He introduced me to the fascinating area of the computational general equilibrium theory. This is the cornerstone in my PhD. In fall 2007, I visited him at the CityU of Hong Kong. I really enjoyed the chatting with him, which provided me tons of inspirations and interesting ideas. I also want to thank him for inviting me to serve as referee for a few conferences and journals.

I am also grateful to Romesh Saigal, Rahul Sami, and Mike Wellman for kindly serving in my thesis committee. I learned a lot about optimization from Romesh. He can always give intuitive explanations for the complex techniques and concepts in linear and nonlinear programming. With Rahul, I got to know the area of mechanism design and worked out my first paper in algorithmic game theory. Mike is very knowledgable in AI and computational economics. I took my AI course with him and he gave me many valuable suggestions when I needed them.

I feel very fortunate to work with many brilliant people: Xi Chen, Decheng Dai, James Leung, Shanghua Teng, and Xin Zhao. I also would like to thank my colleagues here in the EECS department, Ran Duan, Wei Huang, Yi Li, Xiaolin Shi, Denny VandenBerg, Xiaodi Wu, Yufan Zhu. They made the theory group a wonderful place to stay for five years.

Confucius says "You should not go on a long journey when your parents are still alive; if you have to do so, you must have convincing reasons". I owe a lot to my parents, who give me infinite supports. I also hope that my PhD journey is worth their loves.

This dissertation is partially supported by NSF under Grant 0347078.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS . . . . **ii**

LIST OF FIGURES . . . . **vi**

ABSTRACT . . . . **vii**

CHAPTER

I. Introduction 1

1.1 History and Motivations 1
1.2 Contributions of this Dissertation 5
1.3 Organization 7

II. Background and Preliminaries 9

2.1 Some Basic Topology and Fixed Point Theorems 9
2.2 Nash Equilibrium 11
2.3 Competitive Equilibrium 14
2.4 Discrete Markov Chain 16
2.5 PPAD 17

III. The Computational Complexity of Competitive Equilibrium 19

3.1 Approximate Competitive Equilibrium 19
3.1.1 Motivations 19
3.1.2 The Main Theorem 20
3.2 Equilibria in Markets with Additively Separable Utility Functions 25
3.2.1 Motivations 25
3.2.2 The Main Theorem 26
3.3 Related Works 30

IV. On the Complexity of Deciding Degeneracy in Games 32

4.1 Motivations 32
4.2 Degeneracy in Linear Programming and in Bimatrix Games 33
4.3 The Main Theorem 34

V. Path Auctions with Multiple Edge Ownership 42

5.1 Motivations 42
5.2 Definitions and Problem Statement 44
5.2.1 Simple path auctions 44
5.2.2 Extended path auctions 45

5.3 The Nonexistence of Individually Rational Strategyproof Mechanisms 47
5.4 Existence of a Pareto Efficient Pure Strategy ϵ-Nash Equilibrium 54
5.5 Related works 59

VI. Using Spam Farms to Boost PageRank 61

6.1 Motivations 61
6.2 Single-Target Spam Farm Model 62
6.3 Our Results 63
6.4 The PageRank Algorithm 63
6.5 Sensitivity Analysis of Markov chain 64
6.6 Characterization of an Optimal Spam Farm 66
6.7 Optimal Spam Farm Under Constraints 74
6.8 Related Works 78

VII. Ranking via Arrow-Debreu Equilibrium 80

7.1 Motivations 80
7.2 Preliminaries 81
7.2.1 The Ranking Problem 81
7.2.2 Arrow-Debreu equilibrium of exchange markets 82
7.2.3 CES Utility Functions 82
7.3 PageRank/Invariant Method V.S. a Cobb-Douglas Market 83
7.4 Ranking via Arrow-Debreu Equilibrium 86
7.5 Conclusion and Future Works 91

VIII. Conclusion 93

BIBLIOGRAPHY **95**

LIST OF FIGURES

Figure

3.1	The reduction idea	27
5.1	VCG mechanism is not strategyproof for this game	44
5.2	No false-name-proof mechanism that satisfies Pareto efficiency	53
5.3	Auction with no Pareto efficient pure-strategy ϵ-Nash equilibrium	57
6.1	Optimal spam farm	67
6.2	Optimal spam farm when the target page points to non generous pages	75
6.3	Optimal spam farm when some boosting page cannot point to the target page	76
6.4	Optimal spam farm when the hijacked pages cannot point to the target page	77

ABSTRACT

The computation of economic equilibria is a central problem in algorithmic game theory. In this dissertation, we investigate the existence of economic equilibria in several markets and games, the complexity of computing economic equilibria, and its application to rankings.

It is well known that a competitive economy always has an equilibrium under mild conditions. In this dissertation, we study the complexity of computing competitive equilibria. We show that given a competitive economy that fully respects all the conditions of Arrow-Debreu's existence theorem, it is PPAD-hard to compute an approximate competitive equilibrium. Furthermore, it is still PPAD-Complete to compute an approximate equilibrium for economies with additively separable piecewise linear concave utility functions.

Degeneracy is an important concept in game theory. We study the complexity of deciding degeneracy in games. We show that it is NP-Complete to decide whether a bimatrix game is degenerate.

With the advent of the Internet, an agent can easily have access to multiple accounts. In this dissertation we study the path auction game, which is a model for QoS routing, supply chain management, and so on, with multiple edge ownership. We show that the condition of multiple edge ownership eliminates the possibility of reasonable solution concepts, such as a strategyproof or false-name-proof mechanism or Pareto efficient Nash equilibria.

The stationary distribution (an equilibrium point) of a Markov chain is widely used for ranking purposes. One of the most important applications is PageRank, part of the ranking algorithm of Google. By making use of perturbation theories of Markov chains, we show the optimal manipulation strategies of a Web spammer against PageRank under a few natural constraints. Finally, we make a connection between the ranking vector of PageRank or the Invariant method and the equilibrium of a Cobb-Douglas market. Furthermore, we propose the CES ranking method based on the Constant Elasticity of Substitution (CES) utility functions.

CHAPTER I

Introduction

1.1 History and Motivations

An equilibrium point, generally referred to as a balanced situation, is a central solution concept for complex systems. In particular, the economic equilibrium is extensively studied in microeconomic theory. There are two major equilibrium concepts in microeconomics: *Nash equilibrium* [62] and *competitive equilibrium* [7]. It is well known that under mild conditions, either a Nash equilibrium or a competitive equilibrium always exists, which lays the foundations for the *game theory* and *general equilibrium theory* respectively.

A more general concept than economic equilibria is the *Kakutani's fixed point* x^* such that $x^* \in \Psi(x^*)$, where Ψ is a correspondence. The existence of Nash equilibria, proved by Nobel laureate John Nash [62], and the existence of competitive equilibrium, proved by Nobel laureates Kenneth Arrow and Gerard Debreu [7], were shown via Kakutani's fixed point theorem. However, both proofs are nonconstructive.

In general, the power of economic equilibrium theories depends on their analytical and predictive power on economic policies. Shoven and Whalley [74] investigated the applications of general equilibrium theory to analyze tax, trading, and price-control policies. As argued by Kamal Jain [48], "If a Turing machine cannot compute [an equilibrium] then an economic system cannot compute [it] either. Hence, a lack of computational result decreases the applicability of equilibrium concept itself." The computation of equilibria is essential to make the theories meet practice.

With respect to the computational aspects of fixed points, people ask questions related to the computability, computational complexity, and polynomial time computability of economic equilibria. Unfortunately, Wong and Richter [81] showed that Brouwer's fixed points are uncomputable in the strong real computing model. Hence, a competitive equilibrium is not computable in that sense. On the other hand, Nash equilibria of a finite matrix game can be computed to any arbitrary precision we like [56]. However, for a three-player game, even if the payoff matrices are all rational, an equilibrium of the game can be irrational. Blum, Cucker, Shub, and Smale [9] studied the real computational complexity, which allows a solution of a problem to be irrational. On the other hand, irrationality is not proper for the standard Turing machine model, which permits only computational problems with finite size input and output. Moreover, even if an economic equilibrium is computable, a decision-maker may lack the resources to reach an optimal solution. This results in the decision-maker seeking for a satisfactory solution rather than an optimal one, and is referred to as *bounded rationality*. Hence, certain approximation concepts must be introduced. Usually, two possible approximate solution concepts are used in economic analysis: *weak approximation*, which approximates certain function value of an exact solution, and *strong approximation*, which approximates an exact solution itself geometrically.

Since the introduction of these approximate solution concepts, a few great results have been achieved regarding the computation of economic equilibria. In his seminal paper [68], Papadimitriou introduced a new complexity class: Polynomial Parity

Argument Directed case (PPAD). It is unlikely that there is a polynomial time algorithm for the PPAD-Complete problems [65]. On the complexity side, Etessami and Yannakakis showed [35] that it is FIXP-Complete to compute a strong approximation of Nash equilibrium. Daskalakis, Goldberg, and Papadimitriou [23] showed that it is PPAD-Complete to compute a weak approximate equilibrium of four-player games. Chen and Deng later extended the PPAD-Completeness result to two-player games [14] and $\frac{1}{n^{\Theta(1)}}$-approximate Nash equilibrium [15]. Moreover, Codenotti *et al.* [21] showed that the computation of competitive equilibrium for a special class of pairing Leontief economy is PPAD-hard. On the algorithmic side, Lemke and Howson [55] developed an algorithm to compute Nash equilibria of bimatrix games. Scarf [71] further designed a more general, discrete algorithm to compute economic equilibria. In addition, Eaves [31] and Eaves & Saigal [32] proposed homotopy methods to compute fixed points. Moreover, based on the proof of Brouwer's fixed point given by Hirsch [44], Kellogg, Li, and Yorke [51] designed a method of continuation to compute Brouwer's fixed points. Unfortunately, none of those algorithms can guarantee to find a fixed point in polynomial time. The state-of-art about the computation of economic equilibria is: for Nash equilibria of bimatrix games, a 0.3393-approximate Nash equilibrium can be computed in polynomial time [78]; for competitive equilibria, polynomial time algorithms are only available for those economics with convex equilibria sets [48, 19, 20].

On the other hand, although Nash equilibrium is an important solution concept in game theory, it has several shortcomings. In particular, a game may have multiple Nash equilibria points. Moreover, a Nash equilibrium may be difficult to reach and may not be Pareto efficient. However, as argued by Nobel laureate Eric Maskin [58], these drawbacks are less severe if a Nash equilibrium is the outcome of a specifically designed mechanism. This is the "engineering" part of economic theories. A mechanism designer starts with a specific objective and designs an economic mechanism such that the outcome of it indeed meets the objective when rational agents play by its rules.

In general, a mechanism has two elements: an *allocation rule*, which maps strategies of players in the economic system to an outcome, and a *pricing rule*, which either charges or pays a certain price to each player. A widely studied solution concept of a mechanism is called the dominant strategy, where each player makes his decision without respect to the strategies of the other players. By definition, a dominant strategy is a stronger economic equilibrium than a Nash equilibrium. A desirable property in mechanism design is strategyproofness, under which the best strategy of each player is to reveal its true information. A well known strategyproof mechanism is the VCG mechanism [49, 65]. With the advent of the Internet and e-commerce applications such as eBay, an agent can easily have access to multiple accounts. He may further improve his own payoff by manipulating his multiple identities. Thus, it is important to study whether strategyproof mechanisms or Nash equilibria are reasonable solution concepts under the constraint that each agent can have multiple identities.

On the application side, ranking, which aggregates the preferences of individual agents over a set of alternatives, is an important ingredient of a reputation system. The stationary distribution – an equilibrium point – of a Markov chain is widely used for ranking purposes, such as PageRank [12] and the Invariant Method [77]. PageRank is part of the Web ranking algorithm used by Google. In Web ranking, Web pages with high rankings probably have great economic values. Thus, Web users may have incentives to manipulate the ranking algorithms in order to boost the rankings of certain pages. Those Web users are called *Web spammers*. From the perspective of game theory, it is worthwhile to study the best manipulation strategies of Web spammers. In addition, on the theoretical aspects, a few works [66, 3, 77, 11, 4] provide mathematical characterizations for those ranking methods. It would be of great interest to further investigate the economic foundations of ranking methods.

This dissertation is devoted to studying the computation of economic equilibria and its applications. Specifically, we will investigate the existence of economic equilibria in several markets and games, the complexity of computing economic equilibria, and its application to rankings.

1.2 Contributions of this Dissertation

The computation of a competitive equilibrium is one of the most important problems in algorithmic game theory. Codenotti *et al.* [21] showed that the competitive equilibria of a special class of pairing Leontief economy has a one-to-one correspondence with the Nash equilibria of a bimatrix game. However, this special class of paring Leontief economy violates one specific condition of Arrow-Debreu's existence theorem [7], which requires that the initial endowment of each good for each individual is strictly positive. Together with Xiaotie Deng [24], we proved that given a competitive economy that fully respects all the conditions of Arrow-Debreu's existence theorem, for any positive constant $h > 0$, it is PPAD-hard to compute a $\frac{1}{n^h}$-approximate competitive equilibrium. This is the first complexity result about the economies that strictly satisfy the conditions of Arrow-Debreu's existence theorem.

The complexity result for economies with the Leontief utility function, which is nonseparable piecewise linear concave, is rather limited. The computation of equilibria for economies with additively separable concave utility functions is an important open problem [65]. In working with Xi Chen, Decheng Dai, and Shanghua Teng [13], we reduced Nash equilibria of bimatrix games to competitive equilibria of economies with additively separable piecewise linear concave (PLC) utility functions. Hence, it is PPAD-Complete to compute an equilibrium for economies with additively separable PLC utility functions. The result also implies that it is PPAD-hard to compute a quasi-equilibrium of an economy.

Degeneracy is an important concept in game theory. As we know, a nondegenerate bimatrix game has an odd number of equilibria [65]. However, a degenerate bimatrix game may contain an infinite number of Nash equilibria points. In [27], we showed that it is NP-Complete to decide whether a bimatrix game is degenerate while it is Co-NP-hard to decide whether it is nondegenerate. However, for a win-lose bimatrix game, it is in polynomial time to decide whether it is degenerate.

We studied a specific mechanism design problem: the path auction game. In the path auction game, there is a network $G = (V, E)$, in which each edge $e \in E$ is owned by an agent. The true cost of e is private information and known only to the owner. Given two vertices, source s and destination t, the auctioneer's task is to buy a path from s to t. The path auction game can be used to model problems in supply chain management, transportation management, QoS routing, and other domains. With the emergence of e-commerce, a bidder can easily have access to multiple accounts. Hence, we studied the path auction games with multiple edge ownership. Together with Rahul Sami and Yaoyun Shi [29], we showed that, assuming that edges not on the winning path always get 0 payment, there is no individually rational, strategyproof mechanism in which only edge costs are reported. If the agents are asked to report costs as well as identity information, we showed that there is no

Pareto efficient mechanism that is false-name-proof. We then studied a first-price path auction in this model. We showed that, in the special case of parallel-path graphs, there is always a Pareto efficient pure strategy ϵ-Nash equilibrium in bids. However, this result does not extend to general graphs—we constructed a graph in which there is no Pareto efficient pure strategy ϵ-Nash equilibrium.

In Web ranking, highly ranked Web pages potentially have great economic advantages. Thus, Web spamming, which manipulates the ranking of a search engine, becomes a great challenge for ranking algorithms. Web spamming threatens the fairness and accuracy of search engines. In the joint work with Yaoyun Shi and Xin Zhao [30], we studied manipulation strategies under PageRank in the single target spam farm model. By making use of perturbation theories of Markov chains, we analyzed the optimal manipulation strategies of a Web spammer under a few natural constraints.

Both the existence of the ranking vector of the PageRank or Invariant method and the existence of competitive equilibrium can be shown via the Brower's fixed point theorem. In [28], we make a connection between the ranking theory and the general equilibrium theory. We show that the ranking vector of PageRank or Invariant is indeed the equilibrium of a Cobb-Douglas economy. Furthermore, we interpret a link in a reference graph as a "demand." Based on that, we propose a new ranking method, the *CES ranking*, which is minimally fair, strictly monotone and invariant to reference intensity, but not uniform or weakly additive.

1.3 Organization

The organization of this dissertation is as follows: Chapter II introduces some important definitions and theorems. Chapter III studies the computation of competitive equilibria. Chapter IV investigates the degeneracy in games. Chapter V studies the path auction games with multiple edges ownership. Chapter VI analyzes the manipulation strategies of Web spammers. Chapter VII establishes a connection between ranking theory and general equilibrium theory. Finally, we conclude in Chapter VIII.

CHAPTER II

Background and Preliminaries

2.1 Some Basic Topology and Fixed Point Theorems

In this section, we review several versions of fixed point theorems, but omit proofs. Those who are interested can refer to [37, 49] for details. First of all, we introduce some basic definitions in topology.

Definition 2.1. [49] Let (X, d) be a metric space. The open ϵ-ball with center x_0 and radius $\epsilon > 0$ is a subset of points in X :$B_\epsilon(x) \equiv \{x \in \mathbb{R}^n | d(x_0, x) < \epsilon)\}$.

Definition 2.2. [37] Let (X, d) be a metric space. A set S in X is open if for every $x \in S$, there exists an open ball centered at x that is contained in S, that is, $\forall x \in S, \exists \epsilon > 0$, such that $B_\epsilon(x) \subseteq S$. A set C in X is closed if its complement is open.

Definition 2.3. [49] Let (X, d) be a metric space. A set S in X is bounded if there exists some $\epsilon > 0$ such that $S \subseteq B_\epsilon(x)$ for some $x \in X$.

Definition 2.4. [49] [(Heine-Borel) Compact Sets] A set S in $\mathbb{R}^n$ is called compact if it is closed and bounded.

Definition 2.5. [37] A set S in $\mathbb{R}^n$ is convex if given any two points x_1 and x_2 in S, for every $\lambda \in [0, 1]$, the point $x_\lambda = (1 - \lambda)x_1 + \lambda x_2$ is also in S.

Definition 2.6. [37] The function $f : S \subseteq \mathbb{R}^n \to \mathbb{R}$ is convex if given any two points x_1 and x_2 in S, for every $\lambda \in [0, 1]$, we have $f(x_\lambda) \leq (1 - \lambda)f(x_1) + \lambda f(x_2)$.

There is a nice characterization of convex functions if f is twice differentiable and the domain S of f is an open set. Let us define the Hessian matrix $\nabla^2 f$ of f to be $(\nabla^2 f)_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}$. Then,

Theorem 2.7. *[10] The function f is convex iff S is convex and its Hessian is positive semidefinite for all $x \in S$.*

A function f is concave if $-f$ is convex. Next, we introduce a weaker concept than concavity.

Definition 2.8. [37] The function $f : S \subseteq \mathbb{R}^n \to \mathbb{R}$ is quasiconcave if given any two points x_1 and x_2 in S, we have $\forall \lambda \in [0, 1]$, $f(x_\lambda) \geq \min\{f(x_1), f(x_2)\}$.

Now we are ready to present the Brouwer's fixed point theorem.

Theorem 2.9. *[37] [Brouwer's fixed point theorem] Let $f : S \to S$ be a continuous function mapping a compact and convex set into itself. Then there exists at least one point $x^* \in S$ such that $f(x^*) = x^*$, where x^* is called a Brouwer's fixed point of f.*

In the next theorem, the fixed point theorem is extended to correspondences (set valued functions). Let us extend the definitions of function and continuity to correspondences first.

Definition 2.10. [37] A correspondence from S to Y is a function that maps each element x of the set S to a subset of Y.

Definition 2.11. [37] Let S and Y be finite dimensional Euclidean spaces, and let $\Psi : S \to\to Y$ be a correspondence. Then:

(1). Ψ is upper-hemicontinuous(uhc) at a point $x \in S$ if for every open set V containing $\Psi(x)$, there exists a neighborhood U of x such that $\Psi(x_1) \subseteq V$ for every $x_1 \in U$.

(2). Ψ is lower-hemicontinuous(lhc) at a point $x \in S$ if for every open set V in Y with $\Psi(x) \bigcap V \neq \emptyset$, there exists a neighborhood U of x such that $\Psi(x_1) \bigcap V \neq \emptyset$ for every $x_1 \in U$.

(3). Ψ is continuous at x if it is both uhc and lhc at this point.

Now we are ready to present Kakutani's fixed point theorem.

Theorem 2.12. *[37] [Kakutani's fixed point theorem] Let Ψ be a correspondence from a set $S \subseteq \mathbb{R}^n$ to itself. Assume that S is compact and convex, and Ψ is upper-hemicontinuous, nonempty, compact, and convex-valued for all $x \in S$. Then $\exists x^* \in S$ such that $x^* \in \Psi(x^*)$, where x^* is called a Kakutani's fixed point of Ψ.*

2.2 Nash Equilibrium

In this section, we review the definitions and some elementary results of the non-cooperative game theory by following the textbook [49].

Definition 2.13. [49] A strategic form game is a triple $G = (I, S_i, u_i)_{i=1}^N$, where $I = \{1, ..., N\}$ is the set of players, S_i is the set of strategies available to player $i \in I$, and $u_i : S = \prod_{j=1}^N S_j \to \mathbb{R}$ describes player i's payoff as a function of the strategies chosen by all players. A strategic form game is finite if each player's strategy set contains a finite number of elements.

A pure strategy profile $s = \{s_1, ..., s_N\}$ is an element of S, which specifies the strategy chosen by each player. Moreover, following standard notation in game theory, let s_{-i} be the vector $\{s_1, ...s_{i-1}, s_{i+1}, ..., s_N\}$, which specifies the strategies chosen by all players except player i. Similarly, we can define $S_{-i} = \prod_{j=1,j\neq i}^{N} S_j$. In general, a player can play the strategies available to him randomly. Hence, a mixed strategy for player i is a probability distribution over S_i. Let M_i denote the set of mixed strategies of player i and $M = \prod_{j=1}^{N} M_i$ denote the set of mixed strategy profiles. We extend the utility function over pure strategies to mixed strategies. Given a mixed strategy $m \in M$ that is a distribution over pure strategy space S, $u_i(m) = \sum_{s\in S} u_i(s)m(s)$, where $m(s)$ is the probability of playing pure strategy s under m. Given a strategic form game, we define two types of solution concepts—*dominant strategy* and *Nash equilibrium*—below.

Definition 2.14. [49] A strategy m_i for player i is (weakly) dominant if $u_i(m_i, m_{-i}) \geq u_i(m_i^{'}, m_{-i})$ for all $m_i^{'} \in M_i$ and for all $m_{-i} \in M_{-i}$. A strategy profile m is a dominant strategy profile if for each player i, m_i is a dominant strategy.

Definition 2.15. [49] Given a finite strategic form game $G = (I, S_i, u_i)_{i=1}^{N}$, a strategy profile $m \in M$ is a Nash equilibrium if for every player i, $u_i(m_i, m_{-i}) \geq u_i(m_i^{'}, m_{-i})$ for any $m_i^{'} \in M_i$.

It is obvious that a dominant strategy profile is indeed a Nash equilibrium. The concept of *Nash equilibrium* was introduced by Nash in his seminal paper [62]. Moreover, he showed that:

Theorem 2.16. *[62, 49] Every finite strategic form game has at least one Nash equilibrium.*

Note that if players have infinite strategy space, a strategic form game may not have an equilibrium [70]. Furthermore, even if a game has an equilibrium, it may not be a pure strategy Nash equilibrium. The famous rock-scissor-paper game is an example of a game that only has a mixed strategy Nash equilibrium. Furthermore, in a Nash equilibrium, players may not achieve optimal utilities. Actually, a Nash equilibrium can make players achieve the worst possible utilities, which is shown in the famous *prisoners' dilemma* [49]. Therefore, Nash equilibrium may not be a perfect solution concept in game theory. In the section that follows, we will introduce some weaker solution concepts than Nash equilibrium.

First of all, we introduce the concept of approximate Nash equilibrium.

Definition 2.17. Given a finite strategic form game $G = (I, S_i, u_i)_{i=1}^N$, a strategy profile $m \in M$ is an ϵ-approximate Nash equilibrium if for each player i, $u_i(m_i, m_{-i}) \geq u_i(m_i^{'}, m_{-i}) - \epsilon$ for any $m_i^{'} \in M_i$.

In a Nash equilibrium, each player plays his strategy independently according to the mixed strategy chosen by himself. We can generalize the concept of Nash equilibrium by allowing players to play correlated mixed strategies.

Definition 2.18. [67] Let x be a distribution over S. The probability distribution x is a correlated equilibrium if for any player i and any strategies $p, q \in S_i$, $\sum\limits_{s_{-i} \in S_{-i}} [u_i(p, s_{-i}) - u_i(q, s_{-i})] x_{ps_{-i}} \geq 0$, where $x_{ps_{-i}}$ is the probability that player i plays strategy p while other players play s_{-i}.

It is obvious that a mixed strategy Nash equilibrium is also a correlated equilibrium.

At last, we introduce a special case of noncooperative strategic games: the Bimatrix Games. Bimatrix games are the simplest types of normal form games. In a bimatrix game there are two players. One is called the *row player* and the other is called the *column player* with pure strategy spaces $\mathcal{R}$ and $\mathcal{C}$ respectively. We can use two matrices, A and B, to represent the payoff matrices to the row player and the column player when they play different combinations of pure strategies. Specifically,

A_{ij} is the payoff to the row player when he plays its ith pure strategy while the column player plays its jth pure strategy; vice versa for matrix B. A mixed strategy is a probability distribution on the strategy space. Let $\Delta_m = \{(x_1, ..., x_m) | \sum_i x_i = 1$ and $\forall i, x_i \geq 0\}$ and $\Delta_n = \{(y_1, ..., y_n) | \sum_j x_j = 1$ and $\forall j, y_j \geq 0\}$ be the mixed strategy spaces of the row and the column player.

Definition 2.19. [14] Given the payoff matrices (A, B) of the row and the column players, a strategy profile (x^*, y^*) is a Nash equilibrium iff $\forall x \in \Delta_m$, $(x^*)^T Ay^* \geq x^T Ay^*$ and $\forall y \in \Delta_n$, $(x^*)^T By^* \geq (x^*)^T By$.

2.3 Competitive Equilibrium

In this section, we will define the competitive economy in a general sense according to Arrow-Debreu [7] and review the existence theorem of competitive equilibrium [7]. A competitive economy can be represented as $E = \{(i)_{i\in\mathcal{I}}, (w_i)_{i\in\mathcal{I}}, (X_i)_{i\in\mathcal{I}}, (u_i)_{i\in\mathcal{I}}, (j)_{j\in\mathcal{J}}, (Y_j)_{j\in\mathcal{J}}, (\alpha_{ij})_{i\in\mathcal{I},j\in\mathcal{J}}\}$, where $\mathcal{I}$ is the set of consumers, w_i is the initial endowment for individual i, X_i is consumption set for i, u_i is the utility function of i, $\mathcal{J}$ is the set of production units, Y_j is the set of possible production plans for production unit $j \in \mathcal{J}$ while α_{ij} is the share of profit owned by consumer i for production unit j. Moreover, let l be the number of commodities in the economy. Given a competitive economy, a competitive equilibrium is defined below.

Definition 2.20. [7] A set of vectors $(x_1^*, ..., x_m^*, y_1^*, ..., y_n^*, p^*)$ is a competitive equilibrium if it satisfies: [1]

(1). for each $j \in \mathcal{J}$, y_j^* maximizes $p^* \cdot y_j$ over the set Y_j

(2). x_i^* maximizes $u_i(x_i)$ over the set $\{x_i | x_i \in X_i, p^* \cdot x_i \leq p^* \cdot w_i + \sum_{j=1}^{|\mathcal{J}|} \alpha_{ij} p^* \cdot y_j^*\}$

[1] A competitive equilibrium is also referred as *Arrow-Debreu equilibrium* or *market equilibrium* as well.

(3). $p^* \in P = \{p | p \in \mathbb{R}^l, p \geq 0, \sum_{h=1}^{l} p_h = 1\}$

(4). Let $x = \sum_{i \in \mathcal{I}} x_i$, $y = \sum_{j \in \mathcal{J}} y_j$, $w = \sum_{i \in \mathcal{I}} w_i$ and $z = x - y - w$. Then $z^* \leq 0$ and $p^* \cdot z^* = 0$.

Under a competitive equilibrium, each production unit maximizes its profit and each consumer maximizes his own utility under the budget constraint while demands are satisfied by supplies. In particular, the condition $p^* \cdot z^* = 0$ in (4) is called *Walras' Law*. It implies that for any commodity, if its price is strictly positive, its demand is equal to its supply; if its price is 0, its demand is no more than its supply. This captures the essence of the equilibrium concept. Arrow and Debreu [7] showed that a competitive economy always has an equilibrium under some mild assumptions.

Theorem 2.21. *[7] Suppose the competitive economy satisfies the following conditions:*

(1). (a) Y_j is a closed convex subset of $\mathbb{R}^l$ containing 0 for each $j \in \mathcal{J}$.

(b) Let $Y = \sum_{j=1}^{|\mathcal{J}|} Y_j = \{y | y = \sum_{j=1}^{|\mathcal{J}|} y_j$, where $y_j \in Y_j\}$, and $\Omega = \{x | x \in \mathbb{R}^l, x \geq 0\}$. Then $Y \bigcap \Omega = 0$.

(c) $Y \bigcap (-Y) = 0$.

(2). The set of consumption vectors X_i available to individual $i \in \mathcal{I}$ is a closed convex subset of $\mathbb{R}^l$ which is bounded from below.

(3). (a) $u_i(x_i)$ is a continuous function on X_i.

(b) For each $x_i \in X_i$, there is an $x_i^{'} \in X_i$ such that $u_i(x_i^{'}) > u_i(x_i)$.

(c) If $u_i(x_i) > u_i(x_i^{'})$ and $0 < t < 1$, then $u_i[tx_i + (1-t)x_i^{'}] > u_i(x_i^{'})$.

(4). (a) $w_i \in \mathbb{R}^l$; for some $x_i \in X_i$, $x_{ih} < w_{ih}$ for each h.

(b) For all i, j, $\alpha_{ij} \geq 0$; for all j, $\sum_{i=1}^{|\mathcal{I}|} \alpha_{ij} = 1$.

then there is a competitive equilibrium.

2.4 Discrete Markov Chain

In this section, we review discrete Markov chain. Given an N-state discrete Markov chain, let $S = \{1, ..., N\}$ be the set of states and X_m be the state of the random process at time m. The basic property of Markov chain is that, $P(X_m|X_{m-1}, ..., X_0) = P(X_m|X_{m-1})$. This is called the *memoryless property* or *Markov property*, that is, the probability that the random process reaches a specific state at time m depends only on its state at time $m-1$. Let p_{ij} be the probability that the random process goes to state j given that its previous state is i, where for every i, $\sum_{j=1}^{N} p_{ij} = 1$. Then, a transition probability matrix $P = (p_{ij})_{N \times N}$ can completely define a discrete Markov chain. We define a simplex $\Delta = \{\pi | \pi \geq 0 \text{ and } \sum_{i=1}^{N} \pi_i = 1\}$. Furthermore, we define a function $f : \Delta \to \Delta$ where $f = P^T \pi$. It is easy to check that Δ is convex and compact; meanwhile f is a continuous function. Therefore, Brouwer's fixed point theorem guarantees that there is at least one $\pi \in \Delta$ such that $\pi = P^T \pi$, which is called *stationary distribution*. However, there might be multiple stationary distributions. Actually, we can go further than the existence result. Given the transition probability matrix P, we can define a graph $G = (V, E)$, where $V = \{1, 2, ..., N\}$ and $(i, j) \in E$ iff $p_{ij} > 0$. If G is strongly connected and aperiodic, then the Markov chain defined by P is called *ergodic*. It is well known that an ergodic Markov chain has a unique stationary distribution [52].

2.5 PPAD

In this section, we briefly review the PPAD complexity class by following [23, 14, 15].

Let $R \subset \{0,1\}^* \times \{0,1\}^*$ be a binary relation over $\{0,1\}^*$. If for every string $x \in \{0,1\}^*$, there exists a $y \in \{0,1\}^*$ such that $(x,y) \in R$, R is a total relation. TFNP is the complexity class that contains all the NP search problems defined by total functions. Different from the way to define most of the complexity classes, in the next, we first define a complete problem of PPAD, which is called LEAFD, and then we give a formal definition of PPAD.

Definition 2.22. [68, 15] The input of LEAFD is a pair $(C, 0^n)$, where C is a circuit with polynomial size. The input of C is $\{0,1\}^n \bigcup \{\text{“}end\text{”}\}$. For every $u_0 \in \{0,1\}^n \bigcup \{\text{“}end\text{”}\}$, the output $C(u_0)$ is an ordered pair (u_{-1}, u_1), where $u_{-1}, u_1 \in \{0,1\}^n \bigcup \{\text{“}end\text{”}\}$. The first element of $C(0^n)$ is “*end*.” Actually, a pair $(C, 0^n)$ implicitly defines a graph $G = (V, E)$ where $V \subseteq \{0,1\}^n$, and $(u,v) \in E$ iff u is the first element of $C(v)$ while v is the second element of u. For every input u, if the first element of $C(u)$ is “*end*,” u is a source; if the second element of $C(u)$ is “*end*,” u is a sink. A node is called a *directed leaf* if it is either a source or a sink. The output of the LEAFD problem is a directed leaf other than 0^n.

Now we are ready to define PPAD.

Definition 2.23. If a search problem is polynomial time reducible [76] to the LEAFD problem, it is in PPAD.

The G defined by the input instance of LEAFD is a directed graph in which the outdegree and the indegree of every vertex are at most 1. The string 0^n corresponds to a source vertex, which has indegree 0 and outdegree 1. Since the sum of degrees of all vertices in any graph is even, there must be at least one other leaf except 0^n in G. Therefore, a PPAD instance is a total relation, and the complexity class PPAD is a subclass of TFNP.

The definition of PPAD is motivated from the proof of the famous Sperner's lemma [68]. In a properly colored subdivided simplex, there always exists a panchromatic subdivision by Sperner's lemma. Actually, it is PPAD-complete [68] to find it. Hence, a more intuitive way to think about PPAD could be: it captures the computation power of finding a panchromatic subdivision in Sperner's lemma. The key difficulty in proving that a problem belongs to PPAD is showing that the orientation of u_{-1} and u_1 can be determined in polynomial time. It was shown that computing a Nash equilibrium is in PPAD by extending Shapley's index method [73, 79]. Most significantly, in 2005, it was proved that it is indeed PPAD-hard [23, 14] to compute a Nash equilibrium. This settled down one of the most important open problems in complexity theory. Furthermore, a computational version of the Brouwer's fixed point is also PPAD-complete [68].

CHAPTER III

The Computational Complexity of Competitive Equilibrium

It is not from the benevolence of the butcher, the brewer, or the baker, that we expect our dinner, but from their regard to their own self-interest. We address ourselves, not to their humanity but to their self-love, and never talk to them of our own necessities but of their advantages.

Adam Smith, *The Wealth of Nations*

3.1 Approximate Competitive Equilibrium

3.1.1 Motivations

Theoretical computer scientists have studied the complexity of computational economic problems for decades. In 2005, there were some breakthroughs in this area. Daskalakis, Goldberg, and Papadimitriou [23] showed that computing an approximate Nash equilibrium for a four-player game is PPAD-Complete. A few months later, the PPAD-hardness result was extended to two-player games by Chen and Deng [14]. It turns out that the class of PPAD is powerful enough to characterize the computation of Nash equilibria. However, we do not know whether PPAD class is the right class to characterize the computation of competitive equilibria. Building upon a link from Nash equilibrium to a special class of pairing Leontief economy established by Ye [83], Codenotti *et al.* [21] showed that the computation of competitive equilibria for a special class of pairing Leontief economy is PPAD-hard.

However, this special class of pairing Leontief economy violates one specific condition of Arrow-Debreu's existence theorem [7]. The theorem requires that the initial endowment of each commodity for each individual is strictly positive. Note that, although that is the case, the existence of competitive equilibria for that special class of pairing Leontief economy is still guaranteed by the existence of Nash equilibria. This is not surprising because Arrow-Debreu's theorem gives only a set of sufficient (not necessary) conditions for the existence of competitive equilibria. Furthermore, a Leontief economy (even a pairing one) does not always have a competitive equilibrium. Actually, it is NP-Complete to decide whether a Leontief economy has an equilibrium [21]. This implies that Leontief economies do not satisfy all the conditions of Arrow-Debreu's existence theorem [7], either. Thus, for those economies that do fully satisfy the conditions of Arrow-Debreu's existence theorem, the computational complexity of its equilibrium is still open. In this section, we will study the complexity of computing an approximate competitive equilibrium for the economy that exactly satisfies the conditions of Arrow-Debreu's existence theorem.

3.1.2 The Main Theorem

The pairing Leontief economy introduced in [21] does not always have an equilibrium, and it is NP-Complete to decide whether a pairing Leontief economy has an equilibrium. This implies that the pairing Leontief economy violates some conditions of Arrow-Debreu's existence theorem. In particular, the condition (4)a in theorem 2.21 is violated by the pairing Leontief economy. Recall that the initial endowment for each individual i is $w_i = (0, ..., 1, ..., 0)$ while the consumption set for each individual i is $X_i = \{x | x \in \mathbb{R}^l, x \geq 0\}$. We cannot find a consumption bundle $x_i \in X_i$ such that for each h, $x_{ih} < w_{ih}$.

In this section, we will show that the computation of an ϵ-approximate competitive equilibrium is PPAD-hard. This result is achieved by reducing the problem of computing a $\frac{1}{n^{\Theta(1)}}$-approximate Nash equilibrium, which is known to be PPAD-Complete [15], to an ϵ-approximate competitive equilibrium. We define approximate competitive equilibrium as follows:

Definition 3.1. A price vector $p = (p_1, p_2, ..., p_l)$ is an ϵ-approximate competitive equilibrium if given conditions (1), (2), (3) of definition 2.20 are satisfied and for any $1 \leq h \leq l$, when $p_h > 0$, $-\epsilon \sum\limits_{i \in \mathcal{I}} w_{ih} \leq z_h(p) \leq \epsilon \sum\limits_{i \in \mathcal{I}} w_{ih}$; when $p_h = 0$, $z_h(p) \leq \epsilon \sum\limits_{i \in \mathcal{I}} w_{ih}$.

We will make use of the idea in [21]. Let (A, B) be the payoff matrices of the row and column players in the bimatrix game. W.L.O.G., we assume that $A \in [1,2]_{n \times n}$ and $B \in [1,2]_{n \times n}$.

Definition 3.2. We define a competitive economy L as follows: $\mathcal{I} = \{1, ..., 2n\}$; for each i, $w_i = \{\gamma, ..., \gamma, 1+\gamma, \gamma, ..., \gamma\}$, where each element of w_i is $\gamma > 0$ except for the ith element that is $1 + \gamma$; $X_i = \{x | x \in \mathbb{R}^l, x \geq 0\}$; $u_i(x_i) = \min\limits_{k: f_{ki} \neq 0} \{\frac{x_{ik}}{f_{ki}}\}$, where f_{ki} is an element of F, and there is only one production unit such that $Y_1 = Y = 0$ and $\alpha_{i1} = \frac{1}{2n}$. The matrix F is defined to be

$$F = \begin{pmatrix} \overrightarrow{1}_{n \times n} & \overrightarrow{1}_{n \times n} + A \\ \overrightarrow{1}_{n \times n} + B^T & \overrightarrow{1}_{n \times n} \end{pmatrix} \tag{3.1}$$

Lemma 3.3. *The economy L fully respects all the conditions of theorem 2.21. Thus it is guaranteed to have a competitive equilibrium.*

Proof. Since $Y = \{0\}$, it is closed and convex. It is obvious that condition (1) is trivially satisfied. The consumption set $X_i = \{x | x \in \mathbb{R}^l, x \geq 0\}$ is closed and convex, $\overrightarrow{0}$ is its lower bound. Thus, condition (2) is satisfied. $u_i(x_i)$ is continuous since for any i, j, $1 \leq f_{ij} \leq 3$, thus for any $\epsilon > 0$, if $\|x_i^{'} - x_i\|_\infty \leq \epsilon$, $|u_i(x_i^{'}) - u_i(x_i)| \leq \epsilon$. Therefor, condition (3)a is satisfied. Moreover, for any $x_i \in X_i$, $x_i + \overrightarrow{1} \in X_i$, thus $u_i(x_i + \overrightarrow{1}) > u_i(x_i)$. Hence, condition (3)b is satisfied. Furthermore, suppose $u_i(x_i) > u_i(x_i^{'})$ and $0 < t < 1$, $u_i[tx_i + (1-t)x_i^{'}] = \min_{i: f_{ki} \neq 0} \{\frac{tx_{ik}+(1-t)x_{ik}^{'}}{f_{ki}}\}$. W.L.O.G., we assume that $u_i[tx_i + (1-t)x_i^{'}] = \frac{tx_{ih}+(1-t)x_{ih}^{'}}{f_{gi}} > tu_i(x_i^{'}) + (1-t)u_i(x_i^{'}) = u_i(x_i^{'})$. Thus, condition (3)c is satisfied. Finally, $\gamma > 0$, condition (4)a is satisfied. According to the definition of L economy, condition (4)b is automatically satisfied. The lemma is proved. □

Theorem 3.4. *Given a competitive economy that fully respects all the conditions of Arrow-Debreu's existence theorem, for any positive constant $h > 0$, it is PPAD-hard to compute a $\frac{1}{n^h}$-approximate competitive equilibrium, i.e., it is PPAD-hard to compute a $\frac{1}{n^{\Theta(1)}}$-approximate competitive equilibrium.*

Proof. Suppose price vector $p = (p_1, ..., p_{2n})$ is an ϵ-approximate competitive equilibrium. Given this price vector, each individual should maximize its utility. For the utility functions of L, we know that for individual i, the optimal consumption bundle x_i should satisfy that for any $1 \leq k \leq 2n$, $\frac{x_{ik}}{f_{ki}} = \beta_i$, where β_i is the maximum utility for i under p. Moreover, we know that when the utility of individual i is maximized, $\sum_{k=1}^{2n} p_k x_{ik} = \sum_{k=1}^{2n} p_k f_{ki} \beta_i = \gamma + p_i$. Thus $\beta_i = \frac{\gamma + p_i}{\sum_{k=1}^{2n} p_k f_{ki}}$. In particular, if $p_i = 0$, $\beta_i = \frac{\gamma}{\sum_{k=1}^{2n} p_k f_{ki}} \leq \gamma$. According to the definition of approximate competitive equilibrium, we know that for any k, if $p_k > 0$, $(1-\epsilon)(1+2n\gamma) \leq \sum_{i=1}^{2n} x_{ik} = \sum_{i=1}^{2n} f_{ki}\beta_i \leq (1+\epsilon)(1+2n\gamma)$; if $p_k = 0$, $\sum_{i=1}^{2n} x_{ik} = \sum_{i=1}^{2n} f_{ki}\beta_i \leq (1+\epsilon)(1+2n\gamma)$. The parameters γ and ϵ will be determined later.

In the section that follows, we claim that when γ and ϵ are small enough, both $\sum_{i=1}^{n} \beta_i$ and $\sum_{i=n}^{2n} \beta_i$ should be sufficient large positive numbers. First we would like to show that there exists at least one commodity $1 \leq k_1 \leq n$ and another commodity $n+1 \leq k_2 \leq 2n$ with positive prices. Otherwise, suppose, for every $n+1 \leq k \leq 2n$,

$p_k = 0$, then for every $n+1 \leq i \leq 2n$, $\beta_i = \frac{\gamma + p_i}{\sum_{k=1}^{2n} p_k f_{ki}} \leq \gamma$. By the definition of approximate competitive equilibrium, for a commodity k_1 with positive price, we can get the following inequality,

$$(1-\epsilon)(1+2n\gamma) \leq \sum_{i=1}^{2n} f_{k_1 i}\beta_i \leq \sum_{i=1}^{n} \beta_i + 3n\gamma \tag{3.2}$$

For every commodity $n+1 \leq k_2 \leq 2n$,

$$\sum_{i=1}^{2n} f_{k_2 i}\beta_i \geq 2\sum_{i=1}^{n} \beta_i + \sum_{i=n+1}^{2n} \beta_i \tag{3.3}$$

Moreover, for each i, $\beta_i \geq \frac{p_i}{3}$. Thus $\sum_{i=1}^{n} \beta_i \geq \frac{1}{3}$ in the above situation. When $\frac{1-3\epsilon}{1+\epsilon} > 9n\gamma$, we can get

$$\begin{aligned} \sum_{i=1}^{2n} f_{k_2 i}\beta_i &\geq 2\sum_{i=1}^{n} \beta_i + \sum_{i=n+1}^{2n} \beta_i && (3.4) \\ &\geq \frac{1+\epsilon}{1-\epsilon}\left(\sum_{i=1}^{n} \beta_i + 3n\gamma\right) && (3.5) \\ &\geq (1+\epsilon)(1+2n\gamma) && (3.6) \end{aligned}$$

This contradicts the definition of approximate equilibrium. Similarly, we can get a contradiction when for every $1 \leq k \leq n$, $p_k = 0$. Therefore, when γ and ϵ are small enough, there must be a commodity $1 \leq k_1 \leq n$ and another commodity $n+1 \leq k_2 \leq 2n$ with positive prices. By the definition of approximate competitive equilibrium, we can get

$$\sum_{i=1}^{n} \beta_i + 2\sum_{i=n}^{2n} \beta_i \leq \sum_{i=1}^{2n} f_{k_1 i}\beta_i \leq (1+\epsilon)(1+2n\gamma) \tag{3.7}$$

and

$$3\sum_{i=1}^{n} \beta_i + \sum_{i=n}^{2n} \beta_i \geq \sum_{i=1}^{2n} f_{k_2 i}\beta_i \geq (1-\epsilon)(1+2n\gamma) \tag{3.8}$$

Thus we can get $\sum_{i=1}^{n} \beta_i \geq \frac{(1-3\epsilon)(1+2n\gamma)}{5} > \frac{1}{6}$ when $\epsilon < \frac{1}{18}$. Symmetrically, we can get $\sum_{i=n+1}^{2n} \beta_i \geq \frac{(1-3\epsilon)(1+2n\gamma)}{5} > \frac{1}{6}$ when $\epsilon < \frac{1}{18}$. Hence, the claim is proved.

For $1 \leq k \leq n$, let $x_k = \frac{\beta_k}{\sum\limits_{i=1}^{n} \beta_i}$ while for $n+1 \leq k \leq 2n$, let $y_{k-n} = \frac{\beta_k}{\sum\limits_{i=n+1}^{2n} \beta_i}$. We would like to show that when when γ and ϵ are small enough, (x, y) is a good approximate Nash equilibrium of bimatrix game (A, B). According to the definition of approximate competitive equilibrium, if $p_k > 0$, $A_{k*}y \geq \frac{(1-\epsilon)(1+2n\gamma)-\sum\limits_{i=1}^{2n}\beta_i}{\sum\limits_{i=n+1}^{2n}\beta_i}$, where A_{k*} stands for the kth row of matrix A; if $p_k = 0$, $x_k = \frac{\beta_k}{\sum\limits_{i=1}^{n}\beta_i} \leq 6\gamma$. Let

$$P = \frac{(1+\epsilon)(1+2n\gamma) - \sum\limits_{i=1}^{2n} \beta_i}{\sum\limits_{i=n+1}^{2n} \beta_i} \tag{3.9}$$

and

$$Q = \frac{(1-6n\gamma)[(1-\epsilon)(1+2n\gamma) - \sum\limits_{i=1}^{2n} \beta_i]}{\sum\limits_{i=n+1}^{2n} \beta_i} \tag{3.10}$$

Thus, we know that

$$\begin{aligned} \max(Ay) - x^T Ay &\leq P - Q \\ &\leq \frac{(2\epsilon + 6n\gamma(1-\epsilon))(1+2n\gamma)}{\sum\limits_{i=n+1}^{2n} \beta_i} \\ &\leq (12\epsilon + 36n\gamma(1-\epsilon))(1+2n\gamma) \end{aligned}$$

So, for any positive constants $t, h > 0$, when $\gamma = \frac{1}{n^{1+t}}$ and $\epsilon = \frac{1}{n^h}$, $\max(Ay) - x^T Ay = O(\frac{1}{n^h})$. Similarly, we can get $\max(B^T y) - x^T By = O(\frac{1}{n^h})$. It has been shown that, for any $h > 0$, the problem of computing a $\frac{1}{n^h}$-approximate Nash equilibrium is PPAD-Complete [15]. Therefore, computing of an ϵ-approximate competitive equilibrium when $\epsilon = \frac{1}{n^h}$ is PPAD-hard. □

Remark: Actually lemma 4.3 of [45] implies that it is PPAD-hard to compute $O(\frac{1}{n^2})$-approximate Arrow-Debreu equilibrium when $\gamma = O(\frac{1}{n^3})$. Our result extends the hardness result in [45] by allowing larger approximation error and perturbations.

In this section we have shown that, given a competitive economy that fully respects all the conditions of Arrow-Debreu's existence theorem, for any positive constant $h > 0$, it is PPAD-hard to compute a $\frac{1}{n^h}$-approximate competitive equilibrium. To the best of our knowledge, this is the first complexity result about the economy that strictly satisfies the conditions of Arrow-Debreu's existence theorem. Moreover, since our result allows larger approximation error(smaller h), it improves the main result in [45].

3.2 Equilibria in Markets with Additively Separable Utility Functions

3.2.1 Motivations

In this section, we will study the computational complexity of Arrow-Debreu equilibria in markets with additively separable utility functions. In particular, we focus on the exchange markets, which do not have production units as do competitive markets. Despite the progress both on algorithms for and on the complexity-theoretic understanding of market equilibria, several fundamental questions concerning market equilibria, including some seemingly simple ones, remain unsettled. In particular, Jain [48] designed a polynomial time algorithm to compute Arrow-Debreu equilibria for an exchange market with linear utility function. Additively separable piecewise linear concave (PLC) utility functions are probably the simplest function besides the linear utility function. Thus, Vijay Vazirani [65] wrote:

> "Concave utility functions, even if they are additively separable over the goods, are not easy to deal with algorithmically. In fact, obtaining a polynomial time algorithm for such functions is a premier open question today."

A function $u(x_1, ..., x_n)$ is an additively separable and concave function if there exist n real-valued concave functions $f_1, ..., f_n$ such that $u(x_1, ..., x_n) = \sum_{j=1}^{n} f_j(x_j)$. Noting that every concave function can be approximated by a PLC function, Vazirani [65] further asked whether one can find an equilibrium in a market with additively separable PLC utility functions in polynomial time or if the problem is PPAD-hard.

3.2.2 The Main Theorem

Given an additively separable PLC utility function $u(x_1, ..., x_n) = \sum_{j=1}^{n} f_j(x_j)$, we call it *k-linear* if for every j, the PLC function $f_j(.)$ has at most k linear segments. Together with Xi Chen, Decheng Dai, and Shanghua Teng [13], we show that:

Theorem 3.5 (Main). *It is PPAD-Complete to compute an n^{-13} approximate Arrow-Debreu equilibrium in an exchange market with 2-linear additively separable PLC utility functions, where each trader only owns and wants $O(1)$ goods.*

In order to prove the above theorem, we introduce the definition of well-supported Nash equilibria.

Definition 3.6 (Well-Supported Nash Equilibria). For $\epsilon > 0$, $(\mathbf{x}, \mathbf{y})$ is an ϵ-well-supported Nash equilibrium of a bimatrix game $(\mathbf{A}, \mathbf{B})$, if $\mathbf{x}, \mathbf{y} \in \Delta^n$ and for all $i, j \in [1...n]$,

$$\mathbf{A}_i \mathbf{y}^T + \epsilon < \mathbf{A}_j \mathbf{y}^T \implies x_i = 0, \quad \text{and} \quad \mathbf{x}\mathbf{B}_i + \epsilon < \mathbf{x}\mathbf{B}_j \implies y_i = 0. \tag{3.11}$$

The basic idea to prove the above theorem is to reduce the well-supported Nash equilibria of a bimatrix game to the equilibria of a market with additively separable PLC utility functions. It is well known that it is PPAD-Complete to compute an approximate well-supported Nash equilibrium [15]. Note that, the existence theorem in [7] is rather limited, since it requires that the initial endowment of every good of every trader is strictly positive. In order to make our result more general and the reduction cleaner, we will prove a general theorem for the existence of Arrow-Debreu equilibria for additively separable PLC markets.

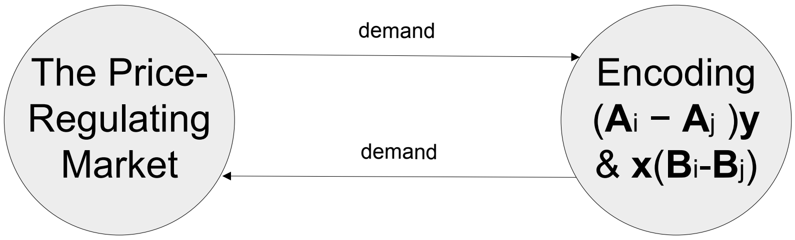

Figure 3.1: The reduction idea

To this end, we introduce the definition of economy graphs.

Definition 3.7 (Economy Graph). Given an exchange market, we define a directed graph $G = (\mathcal{T}, E)$ as follows. The vertex set of G is exactly $\mathcal{T}$, the set of traders in the market. For every two traders $T_i \neq T_j \in \mathcal{T}$, we have an edge from T_i to T_j if T_j possesses a good for which T_i has a demand. G is called the *economy graph* of the market. We say the market is strongly connected if G is strongly connected.[1]

Theorem 3.8. *Let $\mathcal{M}$ be a market with additively separable* PLC *utilities. If it is strongly connected, then a market equilibrium* $\mathbf{p}$ *exists. Moreover, every quasi-equilibrium of $\mathcal{M}$ is indeed an Arrow-Debreu equilibrium.*[2]

Theorem 3.8 is actually a corollary of Theorem 2 in [59]. A quasi-equilibrium is an Arrow-Debreu equilibrium except that it does not require any trader with zero income to maximize his utility.

Now we can describe the idea of our reduction from the well-supported equilibria of a bimatrix game to the equilibria of markets with additively separable PLC utility functions. The reduction consists of two major components as shown in Figure 3.1. One component is the price-regulating market, which enforces the equilibrium price of each good to remain between 1 and 2. The other component is deliberately constructed to encode the well-supported Nash equilibria. Interested audience can

[1] The direction of the edge in the economy graph here is different from the definitions of economy graphs in [19, 59].

[2] In [19], it has a similar theorem about the existence of equilibria for markets with CES utility functions. However, it does not cover linear or piecewise linear utility functions.

refer to [13] for the details.

First, we introduce a simple linear market $\{\mathcal{M}_{2n+2}\}$ with $2n+2$ goods, which we refer to as the price-regulating market. Among the $2n+2$ goods, there are two blocks of n goods corresponding to the row player and the column player respectively, while the other two goods are for cleanup purposes. There are $O(n^2)$ traders in $\{\mathcal{M}_{2n+2}\}$. Each trader is only interested in two goods. The marginal utility per unit of one good is twice as much as the other. $\mathcal{M}_{2n+2}$ has the following nice price-regulating property: If $\mathbf{p}$ is a *normalized*[3] approximate equilibrium price vector of $\mathcal{M}_{2n+2}$, then $p_k \in [1, 2]$ for all $k \in [1...2n+2]$. This price-regulating property allows us to encode $2n$ free variables $x_1, ..., x_n$ and $y_1, ..., y_n$ between 0 and 1.

As a key feature in our analysis, the price-regulation property is stable with respect to "small perturbations" to $\mathcal{M}_{2n+2}$. That is, when new traders are added to $\mathcal{M}_{2n+2}$ (without introducing new goods), this property remains hold as long as the total amount of goods these new traders initially have is small compared to those of the traders in $\mathcal{M}_{2n+2}$. Using the stability of the price-regulating markets $\{\mathcal{M}_{2n+2}\}$, given an $n \times n$ bimatrix game $(\mathbf{A}, \mathbf{B})$, we can construct an additively separable PLC market by adding new traders — whose initial endowments are relatively small — to $\mathcal{M}_{2n+2}$, the price-regulating market with $2n+2$ goods.

Remember that we have two blocks of goods in the market, which correspond to the row player and the column player respectively. We use the prices of the $2n$ goods to encode a pair of probability vectors $(\mathbf{x}, \mathbf{y})$ (after normalization): $x_k = p_k - 1$ and $y_k = p_{n+k} - 1$, $k \in [1...n]$. There are $2n(n-1)$ traders to add (besides some cleanup traders). Each trader corresponds to a pair of pure strategies of the row (or the column) player. We will deliberately set up the initial endowments and utility

[3] In this section, we say a price vector $\mathbf{p}$ is normalized if the smallest nonzero entry of $\mathbf{p}$ is equal to 1.

functions of each trader such that the trader can encode the inequality $(\mathbf{A}_i - \mathbf{A}_j)^T\mathbf{y}$ (or $\mathbf{x}(\mathbf{B}_i - \mathbf{B}_j)$). For a trader (i, j) corresponding to the row player, he has small initial endowment of the second block of goods, the amount of each good is proportional to $\max\{A_{ik} - A_{jk}, 0\}$, and a relatively large endowment of good i, which belongs to the first block of goods. However, trader (i, j) has no initial endowment of other goods (except the two cleanup goods).

The additively separable utility functions of trader (i, j) with respect to the second block of goods as well as good i are PLC with 2 segments. The amount of each good, which corresponds to the turning point of each 2-segment PLC utility function of the second block of goods, is proportional to $\max\{A_{jk} - A_{ik}, 0\}$, and that value corresponding to good i is trader (i, j)'s initial endowment of good i. Moreover, the marginal utility per unit of each good, which corresponds to the first segment of each PLC utility function of the second block of goods, is three times that of good i.

Now suppose $(\mathbf{A}_i - \mathbf{A}_j)^T\mathbf{y} < \epsilon$. In trading, trader (i, j) will sell his initial endowments to generate revenue. Note that the equilibrium price of each good is between 1 and 2. Thus, whatever the equilibrium price is, the marginal utility per unit cost of each good, which corresponds to the first segment of each PLC utility function of the second block of goods, is always greater than that of good i. Therefore, trader (i, j) will first buy the second block of goods proportionally up to $\max\{A_{jk} - A_{ik}, 0\}$ for each good, and can only buy good i with the revenue left. We can further show that: the difference between trader (i, j)'s initial endowment and consumption of good i is proportional to $(\mathbf{A}_i - \mathbf{A}_j)^T\mathbf{y}$ divided by p_i. Thus, given $\mathbf{y}$, if the payoff of strategy i of the row player is less than that of strategy j by more than ϵ, then trader (i, j) does not have enough money to buy much of good i (besides his initial endowment). Instead, the amount of good i left is bought by the traders in the linear market $\mathcal{M}_{2n+2}$, who have large incomes. Intuitively, this can happen only if the price of good i is very low. Actually, we show that $(\mathbf{A}_i - \mathbf{A}_j)^T\mathbf{y} < \epsilon$ indicates that $p_i = 1$. Thus, by definition, $x_i = p_i - 1 = 0$. This encodes the condition of the approximate well-supported Nash equilibria. Therefore, we get a market $\mathcal{M}$ with the property that from every approximate market equilibrium $\mathbf{p}$ of $\mathcal{M}$, the pair $(\mathbf{x}, \mathbf{y})$ obtained above (after normalization) is an approximate Nash equilibrium of $(\mathbf{A}, \mathbf{B})$. Moreover, if $(\mathbf{A}, \mathbf{B})$ is a sparse game, each trader in $\mathcal{M}$ initially has $O(1)$ goods and is only interested in $O(1)$ goods. Thus, the main theorem is proved.

Combined with Theorem 3.8, an important corollary of the main theorem is that:

Theorem 3.9. *It is PPAD-hard to compute a quasi-equilibrium of a competitive economy.*

3.3 Related Works

In the past decade, people have designed a few algorithms to compute competitive equilibrium and its variants. Devanur *et al.* [26] gave the first algorithm for linear utility functions in Fisher market based on convex programming. Later, Jain [48] gave the first algorithm to compute Arrow-Debreu equilibrium for linear utility functions also based on convex programming. Codenotti *et al.* [20] developed an algorithm to compute approximated competitive equilibria for economies that satisfy weak gross substitutability. They [19] further gave an algorithm for economies of CES utility functions with elasticity no less than 1/2. Ye [84] designed an interior-point algorithm for solving Fisher and Arrow-Debreu market equilibrium with linear utility functions. Ye [83] also designed an algorithm to compute equilibrium for Leontief utility functions in Fisher market. All those algorithms are polynomial time. However, they only work for rather restricted classes of utility functions.

On the complexity side, Papadimitriou [68] showed that the computation of exchange equilibrium given aggregated excess demand functions is PPAD-Complete. However, the complexity question is unclear for the computation of competitive equilibrium, which concerns a set of utility-maximizing individuals, but the excess demand functions are not given explicitly. Deng *et al.* [25] studied the complexity, approximability, and inapproximability of competitive equilibrium for linear utility functions and indivisible commodities. Codenotti *et al.* [21] showed that the computation of competitive equilibrium for a special class of pairing Leontief economy is PPAD-hard. Recently, based on [21], Huang and Teng [45] showed that the computation of approximate equilibrium for the Leontief economy is PPAD-hard while its smooth complexity cannot be in polynomial time unless $PPAD \subset RP$. In particular, they showed that if the utility matrix and initial endowments matrix of the Leontief economy are perturbed by a magnitude of σ, given an ϵ-approximate equilibrium of the Leontief economy, an $O(n\sqrt{\epsilon} + n^{1.5}\sqrt{\sigma})$-approximate Nash equilibrium can be constructed in polynomial time. Thus, according to Chen, Deng, and Teng [15], it is PPAD-hard to compute a $\frac{1}{n^h}$-approximate competitive equilibrium for any $h > 2$. Note that the definition of the approximate equilibrium in [45] for Leontief economy is weak in the sense that it does not guarantee that the demand for any commodity with positive price is not much less than its supply.

CHAPTER IV

On the Complexity of Deciding Degeneracy in Games

4.1 Motivations

Game theory is a subject of studying and predicting the behavior of rational decision makers. In 1950, Nash [62] showed that under mild conditions, a noncooperative game always has a Nash equilibrium. However, Nash's proof of the existence of Nash equilibrium is nonconstructive. Researchers have become devoted to studying the computation of Nash equilibria since then. On the complexity side, it has been shown that it is PPAD-Complete to compute a Nash equilibrium, even for bimatrix games [14, 23]. On the algorithmic side, a nice property of bimatrix games is that as long as the payoff matrices are rational, the equilibria must be rational, too. Lemke and Howson [55] designed a combinatorial algorithm, which can not only compute equilibria of a bimatrix game, but also shows the existence of equilibria constructively.

Nevertheless, just like the Simplex algorithm for linear programming, Lemke and Howson's algorithm can fail on degenerate games. In order to deal with degeneracy in the computation of Nash equilibrium, certain perturbation techniques, such as lexicographical perturbation, should be deployed [65]. Besides that, a nondegenerate bimatrix game has a nice property in that the number of its equilibria is odd [65].

However, for a degenerate bimatrix game, its equilibria are the union of the connected union of maximal Nash subsets [65], which contains an infinite number of Nash equilibria points. In all, there are significant differences between degenerate and nondegenerate games.

In this chapter, we will investigate the computational complexity of deciding degeneracy in bimatrix games. We will show that it is NP-Complete to decide whether a bimatrix game is degenerate while it is Co-NP-Complete to decide whether it is nondegenerate. However, we show that, for a win-lose bimatrix game, it is in P to decide whether it is degenerate.

4.2 Degeneracy in Linear Programming and in Bimatrix Games

A closely related concept to Nash equilibrium is the best response condition.

Definition 4.1. [65] (Best Response condition) Let x and y be the strategies of the row player and column player, respectively. x is the best response to y if

$$x_i > 0 \Rightarrow (Ay)_i = u = \max\{(Ay)_k | k \in [1..m]\}$$

Now we are ready to define degeneracy in games.

Definition 4.2. [65] (Nondegenerate) A bimatrix game is *nondenegerate* if there is no mixed strategy of support size k has more than k pure best responses.

Otherwise, if the above condition is violated, the game is called degenerate.

In a linear system,

$$\begin{cases} Ax = b \\ x \geq 0 \end{cases}$$

where A is a matrix of dimension $m \times n$ and rank m. This system is said to be *degenerate*, if there exists a basis B such that at least one component in the vector $B^{-1}b$

is zero. It is well known that if the linear system is degenerate, b can be expressed as a linear combination of at most $m-1$ columns of A. It has been shown [61] that deciding whether a linear programming is degenerate is NP-Complete. However, their proof cannot work for the degenerate problem of bimatrix games. This is because the matrix they constructed [61] is a 0-1 matrix, which is simply degenerate if used as a payoff matrix.

In the next section, we study the relationship between degeneracy in linear programming and degeneracy in games. Let $A = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$ and $b = (2,1)^T$. It is easy to see that this linear system is degenerate. However, if A is the payoff matrix of the row (the column) player, the game is nondegenerate. Conversely, let $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ and $b = (2,1)^T$. If A is the payoff matrix in a bimatrix game, the game is degenerate. However, the linear system is nondegenerate. Therefore, degeneracy in linear programming does not imply degeneracy in games, and vice versa.

4.3 The Main Theorem

Theorem 4.3. *It is NP-Complete to decide whether a bimatrix game is degenerate.*

In order to prove this theorem, we first show our construction in below.

The construction: We will reduce the 3-SAT to our problem. Let f be a 3-SAT formula $f = c_1 \wedge c_2 \wedge ... \wedge c_n$, where n is the number of clauses and each clause $c_i = l_{i1} \vee l_{i2} \vee l_{i3}$ contains three literals. Each literal l_{iw} is either a positive variable x_h or its negation $\overline{x_h}$. Let c_i^k denote the assignment of the true values of the three literals in c_i as the binary representation of the integer k, where $k \in [1..7]$. For example, the binary representation of 5 is 101. Thus, c_i^5 represents the assignment such that $l_{i1} = 1, l_{i2} = 0$ and $l_{i3} = 1$.

The strategy space of the column player is $\mathcal{C} = \{c_i^k | i \in [1..n] \text{ and } k \in [1..7]\}$. And the strategy space $\mathcal{R}$ of the row player is $\mathcal{C} \bigcup \{f\} \bigcup \{c_i^p c_j^q | \; c_i^p \text{ and } c_j^q \text{ are conflicting}\}$, where f is a special strategy. Two clause assignments c_i^p and c_j^q are conflicting iff there are two literals l_{iw} in c_i and l_{js} in c_j corresponding to the same variable x_h but c_i^p and c_j^q make conflicting assignments to x_h. For example, let $c_1 = x_1 \vee \overline{x_2} \vee x_3$ and $c_2 = x_2 \vee x_4 \vee x_5$. c_1^5 will make $\overline{x_2} = 0$, which assigns $x_2 = 1$ and c_2^3 will assign $x_2 = 0$. Thus, c_1^5 and c_2^3 are conflicting clause assignments. Note that the order of c_i^p and c_j^q does not matter, i.e., the pure strategy $c_i^p c_j^q$ implicitly equals the pure strategy $c_j^q c_i^p$. We call $c_i^p c_j^q$ *conflicting-clause-assignments strategy.*

Given the strategy spaces of the row and column players, the payoff function r for the row player is:

(1). $r(c_i^p, c_h^k) = 1$ if $i = h$ and $p = k$; otherwise $r(c_i^p, c_h^k) = 0$.

(2). $\forall h \in [1..n]$ and $k \in [1..7]$, $r(f, c_h^k) = 1/n$.

(3). Suppose we have D conflicting-clause-assignments strategy $c_i^p c_j^q$ and let $\mathcal{O}$ be an arbitrary ordering of the elements in $\{c_i^p c_j^q | \; c_i^p \text{ and } c_j^q \text{ are conflicting}\}$ from 0 to $D-1$. Thus, for the dth conflicting-clause-assignments strategy $c_i^p c_j^q$, $r(c_i^p c_j^q, c_h^k) = \frac{1}{2} + 4^d \epsilon$ if $i = h$ and $p = k$ or $j = h$ and $q = k$; otherwise $r(c_i^p c_j^q, c_h^k) = 0$. Here ϵ is a small positive number whose value will be fixed later. Note that $D \in O(n^2)$.

The following matrix illustrates the payoff function of the row player.

$$
\begin{array}{c|ccccccccccccc}
 & c_1^1 & c_1^2 & \cdots & c_1^7 & c_2^1 & c_2^2 & \cdots & c_2^7 & \cdots\cdots & c_n^1 & c_n^2 & \cdots & c_n^7 \\
c_1^1 & 1 & & & & & & & & & & & & \\
c_1^2 & & 1 & & & & & & & & & & & \\
\vdots & & & \ddots & & & & & & & & & & \\
c_1^7 & & & & 1 & & & & & & & & & \\
c_2^1 & & & & & 1 & & & & & & & & \\
\vdots & & & & & & \ddots & & & & & & & \\
c_2^7 & & & & & & & & 1 & & & & & \\
\vdots & & & & & & & & & \ddots & & & & \\
c_n^1 & & & & & & & & & & 1 & & & \\
\vdots & & & & & & & & & & & \ddots & & \\
c_n^7 & & & & & & & & & & & & & 1 \\
f & \frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \frac{1}{n} \\
c_1^1c_1^2 & \frac{1}{2}+\epsilon & \frac{1}{2}+\epsilon & & & & & & & & & & & \\
\vdots & & & & & & & & & & & & & \\
c_i^pc_j^q & & & & & & \frac{1}{2}+4^d\epsilon & & & \frac{1}{2}+4^d\epsilon & & & & \\
\vdots & & & & & & & & & & & & & \\
c_n^6c_n^7 & & & & & & & & & & & & \frac{1}{2}+4^{D-1}\epsilon & \frac{1}{2}+4^{D-1}\epsilon
\end{array}
$$

Proof. Given the mixed strategy y of the column player with support k, we can multiply y with the payoff matrix of the row player. Then, if the number of pure strategy best responses of the row player is more than k, the game is degenerate. Thus, deciding degeneracy is in NP. We will reduce the 3-SAT problem to the problem of deciding degeneracy in bimatrix games as shown in above. It is obvious that the size of the payoff matrix is in a polynomial of n, and each entry of the matrix can be represented by a polynomial number of bits. Therefore, the construction can be done in polynomial time.

If there is a satisfying assignment for f, the assignment of each clause c_i^k must be in the form of c_i, where $k \in [1..7]$. We call such a corresponding strategy c_i^k *active pure strategy*. We will set the mixed strategy y of the column player to be $\frac{1}{n}$ on those active pure strategies and 0 anywhere else. Thus, the support of y is n. Given y, for each conflicting-clause-assignments strategy $c_i^pc_j^q$, at most one of c_i^p and c_j^q could be active. So, when ϵ is small enough, $c_i^pc_j^q$ cannot be a best response. Hence, given the mixed strategy y, for the row player, the set of pure best responses to y are all the active pure strategies and f. In other words, the number of best responses is $n+1$.

Therefore, if the formula f is satisfiable, the game is degenerate.

In the next section, we prove the reverse direction, *i.e.*, if the game is degenerate, the formula f is satisfiable. Suppose the game is degenerate, which can be shown with the mixed strategy y of the column player. (Note that y cannot be a pure strategy, since each column of the payoff matrix only has one element equal to 1.) Again, let u be the maximum payoff of the row player given y and $\mathcal{S}$ be the support set of y. Assume that the support set $\mathcal{S}$ can be represented as the union of two subsets $\mathcal{M} = \{c_i^k | y[c_i^k] = u\}$ and $\mathcal{N} = \{c_i^k | y[c_i^k] < u\}$. We will show that the set $\mathcal{N}$ is empty by contradiction in the next section. Thus, we assume $\mathcal{N}$ is not empty now.

We will count the number of pure best responses of the row player with respect to y. First, for strategy $c_i^k \in \mathcal{M}$, its corresponding pure strategy c_i^k of the row player is one of the best responses since the maximum payoff of the row player is u. Second, the payoff of the strategy f is always $\frac{1}{n}$ whatever y is. For any conflicting-clause-assignments strategy $c_i^p c_j^q$, if both $c_i^p \in \mathcal{M}$ and $c_j^q \in \mathcal{M}$, its payoff will exceed u, which contradicts the assumption that u is the maximum payoff. Thus, at least one of c_i^p and c_j^q must belong to $\mathcal{N}$. Now we will show that for any pure strategy $c_i^p c_j^q$, if it is a best response strategy respect to y, then at least one of c_i^p and c_j^q will be in $\mathcal{N}$ and that strategy cannot appear in any best-response conflicting-clause-assignments strategy $c_{i'}^{p'} c_{j'}^{q'}$, which has a smaller index than $c_i^p c_j^q$ in $\mathcal{O}$. We will prove this by cases.

(1). $D = 1$. It naturally follows.

(2). $D = 2$. Assume the two best response strategies are $c_{i_1}^{p_1} c_{j_1}^{q_1}$ and $c_{i_2}^{p_2} c_{j_2}^{q_2}$. If both $c_{i_2}^{p_2}$ and $c_{j_2}^{q_2}$ either belong to $\mathcal{M}$ or appear in $c_{i_1}^{p_1} c_{j_1}^{q_1}$, the payoff value of $c_{i_2}^{p_2} c_{j_2}^{q_2}$ is greater than that of $c_{i_1}^{p_1} c_{j_1}^{q_1}$. Contradiction occurs. Thus, either $c_{i_2}^{p_2}$ or $c_{j_2}^{q_2}$ belongs to $\mathcal{N}$ and does not appear in $c_{i_1}^{p_1} c_{j_1}^{q_1}$.

(3). $D \geq 3$. Assume that $c_{i_1}^{p_1}, c_{j_2}^{q_2} \in \mathcal{N}$ and the corresponding conflicting-clause-assignments strategies $c_{i_1}^{p_1} c_{j_1}^{q_1}$ and $c_{i_2}^{p_2} c_{j_2}^{q_2}$ of the row player are the best responses. Moreover, let $c_{i_1}^{p_1} c_{j_2}^{q_2}$ be another conflicting-clause-assignments strategy (if they are not conflicting, $c_{i_1}^{p_1} c_{j_2}^{q_2}$ will not belong to the strategy space of the row player; so, it cannot be a best response). Assume the indices of $c_{i_1}^{p_1} c_{j_1}^{q_1}$, $c_{i_2}^{p_2} c_{j_2}^{q_2}$, and $c_{i_1}^{p_1} c_{j_2}^{q_2}$ in $\mathcal{O}$ are d_1, d_2, and d respectively, where $d_1 < d_2 < d$. Then, we claim that the payoff to $c_{i_1}^{p_1} c_{j_2}^{q_2}$ will be more than u when ϵ is small enough. Note that, by the assumption of best-response condition,

$$(\frac{1}{2} + 4^{d_1}\epsilon)(y[c_{i_1}^{p_1}] + y[c_{j_1}^{q_1}]) = u,$$

we know that $y[c_{j_1}^{q_1}] \leq u$, thus

$$y[c_{i_1}^{p_1}] \geq \frac{1 - 2 \cdot 4^{d_1}\epsilon}{1 + 2 \cdot 4^{d_1}\epsilon} u.$$

Similarly, we can get $y[c_{j_2}^{q_2}] \geq \frac{1-2\cdot 4^{d_2}\epsilon}{1+2\cdot 4^{d_2}\epsilon} u$. Set

$$\epsilon = \frac{1}{4 \cdot 4^{2D}},$$

the payoff of $c_{i_1}^{p_1} c_{j_2}^{q_2}$ would be

$$\begin{aligned}
(\frac{1}{2} + 4^{d}\epsilon)(\frac{1 - 2 \cdot 4^{d_1}\epsilon}{1 + 2 \cdot 4^{d_1}\epsilon} + \frac{1 - 2 \cdot 4^{d_2}\epsilon}{1 + 2 \cdot 4^{d_2}\epsilon})u &> (1 + 2 \cdot 4^{d}\epsilon)\frac{1 - 2 \cdot 4^{d_2}\epsilon}{1 + 2 \cdot 4^{d_2}\epsilon} u \\
&\geq (1 + 8 \cdot 4^{d_2}\epsilon)\frac{1 - 2 \cdot 4^{d_2}\epsilon}{1 + 2 \cdot 4^{d_2}\epsilon} u \\
&= \frac{1 + 6 \cdot 4^{d_2 - 2D - 1} - 4^{2(d_2 - 2D)}}{1 + 2 \cdot 4^{d_2 - 2D - 1}} u \\
&> u
\end{aligned}$$

This contradicts the assumption that u is the maximum payoff to the row player.

Thus, for any conflicting-clause-assignments strategy $c_i^p c_j^q$, if it is a best response respect to y, then at least one of c_i^p and c_j^q will be in $\mathcal{N}$ and that strategy cannot appear in any best-response conflicting-clause-assignments strategy $c_{i'}^{p'}c_{j'}^{q'}$, which has a smaller index than $c_i^p c_j^q$ in $\mathcal{O}$. In other words, if there are l best-response conflicting-clause-assignments strategies respect to y, then the number of corresponding active pure strategies in $\mathcal{N}$ is at least l.

Now, we can count the number of pure best responses with respect to y. First of all, we know that $u \geq \frac{1}{n}$, since the payoff to the strategy f is always $\frac{1}{n}$. If $u > \frac{1}{n}$, the number of pure best responses is at most $|\mathcal{M}| + |\mathcal{N}| = |\mathcal{S}|$. Thus, the game is nondegenerate. Contradiction occurs. If $u = \frac{1}{n}$, let n' be the number of best-response conflicting-clause-assignments strategies. We will show that $n' < |\mathcal{N}|$. Otherwise, assume $n' = |\mathcal{N}|$. (n' cannot be greater than $|\mathcal{N}|$ by the above argument.) Since for each best-response conflicting-clause-assignments strategy $c_i^p c_j^q$, either c_i^p or c_j^q will be in $\mathcal{N}$. W.L.O.G., we assume $c_i^p \in \mathcal{N}$. We also know that $\frac{1-2\cdot4^D\epsilon}{1+2\cdot4^D\epsilon}\frac{1}{n} \leq y[c_i^p] < \frac{1}{n}$. Thus, there must be a strategy $c_h^k \in \mathcal{N}$ such that

$$
\begin{aligned}
0 < y[c_h^k] &\leq 1 - \frac{|\mathcal{M}|}{n} - \frac{1-2\cdot4^D\epsilon}{1+2\cdot4^D\epsilon}\frac{|\mathcal{N}|-1}{n} \\
&\leq \frac{|\mathcal{M}|+|\mathcal{N}|-1}{n} - \frac{|\mathcal{M}|}{n} - \frac{1-2\cdot4^D\epsilon}{1+2\cdot4^D\epsilon}\frac{|\mathcal{N}|-1}{n} \\
&\leq \frac{|\mathcal{N}|-1}{n}\frac{4\cdot4^D\epsilon}{1+2\cdot4^D\epsilon}
\end{aligned}
$$

We know that when $\epsilon = \frac{1}{4\cdot4^{2D}}$,

$$
\frac{|\mathcal{N}|-1}{n}\frac{4\cdot4^D\epsilon}{1+2\cdot4^D\epsilon} < \frac{1-2\cdot4^D\epsilon}{1+2\cdot4^D\epsilon}\frac{1}{n}
$$

This implies that the strategy c_h^k cannot appear in any best-response conflicting-clause-assignments strategy. Thus $n' < |\mathcal{N}|$. Moreover, f is a best response when $u = \frac{1}{n}$. In all, when $u = \frac{1}{n}$, the number of best responses for the row player (with respect to y) cannot exceed $|\mathcal{N}| + |\mathcal{M}| = |\mathcal{S}|$. Thus, the game in nondegenerate.

Since we assume that the game is degenerate, contradiction still occurs. Therefore, $\mathcal{N}$ is empty. In other words, $\mathcal{S} = \mathcal{M}$.

We further claim that $u = \frac{1}{n}$ when $\mathcal{S} = \mathcal{M}$. Otherwise, if $u > \frac{1}{n}$, f cannot be a best response, nor can any conflicting-clause-assignments strategy. The game is still nondegenerate. Contradiction occurs. Thus, $u = \frac{1}{n}$. Moreover, we know that for any pure strategy c_i^p and c_i^q, since they are conflicting, both $y[c_i^p]$ and $y[c_i^q]$ cannot equal to $\frac{1}{n}$ simultaneously. Otherwise, the payoff to the conflicting-clause-assignments strategy $c_i^p c_i^q$ is more than $\frac{1}{n}$. Similarly, two conflicting clause assignments c_i^p and c_j^q cannot be in the support set $\mathcal{S}$ of y simultaneously. Reversely, for every clauses c_i, there must be some $k \in [1..7]$ such that $c_i^k \in \mathcal{S}$ by the pigeon hole principle. Thus, each clause is satisfied. Therefore, the clause assignments $(c_1^{k_1}, c_2^{k_2}, ..., c_n^{k_n})$ corresponding to $\mathcal{S}$ give a satisfying assignment to the formula f.

Finally, we have shown that the formula f is satisfiable iff the game is degenerate. Thus, deciding degeneracy is NP-Complete. $\square$

A straightforward corollary of Theorem 4.3 is that:

Corollary 4.4. *It is Co-NP-Complete to decide whether a game is nondegenerate.*

In the following section, we want to study a special class of bimatrix games: win-lose games. In win-lose bimatrix games, the payoff value is either 0 or 1. The following is a necessary and sufficient condition for a win-lose bimatrix game to be nondegenerate.

Theorem 4.5. *For a win-lose bimatrix game, it is nondegenerate iff for the row player, every column of its payoff matrix A has exactly one nonzero element while for the column player, every row of its payoff matrix B has one nonzero element, too.*

Proof. We prove our claim only for the row player; the proof for the column player follows symmetrically. Suppose a win-lose game is nondegenerate, each column of matrix A cannot have more than one nonzero item and cannot have all zeros. Otherwise, if the column player plays a pure strategy of that column, the row player has more than one best responses. Thus, the game is degenerate. Contradiction occurs. On the other hand, if each column of matrix A has exactly one nonzero element, for each strategy y of the column player, the number of nonzero elements in the vector Ay cannot be more than the size of support in y. Thus, the game is nondegenerate. □

Therefore, it is in P to decide whether a win-lose bimatrix game is degenerate.

CHAPTER V

Path Auctions with Multiple Edge Ownership

5.1 Motivations

In the path auction game, there is a network $G = (V, E)$ in which each edge $e \in E$ is owned by an agent. The true cost of e is private information and known only to the owner. Given two vertices, source s and destination t, the auctioneer's task is to buy a path from s to t. The path auction game can be used to model problems in supply chain management, transportation management, QoS routing, and other domains. Recently, path auctions have been extensively studied [64, 50, 6, 46, 34]; most of the literature has focused on the Vickrey-Clarke-Groves (VCG) class of mechanisms [80, 18, 38]. In the most natural VCG mechanism, the auctioneer pays each agent on the winning path an amount equal to the highest bid with which the agent would still be on the winning path. This mechanism is attractive because it is efficient and strategyproof—the dominant strategy for each agent is to report his true cost.

In the traditional path auction model, each agent only owns one edge in the graph and there is no cooperation between agents. Here, we study a variant of the path auction game in which *each agent may own multiple edges*. In this extended model, if ownership information is publicly available (i.e., the auctioneer knows which agent owns which edge), then the VCG mechanism design approach still yields a strategyproof mechanism.

In practice, however, ownership information is often private – it could be costly for the auctioneer to find out the true ownership information, or an agent may have incentive to hide the true ownership information in order to get a better payoff. For example, in Figure 5.1, there are two agents: a and b. Agent a owns edges (s,i) and (i,t) with true cost 1 each; agent b owns edges (s,j) and (j,t) with true cost 2 each. If the true ownership information was known to the auctioneer, then the most natural VCG mechanism would reduce to a second price auction: it would choose path (s,i), (i,t) as the winning path and pay agent a an amount equal to 4. However, if agent a hides his ownership information, the mechanism will treat edges (s,i) and (i,t) as if they were owned by different agents. When the agents bid their true costs, the winning path stays the same, but the payment to agent a would be $2 \times 3 = 6$. Moreover, when ownership information is not available to the auctioneer, agent a can increase his utility by bidding lower than his true cost. For example, he could bid 0.5 for both edges (s,i) and (i,t). This does not change the winning path, but the payment to agent a would increase to $2 \times 3.5 = 7$. Hence, the straightforward VCG mechanism, which assumes that each edge is owned by a distinct agent, is not strategyproof in this extended model.

In this chapter, we model situations in which each agent can own multiple edges at the same time, but ownership information is private. Thus, the traditional path auction model is a special case of our extended model. The possibility of one agent having multiple identities is inherent in online communities. In an online auction system, each seller/buyer may have multiple accounts in the system. Now, if a buyer wants some combination of goods that can be expressed in path auction form, it will be hard for him to find the true identity of each seller account, so he will be faced

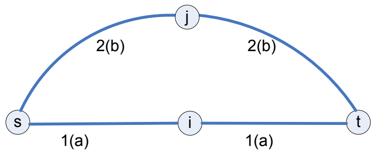

Figure 5.1: VCG mechanism is not strategyproof for this game

with the unknown-ownership scenario.

5.2 Definitions and Problem Statement

We now introduce formal definitions of the path auction game based on set systems defined in [50]. We begin by defining simple path auctions, in which ownership of each edge is known to be distinct. We then define an extended path auction model that incorporates multiple edge ownership.

5.2.1 Simple path auctions

Given a graph $G = (V, E)$, let each edge $e \in E$ be owned by a distinct agent and have a cost c_e, the true cost incurred by the owner if the edge is selected. This value is private (known only to agent who owns e). We define the feasible set $F = \{P|\ P$ is a path from the source s to the destination $t\}$. Note that a path is represented by the set of its edges. The task of the auctioneer is to buy a path from s to t by auction. It consists of the following two steps:

(1). Each agent submits a sealed bid b_e to the auctioneer. The bidding vector b is $(b_{e_1}, b_{e_2}, ..., b_{e_m})$, where b_{e_i} is the bidding price for the edge $e_i \in E$. Moreover, let B denote the bidding space, which is the set of all possible bidding vectors.

(2). Given the bidding vector b, the auctioneer selects a path P from the feasible set F as the winning path, and computes a payment p_e for each edge $e \in P$. An agent wins if he owns an edge e on the winning path P^j; otherwise he loses.

In order to implement the auction, we need to design a mechanism $(f, p_1, p_2, ..., p_m)$ where $f : B \to F$ selects one element in the feasible set as the winning path and $p_i : B \to R$ computes the payment to agent i. Moreover, we assume that:

(1). (G, F) is common knowledge to the auctioneer and all the agents.

(2). The game is *monopoly free*, which means no edge is in all the feasible sets, i.e., $\bigcap_{P^j \in F} P^j = \varnothing$.

(3). The agents are rational and have quasilinear utilities. The utility is defined to be the payment minus the incurred true cost, i.e., $u_e = p_e - c_e$ if e is on the winning path; otherwise, $u_e = p_e$. The agents want to maximize their utilities.

5.2.2 Extended path auctions

We extend the simple path auctions model to allow for the possibility that each agent owns multiple edges. The edge set E can be partitioned as $E = \bigcup_i E_i$, where E_i is the set of edges owned by agent i. We also assume that an agent i that owns k_i edges, i.e., $|E_i| = k_i$, can have up to k_i identities $ID_i = \{ID_{i1}, ID_{i2} ..., ID_{ik_i}\}$ to use in an auction (one for each edge). We assume that E_i and ID_i are private information, which is only known to agent i. Moreover, for two different agents i and j, $ID_i \bigcap ID_j = \varnothing$, meaning that an agent cannot claim an identity that belongs to another agent for one of his own edges. In the extended model, a game is monopoly free if, for any agent i, there is at least one path between s and t in graph $(V, E \setminus E_i)$. Furthermore, given the winning path P, the utility for agent i is $u_i = \sum\limits_{e_i \in E_i} p_{e_i} - \sum\limits_{e_i \in P} c_{e_i}$. According to different formats of bidding languages, we can define two formats of extended path auctions.

- *Format I*: in this type of auction, each agent is only asked to submit the bidding price for each edge he owns, while keeping the identity information private.

- *Format II*: in this type of auction, the agent is asked to submit identity information (possibly false) about the set of edges he owns as well as the bidding price for each edge he claims to own.

Next, we will introduce some basic game theory definitions in the extended path auction model. Since an agent can own multiple edges in the extended model, the agent needs to submit a bidding price for each edge that belongs to him. Thus, let b_i be the bidding vector of agent i while c_i is the true cost vector of agent i. Furthermore, let b_{-i} denote the bidding vector of all agents except agent i.

Definition 5.1. A strategy profile b is an ϵ-Nash equilibrium if for any agent i and any bidding vector $b_i^{'}$ of agent i, given the bidding vector b_{-i} of other agents, $u_i(b_i, b_{-i}) \geq u_i(b_i^{'}, b_{-i}) - \epsilon$.

Definition 5.2. A bidding strategy b_i of agent i is a dominant strategy, if for any bidding vector b_{-i} of other agents and any bidding vector $b_i^{'}$ of agent i, $u_i(b_i, b_{-i}) \geq u_i(b_i^{'}, b_{-i})$.

In auctions of format I, we introduce the concept of strategyproofness.

Definition 5.3. A format I auction mechanism is strategyproof if it is a dominant strategy for each agent i to bid his true value, i.e., $b_i = c_i$.

The VCG mechanism is strategyproof in the simple path auction, that is, the dominant strategy of each agent is to bid his true cost in VCG mechanism. Besides strategyproofness, individual rationality, which requires that no agent should be paid less than the cost he incurs, is another important concept studied in mechanism design.

Definition 5.4. A mechanism is individually rational iff for any agent i, the payment to agent i is at least his bidding price if he wins (so every agent should have nonnegative utility in the mechanism if they bid truthfully).

In auctions of format II, we introduce the solution concept of false-name-proofness.

Definition 5.5. A format II auction mechanism is false-name-proof if for any agent, it is a dominant strategy to report the true identity information of each edge as well as the true cost of each edge he owns.

In both formats of extended path auctions, a mechanism will choose a winning path from feasible paths and make payments to the agents. The definition of Pareto efficiency is:

Definition 5.6. A winning path selection rule is *Pareto efficient* if the mechanism chooses P^i as the winning path such that $\forall k, \sum\limits_{e \in P^i} c_e \leq \sum\limits_{e \in P^k} c_e$.

In economic literature, Pareto efficiency is defined as an allocation in which there is no agent can be better off without making another agent worse off. The definition of Pareto efficiency in this chapter follows from the definition in [85], which is different from the standard definition. However, in the quasilinear utility setting we assume that the two definitions are equivalent.

5.3 The Nonexistence of Individually Rational Strategyproof Mechanisms

In the extended path auction model, one natural question is whether it is possible to design an auction mechanism that asks agents to report only edge costs (format I), such that it is in every agent's best interest to bid the true cost for each edge he owns. In this section, we show that no reasonable mechanism can meet this requirement. Note that the strategyproofness condition defined in this chapter is essentially equivalent to the bribe-proof condition in [72]. However, since path auction games do not satisfy the monotonically closed condition, the results in [72] cannot be applied here.

We begin with a characterization of strategyproof auction mechanisms that is well known from the literature on auctions. We state the theorem in the following, as we will rely on it in our proofs.

Theorem 5.7. *[50, 5, 53] The set of individually rational strategyproof auction mechanisms, in which all the losing agents are paid 0, can be characterized as follows:*

(1). A mechanism is strategyproof only if the selection rule is monotone: No losing agent can become a winner by raising his bid, given fixed bids by all other agents.

(2). Given a monotone selection rule, there is a unique strategyproof mechanism with this selection rule. This mechanism pays each agent his threshold bid, i.e., the supremum of all values he could have bid and won.

Actually, we can construct a trivial strategyproof mechanism, which always selects a fixed path as the winning path and pays a fixed amount of money to the edges on the path. We call such a mechanism the *dictator mechanism.* It is not hard to verify that the dictator mechanism is strategyproof. However, it does not satisfy individual rationality.

Our impossibility proof builds on the characterization of strategyproof mechanisms. We begin with the following lemma, which shows that for any individually rational strategyproof mechanism, any bid vector can be perturbed slightly to ensure that all winners have strictly positive net utilities.

Lemma 5.8. *Consider individually rational strategyproof mechanisms for the extended path auction model in which edges not on the winning path are always paid* 0*. For any such mechanism* $(f, p_{e_1}, ..., p_{e_m})$*, any* $\epsilon > 0$*, and any strictly positive cost* *vector* $b^1 = (b^1_{e_1}, ..., b^1_{e_m})$*, there exists another cost vector* $b^{'} = (b^{'}_{e_1}, ..., b^{'}_{e_m})$ *such that: (1) When the agents have true costs given by* $b^{'}$ *and bid truthfully, every edge on the winning path has strictly positive utility; (2) for all* j*,* $|b^1_{e_j} - b^{'}_{e_j}| \leq \epsilon$*.*

Proof. Suppose that b^1 is a given cost vector, and that when the agents bid according to b^1, the winning path is P^1, i.e., $f(b^1) = P^1$. By assumption, the edges not on the winning path have utility 0 since they are paid 0 and have 0 incurred cost. If every edge on the winning path is paid strictly more than its bid, we are done. Otherwise, some of the edges on the winning path have utility exactly 0; in this case, we will modify the cost (and bid) vector according to the following procedure.

In the first modification step, for an edge $e_1 \in P^1$, we decrease the costs of e_1 from $b^1_{e_1}$ to $b^1_{e_1} - \epsilon$, and keep the cost of all other edges unchanged. Let $\mathcal{T} = \{e_1\}$. Thus we get a new true cost vector b^2; let $P^2 = f(b^2)$. According to Theorem 5.7, e_1 must be on the new winning path P^2. Moreover, we claim that the payment to the edge e_1 should not change, that is, $p_{e_1}(b^1) = p_{e_1}(b^2)$. Otherwise, if $p_{e_1}(b^1) > p_{e_1}(b^2)$, an edge with true cost $b^1_{e_1} - \epsilon$ could improve its utility by increasing its bidding price to $b^1_{e_1}$. If $p_{e_1}(b^1) < p_{e_1}(b^2)$, an edge with true cost $b^1_{e_1}$ could improve its utility by decreasing its bidding price to $b^1_{e_1} - \epsilon$. Both cases contradict the strategyproofness of the mechanism. Since $p_{e_1}(b^1) = p_{e_1}(b^2)$, when the true cost of the edge e_1 decreases from $b^1_{e_1}$ to $b^1_{e_1} - \epsilon$, its utility will increase by ϵ; thus, e_1 will have strictly positive utility under cost vector b^2.

In the kth step, where $k \geq 2$, we choose an edge $e_k \in P^k \backslash \mathcal{T}$. We decrease the cost of e_k from $b^k_{e_k}$ to $b^k_{e_k} - \frac{\epsilon}{2^{k-1}}$ and keep the bidding prices of all other edges unchanged. Hence, we get a new bidding vector b^{k+1} and $f(b^{k+1}) = P^{k+1}$. Let $\mathcal{T} = \mathcal{T} \bigcup \{e_k\}$. Similar to the argument above, the edge e_k must be on the new winning path P^{k+1}, and its utility must have increased by $\frac{\epsilon}{2^{k-1}}$ because the payment to it does not change.

Moreover, we claim that any edge $e_j \in \mathcal{T}$ is still on the new winning path P^{k+1} and its utility cannot decrease by more than $\frac{\epsilon}{2^{k-1}}$, *i.e.*, $p_{e_j}(b^{k+1}) - p_{e_j}(b^k) \geq -\frac{\epsilon}{2^{k-1}}$. Otherwise, suppose $p_{e_j}(b^{k+1}) - p_{e_j}(b^k) < -\frac{\epsilon}{2^{k-1}}$. It can happen that both edges e_j and e_k are owned by the same agent i, and the true cost of the edge e_k is $b^k_{e_k} - \frac{\epsilon}{2^{k-1}}$. Thus, agent i can increase his utility by increasing the bidding price of the edge e_k to $b^k_{e_k}$. This contradicts strategyproofness. Furthermore, since the utility of the edge $e_j \in \mathcal{T}$ cannot decrease by more than $\frac{\epsilon}{2^{k-1}}$ at the kth step and $\frac{\epsilon}{2^{j-1}} > \sum_{i=j+1}^{N} \frac{\epsilon}{2^{i-1}}$ for any finite number $N > j$, the edge e_j always has positive utility. Therefore, given the assumption that the edges not on the winning path are always paid 0, the edge e_j is always on the new winning path P^k for $k \geq j$. So, all the edges in $\mathcal{T}$ are on the winning path and have positive utilities.

Since the number of edges is finite, the above procedure must stop. When the process is terminated, suppose the winning path $P = \mathcal{T}$ and the final cost vector is b'. According to the argument above, every winning agent must have positive utility when the agents have true costs b' and bid truthfully. $\square$

We now prove the main result. The intuition behind this result is that for strategyproofness to hold in the extended model, the payments made to the set of winning edges must increase if all costs increase; however, this implies that agents can profit by inflating their bids.

Theorem 5.9. *There is no individually rational strategyproof format I mechanism in which all the edges not on the winning path are always paid 0.*

Proof. Suppose there is a strategyproof mechanism. By Lemma 1, consider a bidding vector b such that losing agents have zero utilities while winning agents have positive utilities. For an edge e_j that is not on the winning path P, increasing b_{e_j} to b'_{e_j}

cannot change the winning path. We prove this by contradiction.

Suppose the winning path changes from P to $P^{'}$, then there exists an edge $e \in P$ but $e \notin P^{'}$. According to Theorem 5.7, e_j cannot be on $P^{'}$. Thus its utility is still 0. Now, suppose an agent i owns both e_j and e. If the true cost of e_j is $b^{'}_{e_j}$, agent i can increase his utility by understating e_j's true cost as b_{e_j}, since e has positive utility when it is on winning path P. This contradicts strategyproofness. Moreover, increasing b_{e_j} does not change the payment to any edge. For any edge $e \notin P$, the payment to it is always 0. Suppose increasing b_{e_j} to $b^{'}_{e_j}$ could increase the payment to the edge $e \in P$ from p_e to $p^{'}_e$, *i.e.*, $p_e < p^{'}_e$. It could happen that an agent i owns both e_j and e, and the true cost of the edge e_j is b_{e_j}. Thus, agent i can increase his utility by overstating e_j's cost as $b^{'}_{e_j}$. This contradicts strategyproofness too. Similarly, we can show that increasing b_{e_j} to $b^{'}_{e_j}$ cannot decrease the payment to an edge $e \in P$. Therefore, for an edge e_j that is not on the winning path, increasing its bid cannot change either the winning path or the payment to any agent. Similarly, we can prove that for an edge e_j that is on the winning path, decreasing b_{e_j} cannot change either the winning path or the payment to any edge.

Now we can construct a suitable pair of bid vectors that result in the same path being selected. According to Lemma 5.8, for any individually rational strategyproof mechanism $(f, p_{e_1}, ..., p_{e_m})$ and any strictly positive bidding vector b, we can construct a sequence of bidding vectors $b(r) = (r \times b_{e_1} - \epsilon(e_1, r), ..., r \times b_{e_m} - \epsilon(e_m, r))$ such that the winning agents always have positive utilities, where $r \in \mathbb{N}$ and $\forall j, \epsilon(e_j, r)$ is a small positive number. Let $\mathcal{B} = \{b(r) | r \in \mathbb{N}\}$ denote the set of all such bidding vectors. For each $b(r) \in \mathcal{B}$, $f(b(r))$ will select a winning path. Since the size of $\mathcal{B}$ is infinite, but there are only a finite number of possible winning paths, there must be an infinite subsequence $\mathcal{B}^{'} = \{b(r_1), b(r_2)......\}$ such that $\forall b(r_i) \in \mathcal{B}^{'}$, $f(b(r_i)) = P$.

According to the assumption of individual rationality, and given that the payment to each edge is finite, we can find two bidding vectors $b(p), b(q) \in \mathcal{B}'$ such that for any edge e_j on the winning path P, $b_{e_j}(p) \le p_{e_j}(b(p)) < b_{e_j}(q) \le p_{e_j}(b(q))$.

Given bidding vectors $b(p)$ and $b(q)$, we can construct a new bidding vector b^* such that $b^*_{e_j} = \min\{b_{e_j}(p), b_{e_j}(q)\}$ if e_j is on the winning path P while $b^*_{e_j} = \max\{b_{e_j}(p), b_{e_j}(q)\}$ if e_j is not on the wining path. According to the construction and the above arguments, we can get $\forall e_j$, $p_{e_j}(b(p)) = p_{e_j}(b^*) = p_{e_j}(b(q))$. This contradicts the fact that for any edge e_j on the winning path P, $p_{e_j}(b(p)) < b_{e_j}(q) \le p_{e_j}(b(q))$.

Therefore, given the condition that the edges not on the winning path always get 0 payment, there is no strategyproof mechanism that satisfies individual rationality.

□

Remark: In the proofs of Lemma 5.8 and Theorem 5.9, the contradictions occur even if each agent owns only two edges. Therefore, in the extended path auction game, there is no individually rational strategyproof mechanism in which the edges not on the winning path are paid 0, even if each agent can only have two edges. We believe that if we remove the assumption that the edges not on the winning path always get 0 payment, Theorem 5.9 still holds. It would be interesting to find a simple proof for such an extension of Theorem 5.9.

Given Theorem 5.9, the next natural question to ask is: if the auctioneer asks the agents to submit ownership information as well as the bidding price information (format II), is it possible to get a false-name-proof [85] mechanism? In [85], Yokoo *et al.* showed that:

Proposition 5.10. *[85] In combinatorial auctions, there is no false-name-proof auction protocol that satisfies Pareto efficiency.*

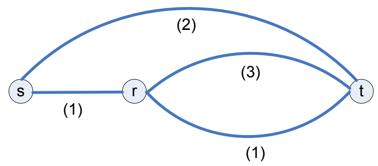

Figure 5.2: No false-name-proof mechanism that satisfies Pareto efficiency

Proposition 5.10 is shown by constructing a generic combinatorial auction example, which does not have a false-name-proof auction protocol that satisfies Pareto efficiency. That example has two items to sell and three bidders, where bidder 1 is interested in packages {1},{2} and the whole package {1,2}, bidder 2 is only interested in the whole package {1,2} and bidder 3 is only interested in {2}. In particular, bidder 1 has two identities to use. Furthermore, the bidders are given an option to quit this game by bidding 0.

We now observe that this counterexample can be viewed as an instance of a path auction (which is a special class of combinatorial auctions). Each single item in the combinatorial auction corresponds to an edge in the path auction while the whole package {1,2} corresponds to a path between s and t. The path auction game is shown in Figure 5.2. The true identity of the agent that each edge belongs to is represented as an integer in the parentheses. Upon the transformation, the proof of Proposition 5.10 still works for the path auction setting. Thus, the following result is immediate.

Corollary 5.11. *In the extended path auction model, there is no false-name-proof mechanism that satisfies Pareto efficiency.*

Proof. In order to prove this corollary, we first transform the generic combinatorial auction example of Yokoo *et al.* [85] to a path auction case as described above. Note that in the combinatorial auction example, the feasible set with the highest bidding price will be the winner. However, path auctions, as a type of procurement auction, will choose the feasible set with the lowest bidding price as the winning path. To accommodate this difference, we change the directions of all the inequalities in the proof of Proposition 5.10 in [85]. Moreover, if a bidding price is perturbed by $+\epsilon$ in [85], it is perturbed by $-\epsilon$ in the path auction case instead. Thus, the modified proof of Proposition 5.10 in [85] works exactly for the path auction case and the corollary is proved. $\square$

5.4 Existence of a Pareto Efficient Pure Strategy ϵ-Nash Equilibrium

Since strategyproofness is not feasible in the extended path auction model, we can consider weakening the solution concept. The concept of ϵ-Nash equilibrium, which has been applied to path auctions in [46], is a natural candidate. In this section, we study the existence of ϵ-Nash equilibria in the extended path auction model under the first-price auction mechanism [46], which elicits the bids from the agents, chooses the cheapest path with respect to the bidding vector as the winning path, and pays each winning agent exactly the bidding price.

Immorlica *et al.* [46] showed that there is always an ϵ-Nash equilibrium in the simple path auction model. One drawback of the concept of Nash equilibria is that an arbitrary ϵ-Nash equilibrium may have low social walfare. Hence, we require that the Nash equilibria studied below satisfies Pareto efficiency. Another drawback of the concept of ϵ-Nash equilibrium is that there exist equilibria in which losing agents bid below their true costs. To eliminate such unnatural equilibria, we assume that the bidding price of each edge is at least its true cost, that is, $\forall e, b_e \geq c_e$. With those natural constraints, we will show that, because of the multiple edge ownership, a Pareto efficient ϵ-Nash equilibrium only exists for a limited class of graphs in extended path auctions.

First, we show the existence of a Pareto efficient ϵ-Nash equilibrium in the parallel-path graph [33], which can be defined as:

Definition 5.12. A parallel-path graph is a graph that can be represented as $\bigcup_k P^k$, where P^k is the k_{th} path from s to t and $\forall i \neq j$, $P^i \bigcap P^j = \varnothing$.

Let $C(P^k) = \sum_{e \in P^k} c_e$ denote the cost of path P^k with respect to true cost vector c. Consider a sorted list of paths from low to high according to these true costs, that is, the path with lower cost has smaller index. For any agent A_i, let $L(A_i)$ be the smallest path index such that path $P^{L(A_i)}$ does not have an edge owned by agent A_i, but all paths with smaller path indices than $L(A_i)$ have at least one edge owned by agent A_i. We compute $L(A_i)$ for each agent that has at least one edge on P^1. Note that $L(A_i)$ must exist under the monopoly-free assumption on the set system. We constructively find a Pareto efficient pure strategy ϵ-Nash equilibrium for parallel-path graph in the following theorem. The proof is motivated by [46].

Theorem 5.13. *If the underlying network is a parallel-path graph, the first-price path auction has a Pareto efficient pure strategy ϵ-Nash equilibrium.*

Proof. The ϵ-Nash equilibrium bidding vector is constructed as follows. Initially, suppose that each agent bid his true cost, i.e., $b = c$. Let $W_b(P^k) = \sum_{e \in P^k} b_e$ denote the cost of path P^k with respect to the bidding vector b. Pick an agent A_k who has at least one edge on P^1, and who has the highest value of $L(A_k)$ of all agents that have edges on P^1. In order to find an ϵ-Nash equilibrium bidding vector, we first pick one edge in $E_{A_k} \bigcap P^1$, and increase its bidding price by $W_b(P^{L(A_k)}) - W_b(P^1) - \epsilon$ if $W_b(P^{L(A_k)}) - W_b(P^1) > \epsilon$; otherwise, the bidding price of the edge remains unchanged. For any path $j \in [2, ..., L(A_k) - 1]$, we pick one edge in $E_{A_k} \bigcap P^j$ and increase its bidding price until $W_{b'}(P^j) = W_b(P^{L(A_k)})$, where b' is the new bidding vector. We call the final bidding vector b^f. Note that the bidding price for each edge in b^f, except those belonging to agent A_k, is exactly its true cost. Since the first-price auction mechanism always selects the path with the minimum cost, P^1 is the winning path under b^f and A_k has at least one edge on it.

We claim that b^f is an ϵ-Nash equilibrium bidding vector. This is proved by analyzing the strategies of all the agents in three cases:

Case I: If an agent A_i is a losing agent, he will bid the true costs for all of the edges in E_{A_i}. Consider the following two subcases: (i) For any path j, if the sum of the bidding prices for the edges in $E_{A_i} \bigcap P^j$ increases, it cannot change the winning path; therefore, the utility of agent A_i cannot be improved; (ii) If the sum of the bidding prices for the edges in $E_{A_i} \bigcap P^j$ decreases such that P^j becomes the winning path, the utility of A_i would be negative. Thus, agent A_i cannot improve his utility through a unilateral deviation.

Case II: For an agent $A_i \neq A_k$ who owns edges on the winning path, according to the definition of b^f, the edges in E_{A_i} would bid their true costs, too. Similar to the above analysis, agent A_i cannot improve his utility by decreasing the bidding prices. On the other hand, for any index $j \in [1, ..., L(A_k)]$, suppose agent A_i increases the sum of the bidding prices for the edges in $E_{A_i} \bigcap P^j$. Again, let $b^{'}$ be the new bidding vector. Note that there always exists an index $r \in [1, ..., L(A_k)]$ such that $E_{A_i} \bigcap P^r = \varnothing$ and $W_{b'}(P^r) = W_{b^f}(P^{L(A_k)})$ remains unchanged. Thus, for any path P^j, the sum of the bidding prices for the edges in $E_{A_i} \bigcap P^j$ cannot increase by more than ϵ; otherwise, P^r will be the winning path and agent A_i has zero utility. Therefore, agent A_i cannot increase his own utility by more than ϵ through a unilateral deviation.

Case III: For agent A_k, his utility is at least $W_{b^f}(P^{L(A_k)}) - W_c(P^1) - \epsilon$. Similar to the analysis in case II, agent A_k cannot improve his utility more than ϵ by increasing

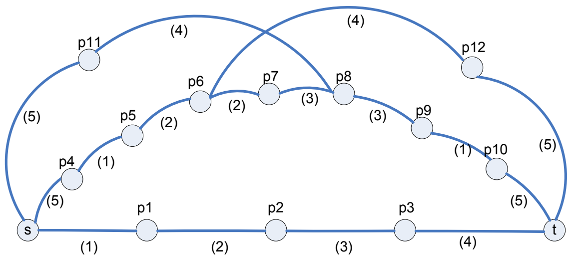

Figure 5.3: Auction with no Pareto efficient pure-strategy ϵ-Nash equilibrium

the bidding prices of the edges he owns. Otherwise, $P^{L(A_k)}$ would be the winning path and the utility of agent A_k would be 0. On the other hand, for any path P^j with an index $j < L(A_k)$, if the sum of the bidding prices for the edges in $E_{A_k} \bigcap P^j$ decreases such that the path P^j is the winning path, then the utility of agent A_k must be less than $W_{b^f}(P^{L(A_k)}) - W_c(P^1)$. Thus, agent A_k cannot improve his utility by more than ϵ through a unilateral deviation.

It is clear that b^f is Pareto efficient since the winning path P^1 has the minimum true cost. Therefore, any parallel-path graph can have a Pareto efficient pure strategy ϵ-Nash equilibrium when the agents bid according to b^f. $\square$

Although there exists a Pareto efficient pure strategy ϵ-Nash equilibrium for path auctions with parallel-path graphs, we can find a non-parallel-path graph that does not have any Pareto efficient pure strategy ϵ-Nash equilibrium. We show this counter example in Figure 5.3; for each edge label, the integer in parentheses denotes the identity of the agent who owns that edge.

Proposition 5.14. *The graph shown in Figure 5.3 cannot have any Pareto efficient pure strategy ϵ-Nash equilibrium in the first-price path auction mechanism.*

Proof. There are 5 agents in this game and 5 paths from s to t:

Path 1: (s, p_1, p_2, p_3, t)

Path 2: $(s, p_4, p_5, p_6, p_7, p_8, p_9, p_{10}, t)$

Path 3: $(s, p_4, p_5, p_6, p_{12}, t)$

Path 4: $(s, p_{11}, p_8, p_9, p_{10}, t)$

Path 5: $(s, p_{11}, p_8, p_7, p_6, p_{12}, t)$

Let b be a Pareto efficient ϵ-Nash equilibrium bidding vector. We assume that the edges on P^1 have small true costs while all other edges have significantly large true costs. By Pareto efficiency, P^1 will be the winning path. Furthermore, we claim that the cost of each path with respect to b can differ at most by ϵ. We prove this by contradiction. Suppose $\exists j \in [2...4], W_b(P^j) > W_b(P^1) + \epsilon$. For any agent $i \in [1...4]$ in Figure 5.3, there is only one path that does not have edges owned by him. We can assume that the path P^j does not have edges owned by agent i, but for four other paths, agent i owns edges on each of them. Note that agent i owns only one edge on the winning path P^1. Thus, agent i can increase the bidding prices of his edges (but still keep P^1 as the winning path) such that his utility is increased by at least ϵ, which leads to a contradiction. Therefore, if b is a Pareto efficient ϵ-Nash equilibrium bidding vector, then the cost of each path with respect to b can differ at most by ϵ. Let

$$A = b_{p_6,p_7} + b_{p_7,p_8}$$

$$B = b_{p_8,p_9} + b_{p_9,p_{10}} + b_{p_{10},t}$$

$$C = b_{p_6,p_{12}} + b_{p_{12},t}$$

Based on the above reasoning, we can get the following two inequalities:

$$|A + C - B| \leq \epsilon \qquad and \qquad |A + B - C| \leq \epsilon$$

Therefore, the following inequality holds:

$$2A \leq 2\epsilon$$

Moreover, according to our assumption that the bidding price of each edge is at least its true cost, that is, $\forall e, c_e \leq b_e$, the following inequality holds:

$$c_{p_6,p_7} + c_{p_7,p_8} \leq \epsilon$$

When ϵ is small enough and the true cost of each edge is large enough, contradiction occurs. So, there is no Pareto efficient pure strategy ϵ-Nash equilibrium for the graph in Figure 5.3 using the first-price path auction mechanism. □

Remark : In the proof of Proposition 5.14, we implicitly assume that $c_{p_6,p_7} = c_{p_7,p_6}$ as well as $c_{p_7,p_8} = c_{p_8,p_7}$, that is, edges (p_6, p_7) and (p_7, p_8) are undirected. However, the proof still works for directed graphs. To see this, suppose (p_6, p_7, p_8) is a directed path from p_6 to p_8 while there is another directed path P' from p_8 back to p_6 in Figure 5.3. Moreover, assume that the edges on the path P' are owned by agents 2 and 3. Then, we can get $|b_{P'} + C - B| \leq \epsilon$, where $b_{P'}$ is the sum of the bidding prices of all the edges on P', and $|A + B - C| \leq \epsilon$. Thus, the final inequality of the above proof is changed to $c_{p_6,p_7} + c_{p_7,p_8} + c_{P'} \leq 2\epsilon$, where $c_{P'}$ is the true cost of the new directed path from p_8 back to p_6. Contradiction still occurs.

5.5 Related works

Path auction games have been extensively studied in recent years. Nisan and Ronen introduced the shortest-path game in their paper on algorithmic mechanism design [64], and showed that the VCG mechanism for this problem is computationally tractable. However, several authors have noted that the VCG mechanism may pay much more than the true cost of the winning path. This has led to the study of the frugality [6] of the VCG mechanism. Archer and Tardos [6] as well as Elkind *et al.* [34] studied frugality in path auctions and showed that payments can be arbitrarily high.

Karlin *et al.* [50] extended the path model to a more general set system model and introduced a new frugality ratio definition. They designed a mechanism that performs better than the VCG in path auctions. The problem of an agent owning multiple edges was mentioned as future work in [50]. Immorlica *et al.* [46] studied first-price path auctions in the traditional single-ownership setting. They showed the existence of a strong ϵ-Nash equilibrium in bids, and bounded the payments in equilibria. J. Schummer [72] studied the bribe-proof auction, in which no agent can pay another one to lie so that both of them are better off. Schummer showed that if the domain is rich, the only bribe-proof mechanism is a constant function. However, path auction games do not satisfy the monotonically closed condition ("richness" condition) in [72]. Yokoo *et al.* [85] introduced the concept of false-name-proof mechanisms, in which the (weakly) dominant strategy for each agent is to report his true values as well as true identities, and showed that in combinatorial auctions there is no false-name-proof mechanism that satisfies Pareto efficiency. The problem of unknown ownership [60] has also been studied in the context of job scheduling by Moulin.

The results we report here were presented in preliminary form at the NetEcon06 Workshop [29]. Following our work, Iwasaki *et al.* [47] recently designed two mechanisms MP and AP for path auction games. The MP mechanism is false-name-proof (can't self-split) when each agent owns only one edge. Moreover, a nice property of the MP mechanism is that its frugality ratio nearly matches a lower bound of any false-name-proof mechanism. The AP mechanism is false-name-proof when an agent can own multiple edges. However, one drawback of the AP mechanism is that it does not always buy a feasible solution.

CHAPTER VI

Using Spam Farms to Boost PageRank

6.1 Motivations

In the past decade, search engines such as Google, Yahoo, and MSN have played a more and more important role in our everyday lives. Therefore, Web sites that show up on the top of query results lists have had an ever increasing economic advantage. This has given people incentives to manipulate the search results by carefully designing the content or link structure of a Web page. This is called Web spamming [41]. The emergence of Web spamming would undermine the reputation of a trusted information resource. A study in 2002 indicated that around 6 to 8 percent of the Web pages in a search engine index were spam [36]. This number increased to around 15 to 18 percent from 2003 to 2004 [42, 1]. This increasing tendency made many researchers believed that Web spamming would become a major challenge to preserving the integrity of Web searches [43].

There are two major categories of Web spamming techniques: *term spamming* and *link spamming*. Term spamming boosts the ranking of a target page by editing a page's textual content. For example, one can add thousands of irrelevant keywords as hidden fields to the target page. A search engine will index those keywords and return the target page as answers to queries that contain those keywords. Link spamming, on the other hand, manipulates the interconnected link structure of controlled pages, called a *link spam farm*, to boost the connectivity based ranking of the target page to be higher than it deserves [40]. PageRank [12], the well-known connectivity based ranking algorithm used by the leading search engine Google, is the most-popular manipulating target for a spammer [1]. Compared to term spamming, link spamming is harder to detect as it can boost the ranking of a target page without changing the content.

6.2 Single-Target Spam Farm Model

In order to boost the ranking of some Web pages in the web graph, a spammer often sets up groups of Web pages with a carefully devised structure. The group of pages fully controlled by a spammer is called a spam farm while non-spam-farm pages are called normal pages. The simplest spam farm model is the single-target spam farm model [39], which has the following characteristics:

(1). Each spam farm has a single target page and a fixed number of boosting pages.

(2). The spammer wants to boost the target page by adding or deleting the outgoing links of the boosting pages and the target page.

(3). It is possible for the spammer to accumulate links from Web pages, such as public bulletins and blogs, outside the spam farm. These links and external pages are called hijacked links and hijacked pages, respectively. The total PageRank score that reaches the spam farm from the hijacked links is refer to as the leakage.

[1] Nowadays, what Google actually uses is a much more sophisticated ranking method, in which PageRank only takes a relatively small factor.

6.3 Our Results

In this chapter, we first characterize the optimal spam farm by using the sensitivity analysis of Markov chains. Under realistic assumptions, the optimal spam farm (ref. Figure 6.1) should have the following features: 1) The boosting pages point to and only to the target page; 2) The target page points to and only to some of the generous pages, which are the Web pages that only point to the target page; 3) the spammer accumulates as many hijacked links as possible.

In an optimal spam farm, the boosting pages, as well as all the hijacked links that the spammer is able to accumulate, must point to the target page. This structure is easy to detect. In order to disguise the spam farm, the spammer may deviate from the optimal one. Thus, we also characterize optimal spam farms under some realistic constraints. We show that in the optimal spam farm, if the target page must point to some non-spam-farm pages, then the target page should point to all the generous pages (ref. Figure 6.2); if some of the boosting pages cannot directly point to the target page, then they should point to some of the generous pages (ref. Figure 6.3); if the hijacked links cannot directly point to the target page, then the spammer should accumulate as many hijacked links pointing to the boosting pages as possible (ref. Figure 6.4).

6.4 The PageRank Algorithm

We follow [54, 12] to define PageRank. Let $G = (V, E)$ [2] be a directed graph with vertex set V and edge set E. We assume that there is no self-loop in G. Let $N = |V|$, and for a vertex $i \in V$, denote by $\mathrm{out}(i)$ the out-degree of i. The transition

[2] We assume that there is no self-loop in the web graph. All our results can be easily extended by following the same proof ideas in this chapter if self-loops are allowed.

matrix of G is $T = [T_{ij}]_{1\le i,j\le N}$:

$$T_{ij} = \begin{cases} \frac{1}{\mathrm{out}(i)} & \text{if } (i,j) \in E \\ 0, & \text{otherwise} \end{cases}$$

Denote by $e \in \mathbb{R}^N$ the all 1 row vector $(1, 1, \cdots, 1)$, and by $E \in \mathbb{R}^{N\times N}$ the all 1 matrix. Let $\bar{T}$ be identical to T except that if a row in P is all 0, it should be replaced by e/N. A page without outgoing links is called a *dangling* page. For some constant c, $0 < c < 1$, the transition matrix for the PageRank Markov chain is

$$P = c\bar{T} + (1-c)E/N.$$

The PageRank π is the stationary distribution, that is, $\pi P = \pi$, of the above Markov chain P. Our definition of PageRank score is different from the definition in [39], where the PageRank score $\bar{\pi}$ is defined as $\bar{\pi} = c\bar{\pi}T + \frac{(1-c)}{N}e$. However, the two definitions yield the same relative PageRank scores [54, 39]. This relation can be represented as $\pi = \alpha\bar{\pi}$, where α is a constant. When the constraints $\sum_i \pi_i = 1$ and $\sum_i \bar{\pi}_i = 1$ are enforced, the two definitions will induce exactly the same PageRank score for each page.

6.5 Sensitivity Analysis of Markov chain

Our theoretical foundation consists of one theorem addressing the mean first passage time of Markov chain [52], two theorems about the fundamental matrix of Markov chain [52] and one theorem relating to the monotone property of Markov chain [17] . Due to the space constraint, we only present the theorem statements. Interested readers can refer to the standard references such as [52, 17, 2] for details. We fix a Markov chain of N states, of which the transition matrix is P.

Definition 6.1. The *mean first passage time from i to j*, denoted by m_{ij}, is the expected number of steps entering State j starting from State i.

Theorem 6.2. *[52] Let P be the transition matrix of a regular Markov chain. We have the following facts:*

(1). For any two states i and j, $m_{ij} = 1 + \sum_{k \neq j} p_{ik} m_{kj}$;

(2). For any state i, the stationary distribution $\pi_i = \frac{1}{m_{ii}}$;

(3). For any two states $i \neq j$, changing the transition probabilities of j to any other states does not change m_{ij}.

Definition 6.3. [52] The *fundamental matrix Z* of the transition matrix P of a Markov chain is defined as:

$$Z \stackrel{\text{def}}{=} (I - (P - B))^{-1}.$$

Here $B \stackrel{\text{def}}{=} \lim_{k \to \infty} P^k$.

Two fundamental results about the fundamental matrix are:

Theorem 6.4. *[52] The fundamental matrix of a regular Markov chain with transition matrix P always exists, and further more,*

$$Z = I + \sum_{k=1}^{\infty} (P - B)^k.$$

Theorem 6.5. *[17] Let P and $\tilde{P}$ be the transition matrices of two Markov chains and $\tilde{P} = P + \Delta$. Suppose $\tilde{\pi}$ and π are the stationary distributions of $\tilde{P}$ and P, while Z is the fundamental matrix of P. We have the following facts:*

(1). $\tilde{\pi}=\tilde{\pi}\Delta Z+\pi$;

(2). Z is diagonally dominant over columns, that is, $z_{jj} \geq z_{ij}$ for all i and j. Furthermore, for all i and j, $j \neq i$, $z_{jj} - z_{ij} = m_{ij}\pi_j$.

Chien *et al.* [17] proves the following useful monotone property of Markov chain.

Theorem 6.6. *[17] Let P be the transition matrix of a finite state regular Markov chain, and let i and j be arbitrary states of P. Let Δ be a matrix that is zero everywhere except in row i, the (i, j) entry is the only positive entry, and $\tilde{P} = P + \Delta$ is also the transition matrix of a regular Markov chain. Let $\tilde{\pi}$ denote the stationary distribution of $\tilde{P}$. Then $\tilde{\pi}_j > \pi_j$.*

6.6 Characterization of an Optimal Spam Farm

In this section, we shall give the characterization of an optimal spam farm under some realistic assumptions. First, we introduce two definitions.

Definition 6.7. A web graph is called a *realistic web graph* if

(1). The number of Web pages N is large enough such that $2c\frac{N-1}{N} > 1$;

(2). The number of dangling pages is at least 2.

In practice, c takes value from 0.8 to 0.9 [63, 12] while Google indexes around 6 billion Web pages, many of which are dangling pages. Therefore, the definition of a realistic web graph is very natural.

Definition 6.8. If a Web page only points to the target page, we call it a *generous page.*

Then we characterize the optimal spam farm as shown in Figure 6.1 in the following theorem.

Theorem 6.9. *In a realistic web graph, we assume that each hijacked page already points to a set of nongenerous pages such that at least two of them are not hijacked pages and do not point to any generous page. Then a spam farm is optimal iff*

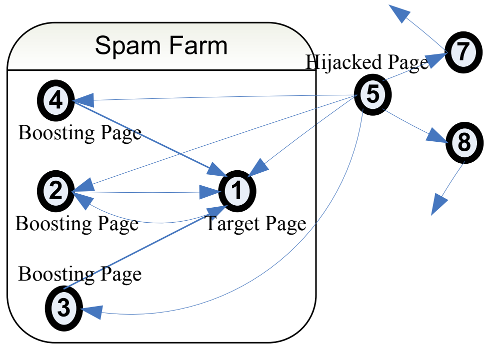

Figure 6.1: Optimal spam farm

(1). The boosting pages point to and only to the target page;

(2). The target page points to and only to some of the generous pages;

(3). The hijacked pages point to the target page and all the boosting pages.

In Theorem 6.9, the assumptions about the hijacked pages are realistic, because in the real world, the hijacked pages are most likely to be online bulletins or blogs. Those pages probably point to a number of normal Web pages that are neither generous pages nor hijacked pages. Moreover, the contents of the hijacked pages are not likely to be relevant to spam farm pages. Based on the belief that one page points to another if they are relevant, it is reasonable to assume that at least two of the Web pages, which the hijacked pages point to, do not point to generous pages.

Although we give the optimal spam farm in the above theorem, a spammer may not achieve the maximum PageRank score for the target page in practice. The reasons are two fold. First, there are enormous bulletin pages and blogs that can be used as hijacked pages and it is impossible for a spammer to add hijacked links to all of them. But, according to the proof of Theorem 6.9, in order to maximize the PageRank score of the target page, a spammer should hijack as many pages that satisfy our assumption as possible. Second, adding a link from a hijacked page to a normal page may boost the target page, too. However, according to the definition of a single-target spam farm model, the hijacked links should point to spam farm pages. Thus, we do not consider such case in our theorem.

In our proof of Theorem 6.9, we find the optimal structure by optimizing outgoing links of the target page and the boosting pages, as well as the hijacked links, step by step. This proof is totally different from the proof in [39], which solves the optimization problem as a whole. Consequently, our method will provide more insights about the effect of adding or deleting links compared to the method of Gyöngyi and Garcia-Molina. In the following proof, let t be the target page, g be a generous page, d be a dangling page, b be a boosting page and h be a hijacked page. First, we study the outgoing links of the boosting pages in the optimal spam farm.

Lemma 6.10. *In the optimal spam farm, the boosting pages should point to and only to the target page.*

Proof. First, we claim that the boosting pages should point to the target page in the optimal spam farm. We will prove this claim in two cases. The first case is that the boosting page b has nonzero out degree. When adding (b, t) to E, $\forall k \neq t$, if edge $(b, k) \in E$, p_{bk} decreases from $c\frac{1}{l} + (1-c)\frac{1}{N}$ to $c\frac{1}{l+1} + (1-c)\frac{1}{N}$, where l is the out degree of page b before adding the link (b, t); if edge $(b, k) \notin E$, p_{bk} does not change. At the same time, p_{bt} increases from $(1-c)\frac{1}{N}$ to $c\frac{1}{l+1} + (1-c)\frac{1}{N}$. The second case is that b has zero out degree. Adding the edge (b, t) to E can increase p_{bt} from $\frac{1}{N}$ to $c + (1-c)\frac{1}{N}$. $\forall k \neq t$, p_{bk} decreases from $\frac{1}{N}$ to $\frac{1-c}{N}$. According to theorem 6.6, adding the edge (b, t) can increase the PageRank score of the target page t. Therefore, the boosting pages should point to the target page in the optimal spam farm.

Next, we claim that the boosting pages cannot point to any other pages besides the target page in the optimal spam farm. We will prove this claim by contradiction. If b points to t and some other pages $k_1, k_2, ..., k_l$ other than t, we will delete all the links from b to $k_1, k_2, ..., k_l$ and see what happens. Before the deletion, $p_{bk_1}, ..., p_{bk_l}$ and p_{bt} should be $c\frac{1}{l+1} + (1-c)\frac{1}{N}$; after deletion, $p_{bk_1}, ..., p_{bk_l}$ would be $(1-c)\frac{1}{N}$ and p_{bt} would be $c + (1-c)\frac{1}{N}$. It is obvious that p_{bt} increases and $p_{bk_1}, ..., p_{bk_l}$ decreases. According to theorem 6.6, the deletion operations can increase the PageRank score of the target page t. Therefore, the boosting pages cannot point to any other pages besides the target page. Finally, putting the two claims together proves the lemma.

□

Lemma 6.10 implies that in the optimal spam farm, the boosting page should be a generous page. However, note that not every generous page is a boosting page since a normal page in the web graph may only point to the target page.

Next, we study the outgoing links of the target page in the optimal spam farm. The basic idea is that according to Theorem 6.2, if we want to maximize the PageRank score of target page t, we need to minimize the mean first passage time m_{tt}. Since the mean first passage time will play an important role in our proofs, the following two lemmas will address the mean first passage times of generous pages and dangling pages to the target page. Recall that m_{ij} stands for the mean first passage time from page i to page j, while g and t stand for the generous page and the target page, respectively.

Lemma 6.11. *For any page k, $m_{gt} \leq m_{kt}$. Moreover, $m_{gt} = m_{kt}$ iff k is a generous page.*

Proof. According to Theorem 6.2,

$$m_{gt} = 1 + \sum_{i \neq t} p_{gi} m_{it} \tag{6.1}$$

$$m_{kt} = 1 + \sum_{i \neq t} p_{ki} m_{it} \tag{6.2}$$

According to the PageRank algorithm, for any page $i \neq t$, $p_{gi} = \frac{1-c}{N} \leq p_{ki}$. It is obvious that $m_{gt} \leq m_{kt}$. Moreover, $m_{gt} = m_{kt}$ iff for any page $i \neq t$, $p_{gi} = p_{ki}$, which implies that k is a generous page too. □

Lemma 6.12. $m_{dt} - m_{gt} \geq c\frac{N-1}{N}m_{gt}$.

Proof. According to equations 6.1 and 6.2, we can get

$$m_{dt} - m_{gt} = \frac{c}{N}\sum_{k\neq t} m_{kt} \tag{6.3}$$

Since for any k, $m_{kt} \geq m_{gt}$, we can get

$$m_{dt} - m_{gt} \geq c\frac{N-1}{N}m_{gt}$$

□

Based on the above two lemmas, we can characterize the outgoing links of the target page in the optimal spam farm.

Lemma 6.13. *In a realistic web graph, the target page should point to and only to some of the generous pages in the optimal spam farm.* [3]

Proof. According to Theorem 6.2, if we want to maximize the PageRank score of the target page t, we need to minimize the mean first passage time m_{tt}. We will prove this lemma by proving the following three claims.

First, we claim that in the optimal spam farm, t has a nonzero out degree. This is because if t has zero out degree,

$$m_{tt} = 1 + \frac{1}{N}\sum_{i\neq t} m_{it}$$

[3] If self-loops are allowed, the target page should only point to itself in the optimal spam farm.

However, if t points to a generous page g,

$$\begin{aligned}\widetilde{m_{tt}} &= 1 + c \cdot m_{gt} + \frac{1-c}{N}\sum_{i \neq t} m_{it} \\ &= 1 + c(m_{gt} - \frac{1}{N}\sum_{i \neq t} m_{it}) + \frac{1}{N}\sum_{i \neq t} m_{it}\end{aligned}$$

According to the assumption of a realistic web graph, there are at least two dangling pages d_1 and d_2. Therefore,

$$N \cdot m_{gt} - \sum_{i \neq t} m_{it} = \sum_{i \neq t, d_1, d_2} (m_{gt} - m_{it}) + (3m_{gt} - m_{d_1 t} - m_{d_2 t})$$

According to Lemma 6.11, $m_{gt} \leq m_{it}$; furthermore, according to Lemma 6.12 and the assumption that $2c\frac{N-1}{N} > 1$, then $3m_{gt} < m_{d_1 t} + m_{d_2 t}$. Therefore, $N \cdot m_{gt} - \sum\limits_{i \neq t} m_{it}$ is negative. Consequently, $\widetilde{m_{tt}} < m_{tt}$. This implies that in the optimal spam farm, t cannot have zero out degree.

Next we claim that in the optimal spam, the target page cannot point to a non-generous page. Suppose the set of pages t points to is $\mathcal{K}$. Then,

$$m_{tt} = 1 + c\frac{\sum\limits_{i \in \mathcal{K}} m_{it}}{|\mathcal{K}|} + \frac{1-c}{N}\sum_{i \neq t} m_{it}$$

If t only points to generous pages,

$$\widetilde{m_{tt}} = 1 + c \cdot m_{gt} + \frac{1-c}{N}\sum_{i \neq t} m_{it}$$

According to Lemma 6.11, when $\mathcal{K}$ contains nongenerous pages, $\widetilde{m_{tt}} < m_{tt}$. This implies that in the optimal spam farm, the target page can only point to a generous page.

Finally, we claim that in the optimal spam farm, when the target page only points to a generous page, the number of generous pages the target page points to does not matter. If t points to $q \in \mathcal{N}$ generous pages,

$$m_{tt} = 1 + \frac{c \cdot q}{q} m_{gt} + \frac{1-c}{N}\sum_{i \neq t} m_{it}$$

If t points to $q+1$ generous pages,

$$\widetilde{m_{tt}} = 1 + \frac{c \cdot (q+1)}{q+1} m_{gt} + \frac{1-c}{N} \sum_{i \neq t} m_{it}$$

It is obvious that $\widetilde{m_{tt}} = m_{tt}$. Therefore, when the target page only points to generous page, the number of generous pages the target page points to does not matter.

Finally, putting the three claims together proves the lemma. $\square$

The last step to prove Theorem 6.9 is studying the hijacked links. The proof of Lemma 6.10 implies that in the optimal spam farm, the hijacked pages should point to the target page. The key question is whether the hijacked pages should point to the boosting pages besides the target page. In order to answer this question, we first prove the following lemma.

Lemma 6.14. *Suppose a hijacked page h already points to the target page t and a set of nongenerous pages $\mathcal{K}$, adding the link (h,g) where g is a generous page can boost the target page iff $\sum_{k \in \mathcal{K}} m_{kt} > (|\mathcal{K}|+1) m_{gt}$.*

Proof. Let π_t and $\widetilde{\pi_t}$ be the PageRank score of the target page before and after adding the link (h,g). According to Theorem 6.5, we can get

$$\widetilde{\pi_t} - \pi_t = \widetilde{\pi_h} \Delta_{h*} Z_{*t}$$

When adding the link (h,g), $\delta_{hg} = -\sum_{i \neq g} \delta_{hi}$ and $\forall i \neq g, \delta_{hi} \leq 0$, then

$$\begin{aligned} \Delta_{h*} Z_{*t} &= \sum_{i \neq g, (h,i) \in E} \delta_{hi} (z_{it} - z_{gt}) \\ &= \sum_{i \neq g, t, (h,i) \in E} \delta_{hi} ((z_{tt} - z_{gt}) - (z_{tt} - z_{it})) + \delta_{ht} (z_{tt} - z_{gt}) \end{aligned}$$

Since $z_{tt} - z_{it} = m_{it} \pi_t$, we can get

$$\begin{aligned} \Delta_{h*} Z_{*t} &= \pi_t \delta_{ht} (\sum_{i \neq b, t, (h,i) \in E} (m_{gt} - m_{it}) + m_{gt}) \\ &= \pi_t \delta_{ht} ((|\mathcal{K}|+1) m_{gt} - \sum_{k \in \mathcal{K}} m_{kt}) \end{aligned}$$

Because $\delta_{ht} < 0$ and $\pi_t > 0$, we know that $\widetilde{\pi}_t > \pi_t$ iff $\sum_{k \in \mathcal{K}} m_{kt} > (|\mathcal{K}| + 1)m_{gt}$. □

Lemma 6.10 and 6.14 give a necessary and sufficient condition for the optimal link structure of the hijacked pages. However, a spammer may not have any knowledge of the mean first passage time. Therefore, in the following lemma, we will address the link structure for those hijacked pages that satisfy some realistic assumptions. The statement of this lemma has nothing to do with mean first passage time.

Lemma 6.15. *In a realistic web graph, suppose a hijacked page h already points to a set of nongenerous pages $\mathcal{K}$; moreover, at least two Web pages in $\mathcal{K}$ do not point to any generous page. In the optimal spam farm, h should point to the target page and all of the boosting pages.*

Proof. Suppose $k_1, k_2 \in \mathcal{K}$ do not point to any generous page. We claim that $m_{k_1t} - m_{gt} \geq c\frac{N-1}{N}m_{gt}$(the same inequality holds for k_2). We prove this claim by considering two cases. If k_1 has zero out degree, Lemma 6.12 proves that $m_{k_1t} - m_{gt} \geq c\frac{N-1}{N}m_{gt}$. If k_1 has nonzero out degree, based on equation 6.1 and 6.2, we can get

$$\begin{aligned} m_{k_1t} - m_{gt} &= \sum_{i \neq t}(p_{k_1i} - \frac{1-c}{N})m_{it} \\ &\geq c \cdot m_{gt} \end{aligned}$$

Therefore, $m_{k_1t} - m_{gt} \geq c\frac{N-1}{N}m_{gt}$.

The proof of Lemma 6.10 implies that in the optimal spam farm, all the boosting pages are generous pages and h should point to the target page. When h already points to the target page, Lemma 6.14 tells us that adding a hijacked link from h to a generous page g can further boost the target page iff $\sum_{k \in \mathcal{K}} m_{kt} > (|\mathcal{K}| + 1)m_{gt}$. Based on our assumption, we know that

$$\begin{aligned} \sum_{k \in \mathcal{K}} m_{kt} &\geq |\mathcal{K}|m_{gt} + 2c\frac{N-1}{N}m_{gt} \\ &> (|\mathcal{K}| + 1)m_{gt} \end{aligned}$$

Therefore, h should point to all of the boosting pages. The lemma is proved. $\square$

Finally, if we summarize Lemma 6.10, 6.13 and 6.15, it would give us a unique configuration of the optimal spam farm. This will complete the proof of Theorem 6.9.

6.7 Optimal Spam Farm Under Constraints

As shown in Theorem 6.9, in the optimal spam farm, the target page only points to generous pages while the boosting pages only point to the target page. This structure is easy to detect. In order to disguise the spam farm, a spammer may require that the target page should point to some nongenerous pages, the boosting pages should not directly point to the target page, or the hijacked pages should not directly point to the target page. Hence, this question arises: what are the optimal spam farm structures under those constraints?

First, we characterize the optimal spam farm when the target page is required to point to some nongenerous pages in the following theorem.

Theorem 6.16. *If the target page t is required to point to a set of pages $\mathcal{K}$, a spam farm is optimal only if*

(1). The boosting pages point to and only to the target page;

(2). The target page points to a set of pages $\mathcal{K} \bigcup \mathcal{L}$ such that $(\sum_{k \in \mathcal{K}} m_{kt} + \sum_{l \in \mathcal{L}} m_{lt})/(|\mathcal{K}| + |\mathcal{L}|)$ is minimized, where $\mathcal{L} \subseteq V$.

Proof. The proof of this theorem directly follows from the proof of Theorem 6.9. $\square$

In order to design the optimal spam farm under the above constraint, a spammer needs to find out the set of Web pages $\mathcal{L}$ such that the average mean first passage time of Web pages in $\mathcal{K} \bigcup \mathcal{L}$ to the target page is minimized. This requires that a

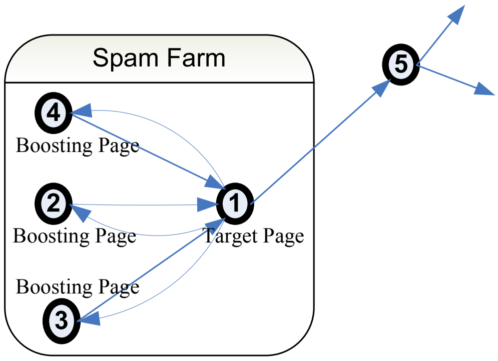

Figure 6.2: Optimal spam farm when the target page points to non generous pages

spammer has knowledge of the mean first passage time of the web graph. Given the limited computing resources of a spammer, it is a nontrivial task for him to find out the set $\mathcal{L}$. However, Theorem 6.16 implies that in the optimal spam farm, the target page should point to all the generous pages, shown in Figure 6.2.

Next, we characterize the optimal spam farm, as shown in Figure 6.3, when some of the boosting pages cannot directly point to the target page in the following theorem.

Theorem 6.17. *In a realistic web graph, suppose $\mathcal{B}$ is the set of boosting pages and a subset of it $\overline{\mathcal{B}} \subset \mathcal{B}$ cannot directly point to the target page, then a spam farm is optimal only if*

(1). For any page in $\mathcal{B} \setminus \overline{\mathcal{B}}$, it points to and only to the target page;

(2). For any page in $\overline{\mathcal{B}}$, it points to some of the generous pages;

(3). The target page points to and only to some of the generous pages.

Proof. For the pages in $\mathcal{B} \setminus \overline{\mathcal{B}}$, Lemma 6.10 implies that they should point to and only to the target page.

For the pages in $\overline{\mathcal{B}}$, we first claim that it cannot have zero out degree in the optimal spam farm. This is because if $b \in \overline{\mathcal{B}}$ has zero degree, adding the link (b, g)

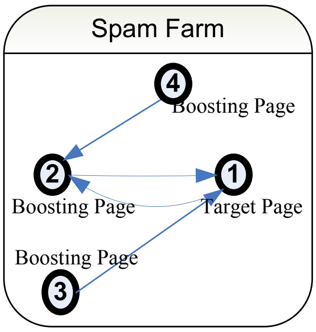

Figure 6.3: Optimal spam farm when some boosting page cannot point to the target page

where g is a generous page can boost the target page t iff

$$\sum_{k \neq t} m_{kt} - N \cdot m_{gt} > 0$$

Given the assumption of a realistic web graph, similar to the proof of the first claim in Lemma 6.13, we know that b should have nonzero out degree in the optimal spam farm.

Next, we claim that b cannot point to nongenerous pages in the optimal spam farm. This is because when b has nonzero out degree, deleting a link from b to another page j (but still keeping that b has non zero out degree) can boost the target page iff

$$\sum_{k \neq j, (b,k) \in E} (m_{kt} - m_{jt}) < 0$$

According to Lemma 6.11, for a generous page g, m_{gt} is minimum. Combined with the previous necessary and sufficient condition, this implies that b cannot point to nongenerous pages and the number of generous pages that b points to does not matter if b only points to generous pages. Consequently, b should point to some of the generous pages in the optimal spam farm.

For the target page, Lemma 6.13 implies that it should point to and only to some of the generous pages. $\square$

Note that in the above two characterizations, we ignore the hijacked links to avoid

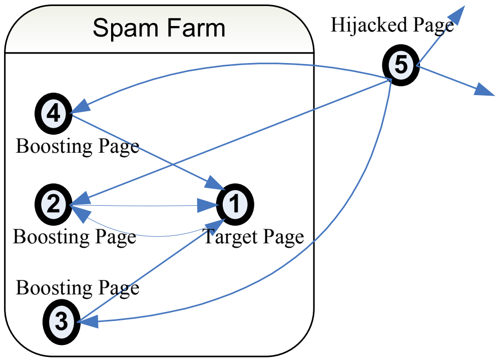

Figure 6.4: Optimal spam farm when the hijacked pages cannot point to the target page

repeatedness. This is because, according to the proof of Theorem 6.9, the structure of the hijacked links in the optimal spam is quite independent of the structure of outgoing links of the target page and the boosting pages. Therefore, when the hijacked links need to be taken into consideration, we can follow almost the same analysis as Theorem 6.9 to find out the structure of the hijacked links in the optimal spam farm.

Finally, we characterize the optimal spam farm, as shown in Figure 6.4, when the hijacked pages cannot directly point to the target page, as we see in the following theorem.

Theorem 6.18. *In a realistic web graph, suppose the hijacked pages already point to some nongenerous pages and the hijacked pages cannot directly point to the target page, then a spam farm is optimal iff*

(1). The boosting pages point to and only to the target page;

(2). The target page points to and only to some of the generous pages;

(3). The hijacked pages point to all of the generous pages.

A similar analysis as the proof of Theorem 6.17 can show correctness of Theorem 6.18. We omit the proof here.

6.8 Related Works

As a first step to detect Web spam, researchers need to identify various spamming techniques. Langville *et al.* introduced the problem of link spam analysis as future work in their comprehensive survey [54]. Bianchini *et al.* [8] studied how to design the structure of a web graph that contains exactly N pages such that a page's PageRank score is maximized. However, in practice, a spammer cannot control all Web pages. When a spammer only controls a small fraction of the web graph, the optimal link structure of the spam pages, especially links from external pages to the spam, is not addressed in [8].

Gyöngyi and Garcia-Molina [39] first introduced the single-target spam farm model. In this model, a spammer wants to boost the PageRank score of the target page by manipulating the outgoing links of the target page and a set of boosting pages. They claimed to have identified a link structure that was *optimal* in maximizing the PageRank score of a single-target page. However, we find that the proof of their paper is flawed by assuming that the PageRank score flowed into the spam farm was constant. Those who are interested can refer to [30] for a counterexample. Nevertheless, given the extremal nature of the optimal link structure, it is not surprising that their conclusion is very close to the correct answer. Moreover, the optimal spam farm structures are easy to detect [82]. The spammer can deviate from the optimal spam farm structure to disguise the spam farm. Unfortunately, this problem was not well addressed in [8, 39].

Adali *et al.* [75] studied the optimal link structure under the assumption that a spammer only has control of the boosting pages, but not the target page. Moreover, the optimality of the disguised attack depends on the forwarding value, which has a flavor of PageRank score. In order to compute forwarding value, a spammer has to solve a system of linear equations like PageRank. Considering the size of the Web, this would require the spammer to have huge computation resources. Thus, such attack strategies are not very practical. Cheng and Friedman [16] quantitatively analyzed PageRank score increase of the target page under optimal sybil attacks [8]. Although the definition of PageRank in [16] is significantly different from the standard definition by Page *et al.* [12], the result in [16] can be easily modified to match the standard case by following exactly the same proof ideas.

CHAPTER VII

Ranking via Arrow-Debreu Equilibrium

7.1 Motivations

Ranking, which aggregates the preferences of individual agents over a set of alternatives, is not only a fundamental problem in social choice theory but also has many applications in real life. For instance, the well-known PageRank algorithm [12] is designed to rank Web pages while the Invariant Method [77, 66] is proposed to evaluate the intellectual influence of academic journals and papers.

Intuitively, the PageRank and the Invariant method share a common property in that the more "vote" an agent gets, the higher ranking he has. Although they work very well in practice, the economic interpretations of their effectiveness are not obvious. Slutzki and Volji [77], as well as Palacios-Huerta and Volji [66], gave the first set of axioms that characterize the Invariant method. Later, Altman and Tennenholtz [3] gave a set of combinatorial axioms to characterize the PageRank algorithm, while Brandt and Fischer [11] interpreted PageRank as a solution of a weak tournament.

General equilibrium theory [57] is one of the most prominent theories in mathematical economics. It studies how a market system, known as the "invisible hand", makes the demands of a market's participants equal to its supplies. Arrow and Debreu [7] showed that under mild conditions, a market always has an equilibrium. The result of this research became known as the *Arrow-Debreu equilibrium*.

In this chapter, we will establish a connection between ranking methods and the Arrow-Debreu equilibrium. Naturally, the preference of one agent for another, which is usually represented as a directed edge in a graph, can be viewed as the demand between agents. Intuitively, the more demands a good gets, the higher price it should have. Therefore, an equilibrium price could be a good candidate for a ranking vector. On the other hand, the PageRank and the Invariant method are the stationary distributions of ergordic Markov chains. Both the existence of a PageRank or an Invariant ranking vector and the existence of Arrow-Debreu equilibrium can be shown via the Brouwer's fixed point theories [7, 57]. We will interpret one form of a fixed point as the other. More specifically, we will show that the ranking vector of the PageRank or the Invariant method is indeed the equilibrium of a Cobb-Douglas market. To the best of our knowledge, this is the first connection between ranking methods and the general equilibrium theory. Based on our observations, we propose a new ranking method, the CES ranking, which is minimally fair, strictly monotone, and invariant to reference intensity, but not uniform or weakly additive.

7.2 Preliminaries

7.2.1 The Ranking Problem

In this subsection, we will follow [3] to define *ranking problems*. Let A be a finite set, representing the set of agents, and M be a $|A| \times |A|$ matrix, representing the preference relationships among the agents. A ranking problem is represented as $\langle A, M\rangle$. For any $n \in \mathbb{N}$, let $\Delta_n = \{(x_1, ..., x_n) | \forall i, x_i \geq 0 \text{ and } \sum_i x_i = 1\}$.

Definition 7.1. A ranking function maps a ranking problem $\langle A, M\rangle$ to a vector

$\pi \in \Delta_{|A|}$.

Invariant Method [77] In the definition of PageRank [12], if the transition matrix T is irreducible (the corresponding graph is strongly connected), its unique stationary distribution is the ranking vector. Thus, the PageRank and the Invariant method are essentially equivalent in mathematics.

7.2.2 Arrow-Debreu equilibrium of exchange markets

In an exchange market, there are m traders and n divisible goods. Let $\mathcal{T} = \{T_1, .., T_m\}$ be the set of traders and $\mathcal{G} = \{G_1, ..., G_n\}$ be the set of goods. Each trader i has an initial endowment of $w_{i,j} \geq 0$ of good j and a utility function $u_i : \mathbb{R}_+^n \to \mathbb{R}$. The individual goal of trader T_i is to obtain a new bundle of goods that maximizes his utility. Let $\mathbf{x}_i \in \mathbb{R}_+^n$ be the bundle of goods of T_i after the exchange, where $x_{i,j}$ is the amount of good j. Naturally, the demand cannot exceed the supply: $\sum_i x_{i,j} \leq \sum_i w_{i,j}$, for every good j.

We use $\mathbf{p} \in \Delta_n$ to denote a price vector, where p_j is the price of G_j. For any trader T_i, given $\mathbf{p}$, we let $\mathbf{x}_i^*(\mathbf{p})$ denote the bundle of goods that maximize his utility under the budget constraint:

$$\mathbf{x}_i^*(\mathbf{p}) = \operatorname{argmax}_{\mathbf{x} \in \mathbb{R}_+^n,\ \mathbf{x} \cdot \mathbf{p} \leq \mathbf{w}_i \cdot \mathbf{p}} u_i(\mathbf{x}).$$

Definition 7.2 (Arrow-Debreu equilibrium). A market equilibrium is a price vector $\mathbf{p} \in \Delta_n$ such that the market clears:

> For every good $G_j \in \mathcal{G}$, $\sum_{i \in [m]} x_{i,j}^*(\mathbf{p}) \leq \sum_{i \in [m]} w_{i,j}$; If $p_j > 0$, then $\sum_{i \in [m]} x_{i,j}^*(\mathbf{p}) = \sum_{i \in [m]} w_{i,j}$.

7.2.3 CES Utility Functions

CES utility functions: [19, 57] The CES (Constant Elasticity of Substitution) function over a bundle of goods $(x_{i1}, ..., x_{in})$ is the family of utility functions $u_i(x_{i1}, ..., x_{in}) = (\sum_{j=1}^{n} \alpha_{ij} x_{ij}^{\rho_i})^{1/\rho_i}$, where $-\infty < \rho_i < 1$, $\rho_i \neq 0$ and $\alpha_{ij} \geq 0$. The parameter $1/(1-\rho_i)$ is called the *elasticity of substitution*. The CES utility function has a very nice property: its demand functions have explicit forms. That is, given a strictly positive price vector $\pi \in \mathbb{R}^n_{++}$, the demand x_{ij} is

$$x_{ij} = \frac{\alpha_{ij}^{1/(1-\rho_i)}}{\pi_j^{1/(1-\rho_i)}} \times \frac{\sum_k \pi_k w_{ik}}{\sum_k \alpha_{ik}^{1/(1-\rho_i)} \pi_k^{-\rho_i/(1-\rho_i)}}$$

There are three important utility functions within the CES category.

(1). $\rho_i \to 1$ corresponds to linear utility functions, where $u_i(x_{i1}, ..., x_{in}) = \sum_j \alpha_{ij} x_{ij}$. In this case, the set of goods that the agent wants are perfect substitutes for each other.

(2). $\rho_i \to -\infty$ corresponds to Leontief utility functions. The Leontief utility function, in general, has the form of $u_i(x_{i1}, ..., x_{in}) = \min_{j:\beta_{ij}>0} \frac{x_{ij}}{\beta_{ij}}$, where $\beta_{ij} \geq 0$. In this case, the set of goods that the agent wants are perfect complement of each other.

(3). $\rho_i \to 0$ corresponds to Cobb-Douglas utility functions. The Cobb-Douglas utility function, in general, has the form of $u_i(x_{i1}, ..., x_{in}) = \prod_j x_{ij}^{\beta_{ij}}$, where $\beta_{ij} \geq 0$. This demand function is a perfect balance of substitution and complementarity effects [19].

7.3 PageRank/Invariant Method V.S. a Cobb-Douglas Market

In this section, we will establish a connection between PageRank/Invariant Method and Arrow-Debreu equilibrium. We do this by showing a more general theorem about Markov chains.

Theorem 7.3. *Given an ergordic Markov chain, there is a mapping from the Markov chain to a Cobb-Douglas market, such that the stationary distribution of the Markov chain is precisely the Arrow-Debreu equilibrium of the Cobb-Douglas market.*

Proof. The general idea of the proof is to reduce the state transition graph of an ergordic Markov chain to the economy graph of a Cobb-Douglas market. Given the state transition graph $G = (V, E, W)$ and the transition matrix P, we can construct a Cobb-Douglas economy graph as follows: for $i \in [1..n]$, there is a trader T_i corresponding to each state s_i, and there is a directed link from T_i to T_j iff $(s_i, s_j) \in E$; for trader T_i, let $N(T_i)$ be the set of outgoing neighbors of T_i. The utility function of T_i is $u_i(x_i) = \prod_{j \in N(T_i)} x_{ij}^{p_{ij}}$, where p_{ij} is the transition probability. Initially T_i has one unit of the good G_i but no other goods. We call such a Cobb-Douglas economy M. We claim that the market equilibrium of M is also the stationary distribution of P.

First of all, since G is strongly connected and the Cobb-Douglas utility function belongs to the CES utility function class, according to Theorem 1 of Codenotti *et al.* [19] M has a strictly positive equilibrium. By the demand function of CES utility function, when $\rho_i \to 0$,

$$x_{ij} = \frac{p_{ij}\pi_i}{\pi_j}.$$

By the definition of Arrow-Debreu equilibrium, for every good with strictly positive price, its demand must be equal to its supply. Thus,

$$\sum_i x_{ij} = \sum_i \frac{p_{ij}\pi_i}{\pi_j} = 1.$$

Equivalently,

$$\sum_i p_{ij}\pi_i = \pi_j.$$

Thus, π is the stationary distribution of the Markov chain P. $\square$

Actually, by the above reduction, we also implicitly show that M has a unique equilibrium. Most importantly, since the PageRank or Invariant method is a special Markov chain, the ranking vector of the PageRank or the Invariant method can also be interpreted as the equilibrium of a Cobb-Douglas market.

Economic Interpretations: It is believed that the validity of PageRank comes from the fact that the Markov chain is a good model for the Web surfing behavior of Web users. In web graph, a link from page p to page q means that a Web user at page p may find the content of page q is useful. Thus, a link in web graph means a vote or reference. Intuitively, the more votes a page gets, the more important it is. Indeed, for a Web user, his goal is to maximize his information needs by following outgoing links of a page to visit other pages. Thus, in our Cobb-Douglas economy graph, each Web page is corresponding to an agent, the content of the page is corresponding to the good the agent initially owns, and a link from p to q means that the agent on page p has a demand for the content of page q. Intuitively, the more "demand" a page gets, the more important it is.

Theorem 7.3 provides a new perspective to view PageRank. That is the substitution and complementarity effects of outgoing links. For instance, suppose we have a directory page of a university, which has outgoing links pointing to the home pages of each of the unversity's the departments. If a Web user clicks one of the outgoing links, it is unlikely that he will click any other. Thus, for this page, its outgoing links are more likely to be substitutes for each other than complements. On the other hand, suppose we have a news page, which has outgoing links pointing to related news pages. A Web user who clicks one of the outgoing links is likely to click another one. Thus, for this page, its outgoing links are more likely to be complements for each other than substitutes. By Theorem 7.3 and the properties of the Cobb-Douglas utility function, the set of pages that a Web page points to is a mix of the substitution and complementarity effects with elasticity of substitution 1 in PageRank.

7.4 Ranking via Arrow-Debreu Equilibrium

The Cobb-Douglas utility function corresponds to the CES utility function with $\rho \to 0$. Thus, by choosing CES utility functions with different elasticities, we can naturally extend the idea of PageRank to a new spectrum of ranking algorithms. We propose the ranking method, which is called *CES ranking*, below.

Algorithm 1 CES ranking

1. Given the agents set A, choose a CES utility function u_i for each agent $i \in A$ and set the initial endowment w_i of it as $\forall j \neq i, w_{ij} = 0$ but $w_{ii} = 1$. The new ranking problem is $\langle A, \{\alpha_{ij}\}_{i \in A, j \in A}, \{\rho_i\}_{i \in A}, \{w_i | i \in A\}\rangle$. W.L.O.G., for any i, let $\sum_j \alpha_{ij} = 1$.
2. If for agent i, $\alpha_{ij} = 0$ for all j, set $\alpha_{ij} = 1/|A|$ for each j. Hence, $u_i = (\sum_{j=1}^{n}(1/|A|)x_{ij}^{\rho_i})^{1/\rho_i}$.
3. For each agent i, update α_{ij} to be $\alpha_{ij} * \beta + (1/|A|) * (1-\beta)$ for every j, where $\beta = 0.85$. Correspondingly, the updated utility function is $u_i = (\sum_{j=1}^{n}(\alpha_{ij} * \beta + (1/|A|) * (1-\beta))x_{ij}^{\rho_i})^{1/\rho_i}$
4. Construct the economy graph G with respect to the CES economy defined above. It is easy to see that G is strongly connected.
5. Compute an equilibrium of G and use it as the ranking vector.

Note that the new CES ranking problem $\langle A, \{\alpha_{ij}\}_{i \in A, j \in A}, \{\rho_i\}_{i \in A}, \{w_i | i \in A\}\rangle$ is a generalization of the ranking problem $\langle A, M\rangle$ defined in the Preliminaries. Now we discuss the existence, uniqueness, and efficiency of the CES ranking, as well as some other properties related to ranking.

Existence of a ranking vector: By Theorem 1 in [19], as long as the economy graph G is strongly connected, there is always a strictly positive equilibrium. By the definition of the CES ranking, it is obvious that the economy graph of it is strongly connected. Thus, a ranking vector always exists.

Uniqueness: According to [19], for CES utility functions with $-1 \leq \rho < 1$, the set of equilibria is convex. We further show that:

claim 7.4. *The CES ranking has a unique ranking vector if* $\forall i, \rho_i \geq 0$.

In order to prove this claim, we first introduce the definition of gross substitute (GS).

Definition 7.5. [57] For any j, let $z_j = \sum_i x_{ij} - \sum_i w_{ij}$ be the excess demand function for G_j. The function $z(.)$ has the gross substitute property if whenever π' and π are such that, for some l, $\pi'_l > \pi_l$ and $\pi'_j = \pi_j$ for $j \neq l$, we have $z_j(\pi') > z_j(\pi)$ for $j \neq l$.

If in the above definition the inequalities are weak, the property is referred to as weak gross substitute (WGS). It is well known that:

Theorem 7.6. *[57] If the aggregate excess demand function of an exchange economy has the GS property, the economy has at most one equilibrium.*

Now we can prove the above claim.

Proof. [1] In the CES ranking, by the Theorem 1 and Lemma 1 in [19], as long as the economy graph G is strongly connected, an equilibrium exists and every equilibrium is strictly positive. Suppose we have two equilibria π' and π such that for l, $\pi'_l > \pi_l$ and $\pi'_h = \pi_h$ for $h \neq l$. Note that $\forall i, j, \alpha_{ij} > 0$. Thus, for any $j \neq l$,

$$\begin{aligned}
\sum_i x_{ij} &= \sum_i \frac{\alpha_{ij}^{1/(1-\rho_i)}}{\pi_j^{1/(1-\rho_i)}} \times \frac{\pi_i}{\sum_k \alpha_{ik}^{1/(1-\rho_i)} \pi_k^{-\rho_i/(1-\rho_i)}} \\
&= \sum_i \alpha_{ij}^{1/(1-\rho_i)} \times \frac{\pi_i/\pi_j}{\sum_k \alpha_{ik}^{1/(1-\rho_i)} (\pi_k/\pi_j)^{-\rho_i/(1-\rho_i)}} \\
&< \frac{\alpha_{lj}^{1/(1-\rho_l)} \times \pi'_l/\pi'_j}{\sum_k \alpha_{lk}^{1/(1-\rho_l)} (\pi'_k/\pi'_j)^{-\rho_l/(1-\rho_l)}} + \sum_{i \neq l} \frac{\alpha_{ij}^{1/(1-\rho_i)} \times \pi'_i/\pi'_j}{\sum_k \alpha_{ik}^{1/(1-\rho_i)} (\pi'_k/\pi'_j)^{-\rho_i/(1-\rho_i)}} \\
&= \sum_i x'_{ij}
\end{aligned}$$

[1] It is well known [19] that the CES utility functions satisfy WGS when $\rho \geq 0$. However, WGS does not imply the uniqueness of equilibrium.

Thus, in the CES ranking, the excess demand function has the GS property. Therefore, its equilibrium and the ranking vector are unique. □

However, when $\rho < -1$, there may be multiple equilibria sets. Actually, for ranking problems, we do not have to insist on the uniqueness of ranking vectors. That is because in most cases, there are different ranking criteria for one ranking problem. It is not surprising that different criteria induce different rankings.

Efficiency of Computation: When $-1 \leq \rho < 1$, an equilibrium of a CES market can be computed via convex programming [19]. Thus, it is in polynomial time. However, for some special utility functions (such as the Leontief utilty function), it is PPAD-hard [21] to compute an equilibrium of it. However, for ranking problems that have relatively small sizes or do not require the real time computation of ranking vectors, the efficiency may not be a serious concern.

In the next section, we study five natural properties, which are satisfied by the PageRank and the Invariant method, with respect to the CES ranking. First, we extend the concept of minimally fair [4] to the CES ranking.

Definition 7.7. A ranking system is minimally fair if when for any i, j, $\alpha_{ij} = 0$, the ranking score of agent i equals that of j for agent $i, j \in A$.

claim 7.8. *The CES ranking is minimally fair.*

Proof. Since initially $\alpha_{ij} = 0$ for any i, j, in order to make the economy graph strongly connected, we set the utility function for each agent as $u_i = ((\sum_j \frac{1}{n} x_{ij}^{\rho})^{1/\rho}$. With this setup, by the market clearing condition, we get

$$\sum_i \frac{(1/n)^{1/(1-\rho)}}{\pi_j^{1/(1-\rho)}} \times \frac{\pi_i}{\sum_k (1/n)^{1/(1-\rho)} \pi_k^{-\rho/(1-\rho)}} = 1$$

Thus, $\forall j$,

$$\pi_j = (1/(\sum_k \pi_k^{-\rho/(1-\rho)}))^{1-\rho}.$$

Note that the right-hand size of the above equation is independent of j. Thus, $\pi = (1/n, ..., 1/n)$ is the only equilibrium for the market. Therefore, the CES ranking is minimally fair. □

Next, we extend the strictly monotone definition in [4] to the CES ranking.

Definition 7.9. A ranking system is strictly monotone iff for any two agents i and j, if for any other agent p, $\alpha_{pi} \leq \alpha_{pj}$ and there exists h such that $\alpha_{hi} < \alpha_{hj}$, the ranking score of agent i is strictly less than that of j.

claim 7.10. *The CES ranking is strictly monotone when the utility functions of all the agents have the same elasticity of substitution.*

Proof. By the market clearing condition, for any two agents i and j,

$$\begin{aligned} \pi_i^{1/(1-\rho)} &= \sum_p \frac{\alpha_{pi}^{1/(1-\rho)} \times \pi_p}{\sum_k \alpha_{pk}^{1/(1-\rho)} \pi_k^{-\rho/(1-\rho)}} \\ &< \sum_p \frac{\alpha_{pj}^{1/(1-\rho)} \times \pi_p}{\sum_k \alpha_{pk}^{1/(1-\rho)} \pi_k^{-\rho/(1-\rho)}} \\ &= \pi_j^{1/(1-\rho)} \end{aligned}$$

Thus, $\pi_i < \pi_j$. □

Actually, the property of strictly monotone corresponds to the intuition that the more demands a good gets, the higher price it is.

Slutzki and Volij showed [77] that,

Theorem 7.11. *[77] If a ranking system satisfies* uniform, weakly additive *and* invariant to reference intensity, *the ranking system must be the Invariant method.*

In the next section, we will study the relationship between the CES ranking and the three properties above.

Definition 7.12. [77] A ranking problem is regular if $\forall i, j, \sum_k \alpha_{ik} = \sum_k \alpha_{jk}$ while $\forall i, j, \sum_k \alpha_{ki} = \sum_k \alpha_{kj}$. A ranking function is uniform if for every regular ranking problem, the ranking score of each agent is $\frac{1}{N}$ where N is the number of agents.

claim 7.13. *The CES ranking is not uniform.*

Proof. Suppose a system has three agents while the parameters of the agents are $\rho_1 = \rho_2 = \rho_3 = \frac{1}{2}$ and $\alpha_{11} = \frac{1}{3}$, $\alpha_{12} = \frac{1}{3}$, $\alpha_{13} = \frac{1}{3}$, $\alpha_{21} = \frac{5}{12}$, $\alpha_{22} = \frac{1}{6}$, $\alpha_{23} = \frac{5}{12}$, $\alpha_{31} = \frac{1}{4}$, $\alpha_{32} = \frac{1}{2}$, $\alpha_{33} = \frac{1}{4}$. By the market clearing condition,

$$\pi_1^2 = \frac{\frac{1}{3}^2 \times \pi_1}{\frac{1}{3}^2\pi_1^{-1} + \frac{1}{3}^2\pi_2^{-1} + \frac{1}{3}^2\pi_3^{-1}} + \frac{\frac{5}{12}^2 \times \pi_2}{\frac{5}{12}^2\pi_1^{-1} + \frac{1}{6}^2\pi_2^{-1} + \frac{5}{12}^2\pi_3^{-1}} + \frac{\frac{1}{4}^2 \times \pi_3}{\frac{1}{4}^2\pi_1^{-1} + \frac{1}{2}^2\pi_2^{-1} + \frac{1}{4}^2\pi_3^{-1}}$$

It is easy to check that $\pi = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ cannot satisfy the above equation. Thus, the CES ranking is not uniform. $\square$

Actually, the uniform property requires that the ranking score of an agent is linearly proportional to the number of "votes" it gets. This assumption may not be reasonable universally. For instance, in the citation analysis, suppose both paper A and B have m citations. In an extreme case, the citations of paper A may only come from one research group while the citations of B come from different research groups. Intuitively, paper B should be more important than paper A. However, any ranking algorithm with the uniform property cannot distinguish those two cases. The CES ranking, as a nonlinear ranking method, may have the potential to find out new signals that were missed by the uniform ranking methods.

The *weakly additive* [77] property says that for a regular ranking problem, the ranking score is still linearly proportional to the "votes" after a symmetric perturbation. Since the CES ranking is not uniform, it cannot satisfy the weakly additive property, either.

Definition 7.14. [77] A ranking system is invariant to reference intensity if for any agent i, when we multiply α_{ij} by a positive constant λ for every j, it cannot change the ranking score of any agent.

claim 7.15. *The CES ranking is* invariant to reference intensity*.*

Proof. Note that

$$\begin{aligned} x_{ij} &= \frac{\alpha_{ij}^{1/(1-\rho_i)}}{\pi_j^{1/(1-\rho_i)}} \times \frac{\sum_k \pi_k w_{ik}}{\sum_k \alpha_{ik}^{1/(1-\rho_i)} \pi_k^{-\rho_i/(1-\rho_i)}} \\ &= \frac{(\lambda\alpha_{ij})^{1/(1-\rho_i)}}{\pi_j^{1/(1-\rho_i)}} \times \frac{\sum_k \pi_k w_{ik}}{\sum_k (\lambda\alpha_{ik})^{1/(1-\rho_i)} \pi_k^{-\rho_i/(1-\rho_i)}} \\ &= \frac{\alpha_{ij}^{1/(1-\rho_i)}}{\pi_j^{1/(1-\rho_i)}} \times \frac{\sum_k \pi_k w_{ik}}{\sum_k \alpha_{ik}^{1/(1-\rho_i)} \pi_k^{-\rho_i/(1-\rho_i)}} \end{aligned}$$

Thus, multiplying α_{ij} by a positive constant λ for every j cannot change the demand function. Therefore, the set of equilibria remains the same. □

When we summarize the above claims together, we get

Theorem 7.16. *The CES ranking is* minimally fair*,* strictly monotone *and* invariant to reference intensity*, but not* uniform *or* weakly additive*.*

7.5 Conclusion and Future Works

In this chapter, we have established a connection between the ranking theory and the general equilibrium theory. First, we showed that the ranking vector of PageRank or Invariant method is actually the equilibrium of a Cobb-Douglas market. This gives a natural economic interpretation for the PageRank and Invariant method. Furthermore, we propose a new ranking method, the CES ranking, which is minimally fair, strictly monotone, and invariant to reference intensity, but not uniform or weakly additive. The new CES ranking, compared to PageRank and the Invariant method, is nonlinear, and could be potentially used to find signals in a system missed by those existing ranking methods.

With the observations in this chapter, we have a complete picture of the encoding power of the three limiting cases of CES utility functions. Pennock and Wellman [69] showed that economies with almost the linear utility functions can encode Bayesian networks. Codenotti *et al.* [21] proved that economies with the Leontief utility functions can encode bimatrix games. Now we demonstrate that economies with the Cobb-Douglas utility functions can encode Markov chains.

We believe that this chapter points to a few promising directions that are worth further exploration.

- Explore more properties that the CES ranking satisfies and make justifications for the properties it does not satisfy.
- For various applications, what is the "right" utility function for each agent? We may go beyond the CES utility functions and explore other ones, such as WGS utility functions [19].
- Design efficient algorithms to compute ranking vectors.
- Further investigate the uniqueness of ranking vectors. If there are multiple equilibria points, do they induce the same ranking? If not, interpret their different economic meanings in the context of ranking.
- Last but not least, design an effective evaluation system for ranking methods and find an application where the CES ranking can outperform existing ranking methods.

CHAPTER VIII

Conclusion

In this dissertation, we have shown that even for an approximate solution or for very simple economies, it is PPAD-Complete to compute a competitive equilibrium. Furthermore, it is NP-Complete to decide degeneracy in bimatrix games. Combined with previous results [14, 15, 22], we can conclude that it is generally hard to compute an economic equilibrium or to decide whether a game or an economy satisfies some natural properties. Thus, the difficulty of computing economic equilibria is a real challenge for making economic theories meet practice. In order to encompass this, one could further investigate other solution concepts or the average case computational complexity. One of the main open problems is the design of a polynomial time approximation scheme (PTAS) for both competitive equilibria and Nash equilibria. In general, the technical challenge to solve this open problem comes from both the nonconvexity of competitive equilibria of an economy and Nash equilibria of a game. Therefore, we expect some novel techniques to be developed before this problem is solved.

With the advent of the Internet and e-commerce applications such as eBay, an agent can easily have access to multiple accounts. He may further improve his own payoff by manipulating his multiple identities. We have studied the path auction games with multiple edge ownership. As we have shown, the condition of multiple edges ownership eliminates the possibility of reasonable solution concepts, such as a strategyproof or false-name-proof mechanism or Pareto efficient Nash equilibria. It would be worthwhile to further investigate the impact of multiple identities on mechanism design.

We investigated PageRank, part of the ranking algorithm of Google. From the perspective of game theory, we have analyzed the optimal manipulation strategies of a Web spammer against PageRank. Our understandings of link spamming techniques against PageRank aids the development of anti-spamming schemes. Furthermore, we have made a connection between the stationary distribution of a Markov chain and the equilibrium of a Cobb-Douglas market. This deepens the understandings of the economic foundations of ranking algorithms such as PageRank and the Invariant method. Finally, we propose the CES ranking method based on the Constant Elasticity of Substitution (CES) utility functions. The CES ranking method is actually a spectrum of ranking algorithms. By violating some linear properties satisfied by PageRank and the Invariant method, the CES ranking could be a useful complement of existing ranking algorithms. In the end, we hope that the connection between Markov chains and Cobb-Douglas markets helps improve the understanding of other applications of Markov chains besides PageRank.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] T. S. A. Benczr, K. Csalogny and M. Uher. Spamrank - fully automatic link spam detection. In *Proc. Int'l Workshop Adversarial Information Retrieval on the Web*, 2005.

[2] D. Aldous and J. A. Fill. Reversible Markov Chains and Random Walks on Graphs. www.stat.berkeley.edu/ aldous/RWG/book.html.

[3] A. Altman and M. Tennenholtz. Ranking systems: The PageRank axioms. In *In EC 05: Proceedings of the 6th ACM conference on Electronic commerce*, pages 1–8. ACM Press, 2005.

[4] A. Altman and M. Tennenholtz. Incentive compatible ranking systems. In *AAMAS '07: Proceedings of the 6th International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 1–8, New York, NY, USA, 2007. ACM.

[5] A. Archer and E. Tardos. Truthful mechanisms for one-parameter agents. *In Proceeding of Symposium on Foundations of Computer Science*, pages 482–491, 2001.

[6] A. Archer and E. Tardos. Frugal path mechanism. *In Proceedings of the 2002 Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 991–999, 2002.

[7] K. J. Arrow and G. Debreu. Existence of an equilibrium for a competitive economy. *Econometrica*, 22:265–290, 1954.

[8] M. Bianchini, M. Gori, and F. Scarselli. Inside PageRank. *ACM Trans. Inter. Tech.*, 5(1):92–128, 2005.

[9] L. Blum, F. Cucker, M. Shub, and S. Smale. Complexity and Real Computation. Springer-Verlag, 1998.

[10] S. Boyd and L. Vandenberghe. Convex Optimization. Cambridge University Press, 2006.

[11] F. Brandt and F. A. Fischer. PageRank as a weak tournament solution. In *Internet and Network Economics, 3rd International Workshop*, pages 300–305, 2007.

[12] S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. In *WWW7: Proceedings of the seventh international conference on World Wide Web 7*, pages 107–117, Brisbane, Australia, 1998.

[13] X. Chen, D. Dai, Y. Du, and S. Teng. Settling the complexity of Arrow-Debreu equilibria in markets with additively separable utilities. In *FOCS '09: Proceedings of the 50th Annual IEEE Symposium on Foundations of Computer Science (FOCS'09)*, 2009.

[14] X. Chen and X. Deng. Settling the complexity of two-player Nash equilibrium. In *FOCS '06: Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS'06)*, pages 261–272, Washington, DC, USA, 2006. IEEE Computer Society.

[15] X. Chen, X. Deng, and S.-H. Teng. Computing Nash equilibria: Approximation and smoothed complexity. In *FOCS '06: Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS'06)*, pages 603–612, Washington, DC, USA, 2006. IEEE Computer Society.

[16] A. Cheng and E. Friedman. Manipulability of PageRank under sybil strategies. 2006. First Workshop on the Economics of Networked Systems.

[17] S. Chien, C. Dwork, R. Kumar, D. R. Simon, and D. Sivakumar. Link evolution: Analysis and algorithms. *Internet Mathematics*, 1(3):277–304, 2003.

[18] E. Clarke. Multipart pricing of public goods. *Public Choice*, 11:17–33, 1971.

[19] B. Codenotti, B. McCune, S. Penumatcha, and K. R. Varadarajan. Market equilibrium for CES exchange economies: Existence, multiplicity, and computation. In *Foundations of Software Technology and Theoretical Computer Science*, pages 505–516, 2005.

[20] B. Codenotti, S. V. Pemmaraju, and K. R. Varadarajan. On the polynomial time computation of equilibria for certain exchange economies. In *SODA*, pages 72–81, 2005.

[21] B. Codenotti, A. Saberi, K. Varadarajan, and Y. Ye. Leontief economies encode nonzero sum two-player games. In *SODA '06: Proceedings of the seventeenth annual ACM-SIAM symposium on Discrete algorithm*, pages 659–667, New York, NY, USA, 2006. ACM.

[22] V. Conitzer and T. Sandholm. Complexity results about Nash equilibria. In *Proceedings of the 18th International Joint Conference on Artificial Intelligence*, pages 765–771, 2003.

[23] C. Daskalakis, P. W. Goldberg, and C. H. Papadimitriou. The complexity of computing a Nash equilibrium. In *STOC '06: Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pages 71–78, New York, NY, USA, 2006. ACM.

[24] X. Deng and Y. Du. The computation of approximate competitive equilibrium is PPAD-hard. In *Inf. Process. Lett.*, volume 108, pages 369–373, Amsterdam, The Netherlands, 2008. Elsevier North-Holland, Inc.

[25] X. Deng, C. Papadimitriou, and S. Safra. On the complexity of equilibria. In *STOC '02: Proceedings of the thiry-fourth annual ACM symposium on Theory of computing*, pages 67–71, New York, NY, USA, 2002. ACM.

[26] N. R. Devanur, C. H. Papadimitriou, A. Saberi, and V. V. Vazirani. Market equilibrium via a primal-dual-type algorithm. In *FOCS '02: Proceedings of the 43rd Symposium on Foundations of Computer Science*, pages 389–395, Washington, DC, USA, 2002. IEEE Computer Society.

[27] Y. Du. On the complexity of deciding degeneracy in games. http://arxiv.org/abs/0905.3012, 2009.

[28] Y. Du. Ranking via Arrow-Debreu equilibrium. under review, 2009.

[29] Y. Du, R. Sami, and Y. Shi. Path auction games when an agent can own multiple edges. In *First Workshop on the Economics of Networked Systems*, 2006.

[30] Y. Du, Y. Shi, and X. Zhao. Using spam farm to boost PageRank. In *AIRWeb '07: Proceedings of the 3rd international workshop on Adversarial information retrieval on the web*, pages 29–36, New York, NY, USA, 2007. ACM.

[31] B. C. Eaves. Homotopies for computation of fixed points. *Mathematical Programming*, 3(1):1–22, 1972.

[32] B. C. Eaves and R. Saigal. Homotopies for computation of fixed points on unbounded regions. *Mathematical Programming*, 3(1):225–237, 1972.

[33] E. Elkind. True costs of cheap labor are hard to measure: Edge deletion and VCG payments in graphs. *In Proceeding of 7th ACM conference on Electronic Commerce*, 2005.

[34] E. Elkind, A. Sahai, and K. Steiglitz. Frugality in path auctions. *In Proceeding of 15th ACM Symposium on Discrete Algorithm*, pages 701– 709, 2004.

[35] K. Etessami and M. Yannakakis. On the complexity of Nash equilibria and other fixed points (extended abstract). In *FOCS '07: Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07)*, pages 113–123, Washington, DC, USA, 2007. IEEE Computer Society.

[36] D. Fetterly, M. Manasse, and M. Najork. Spam, damn spam, and statistics: using statistical analysis to locate spam web pages. In *WebDB '04: Proceedings of the 7th International Workshop on the Web and Databases*, pages 1–6, New York, NY, USA, 2004. ACM Press.

[37] A. Fuente. Mathematical Methods and Models for Economists. Cambridge Press, 2000.

[38] T. Groves. Incentives in teams. *Econometrica*, 41:617–663, 1973.

[39] Z. Gyöngyi and H. Garcia-Molina. Link spam alliances. In *VLDB '05: Proceedings of the 31st International Conference on Very Large Data Bases*, pages 517–528. VLDB, 2005.

[40] Z. Gyöngyi and H. Garcia-Molina. Spam: It's not just for inboxes anymore. *IEEE Computer Magazine*, 38(10):28–34, October 2005.

[41] Z. Gyöngyi and H. Garcia-Molina. Web spam taxonomy. In *First International Workshop on Adversarial Information Retrieval on the Web*, 2005.

[42] Z. Gyöngyi, H. Garcia-Molina, and J. Pedersen. Combating web spam with TrustRank. In *Proceedings of the 30th International Conference on Very Large Databases*, pages 576–587. Morgan Kaufmann, 2004.

[43] M. R. Henzinger, R. Motwani, and C. Silverstein. Challenges in web search engines. *SIGIR Forum*, 36(2):11–22, 2002.

[44] M. W. Hirsch. A proof of the non-retractability of a cell onto its boundary. *Proc. Amer. Math. Soc.*, 14:364–365, 1963.

[45] L.-S. Huang and S.-H. Teng. On the approximation and smoothed complexity of Leontief market equilibria. 2006.

[46] N. Immorlica, D. Karger, E. Nikolova, and R. Sami. First-price path auctions. *In Proceeding of 7th ACM conference on Electronic Commerce*, pages 203–212, 2005.

[47] A. Iwasaki, D. Kempe, Y. Saito, M. Salek, and M. Yokoo. False-name-proof mechanisms for hiring a team. In *Internet and Network Economics, 3rd International Workshop*, pages 245–256, 2007.

[48] K. Jain. A polynomial time algorithm for computing the Arrow-Debreu market equilibrium for linear utilities. In *FOCS '04: Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science (FOCS'04)*, pages 286–294, Washington, DC, USA, 2004. IEEE Computer Society.

[49] G. Jehle and P. Reny. Advanced Microeconomic Theory. Pearson Education, 2003.

[50] A. R. Karlin and D. Kempe. Beyond VCG: Frugality of truthful mechanisms. *In Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science*, pages 615–626, 2005.

[51] R. Kellogg, T. Li, and J. Yorke. A method of continuation for calculating a Brouwer fixed point. *In: S. Karamardian, Editor, Fixed Points: Algorithms and Applications*, pages 133–147, 1977.

[52] J. G. Kemeny and J. L. Snell. Finite Markov Chains, 1960. D. Van Nostrand Company.

[53] V. Krishna. Auction Theory. *Academic Press*, 2002.

[54] A. Langville and C. Meyer. Deeper inside PageRank. *Internet Mathematics*, 1(3):335–380, 2005.

[55] C. E. Lemke and J. J. T. Howson. Equilibrium points of bimatrix games. In *Journal of the Society for Industrial and Applied Mathematics*, volume 12, pages 413–423, 1964.

[56] R. J. Lipton and E. Markakis. Nash equilibria via polynomial equations. In *LATIN*, pages 413–422, 2004.

[57] A. Mas-Colell, M. D. Whinston, and J. R. Green. Microeconomic Theory. In *Oxford Press*, 1995.

[58] E. Maskin. Nash equilibrium and mechanism design. manuscript, 2008.

[59] R. R. Maxfield. General equilibrium and the theory of directed graphs. *Journal of Mathematical Economics*, 27(1):23–51, February 1997.

[60] H. Moulin. On scheduling fees to prevent merging, splitting, and transferring of jobs. *Math. Oper. Res.*, 32(2):266–283, 2007.

[61] K. G. Murty and S. N. Kabadi. Some NP-complete problems in linear programming. *Operations Research Letters*, 1(3):101–104, 1982.

[62] J. Nash. Equilibrium points in n-person games. In *Proceedings of the National Academy of Sciences of the United States of America*, volume 36, pages 48–49, 1950.

[63] A. Y. Ng, A. X. Zheng, and M. I. Jordan. Link analysis, eigenvectors and stability. In *IJCAI*, pages 903–910, 2001.

[64] N. Nisan and A. Ronen. Algorithmic mechanism design. *In Proceeding of 31st Annual ACM Symposium on Theory of Computation*, pages 129–140, 1999.

[65] N. Nisan, T. Roughgarden, E. Tardos, and V. V. Vazirani. Algorithmic Game Theory. Cambridge University Press, 2007.

[66] I. Palacios-Huerta and O. Volij. The measurement of intellectual influence. *Econometrica*, 72(3):963–977, 2004.

[67] C. Papadimitriou and T. Roughgarden. Computing correlated equilibria in multi-player games. 2008.

[68] C. H. Papadimitriou. On the complexity of the parity argument and other inefficient proofs of existence. *J. Comput. Syst. Sci.*, 48(3):498–532, 1994.

[69] D. Pennock and M. Wellman. Toward a market model for Bayesian inference. In *In Proceedings of the 12th National Conference on Uncertainty in Artificial Intelligence*, pages 405–413. Morgan Kaufmann, 1996.

[70] K. Prasad. On the computability of Nash equilibria. *Journal of Economic Dynamics and Control*, 21(6):943–953, June 1997.

[71] H. Scarf. The Computation of Economic Equilibria. Yale University Press, 1973.

[72] J. Schummer. Manipulation through bribes. In *Journal of Economic Theory*, pages 91:180–198, 2000.

[73] L. Shapley. A note on Lemke-Howson algorithm. *Mathematical Programming Study*, 1:175–189, 1974.

[74] J. B. Shoven and J. Whalley. Applying General Equilibrium Theory. Cambridge University Press, 1992.

[75] T. L. Sibel Adali and M. Magdon-Ismail. Optimal link bombs are uncoordinated. *In Proceeding of AIRWeb*, 2005.

[76] M. Sipser. Introduction to the Theory of Computatioin. PWS Publishing Company, 1997.

[77] G. Slutzki and O. Volij. Scoring of web pages and tournaments—axiomatizations. *Social Choice and Welfare*, 26(1):75–92, January 2006.

[78] P. G. Spirakis and H. Tsaknakis. An optimization approach for approximate nash equilibria. In *Internet and Network Economics, 3rd International Workshop*, pages 42–56, 2007.

[79] M. J. Todd. Orientation in complementary pivot algorithms. *Mathematics of Operation Research*, 1(1):54–66, Feb 1976.

[80] W. Vickrey. Counterspeculation, auctions, and competitive sealed tenders. *Journal of Finance*, 16:8–37, 1961.

[81] K.-C. Wong and M. K. Richter. Non-computability of competitive equilibrium. *Economic Theory*, 14:1–28, 1999.

[82] B. Wu and B. D. Davison. Identifying link farm spam pages. In *WWW '05: Special interest tracks and posters of the 14th International Conference on World Wide Web*, pages 820–829, New York, NY, USA, 2005. ACM Press.

[83] Y. Ye. Exchange market equilibria with Leontief's utility: Freedom of pricing leads to rationality. *Theor. Comput. Sci.*, 378(2):134–142, 2007.

[84] Y. Ye. A path to the Arrow-Debreu competitive market equilibrium, 2007.

[85] M. Yokoo, Y. Sakuri, and S. Matsubara. The effect of false-name bids in combinatorial auctions: New fraud in internet auctions. *Games and Economic Behavior*, 46:174–188, 2004.