2012

Essays on social networks

Emre Unlu
Louisiana State University and Agricultural and Mechanical College, eunlu1@lsu.edu

ESSAYS ON SOCIAL NETWORKS

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Economics

by
Emre Unlu
B.S., Middle East Technical University, 2007
M.S., Louisiana State University, 2009
August 2012

Dedicated to my wife Zuhal Unlu.

Acknowledgments

I would like to express my deepest gratitude to my advisor, Dr. Sudipta Sarangi for patiently guiding me through academic process. My association with him has been the most fortunate experience of my entire academic journey. His encouragement, optimism and involvement have been a constant source of inspiration for me. Without his continuous support, this dissertation would not have been possible.

I would like to express my sincere gratitude to my committee member, Dr. Robert Newman for his constant support and encouragement. I have immensely benefited from my discussions with him. I am also deeply indebted to my other committee members, Dr. Yves Zenou Dr. Bulent Unel for their insightful comments and guidance. I am also thankful to Mary Jo Neathery and Judy Collins for taking care of all the professional details during my graduate studies.

I am thankful to Dr. Christophe Bravard, Dr. Surajit Bortotokey and Justin Cook for their comments and suggestions on my dissertation and I have truly enjoyed my discussions with them. I would like to thank my colleagues Duha, Collin, Justin, Chandan, Elif and Cihan for their much needed support, appreciation and warm friendship. I would like to thank my dear wife Zuhal Unlu especially for sharing my complaints, stress and happiness. I would like to thank my brother Dr. Fatih Unlu for his guidance and inspiring me to start an academic career.

My parents, Filiz and Aziz Unlu, have been my inspirations. I am extremely fortunate to have their unconditional love, support and sacrifice.

Table of Contents

Abstract

This dissertation makes a contribution to the social networks literature. The first essay provides a way of evaluating a player's contribution to their team and relates their effort to their market values. We extend the work of Ballester et al. (*Econometrica*, 2006) by incorporating a network outcome component in the players' payoff functions and prove that there is a unique interior Nash Equilibrium in pure strategies. We then illustrate this approach by collecting data from the UEFA Euro 2008 from Quarter Final to the Final matches. The passing data are utilized to construct the interaction network of each team. We only consider (net) successful discounted passes between every player for each game and each team. A discount measure for the passes made in the own half of the field is introduced to capture the importance of the passing effort. This data are used to determine the key player in each game, and also rank all the players in their order of importance in the game by the intercentrality measure. A sensitivity check for the model parameters is then carried out using a number of different simulations. Next, we identify key groups of players to determine which combinations of players played a more important role for their team. This information can be crucial for the clubs, managers and coaches who wonder about the joint performance of the players. In the last part, we investigate the returns of the intercentrality measure to the players. The most current market values and other observable characteristics of the players in the sample are obtained from *transfermarkt.de* and combined with the intercentrality measure from UEFA Euro 2008. The result of the first chapter indicates that players who have higher amount of interaction (high intercentrality value) with their teammates regardless of

their position on the field have significantly higher market values.

The second essay of my dissertation characterizes efficient networks in player and partner heterogeneity models for both the one-way flow and the two-way flow models. Player (partner) dependent network formation allows benefits and costs to be player (partner) heterogenous which is an important extension for modeling social networks in the real world. Employing widely used assumptions, I show that efficient networks in the two way flow model are minimally connected and have star or derivative of star type architectures, whereas efficient networks in the one way flow model have wheel architectures.

The third essay of my dissertation considers a non-cooperative network formation game where identity is introduced as a single dimension to capture the characteristics of a player in the network. Players access to the benefits from the link through direct and indirect connections. We consider cases where cost of link formation paid by the initiator or it is shared between the participants. Each player is allowed to choose their commitment level to their identities. The cost of link formation decreases as the players forming the link share the same identity and higher commitment levels. We then introduce link and node imperfections to the model. Each existing link in the network successfully transmits information with a probability. We consider two cases for reliability probability of existing links: a homogenous probability, p and heterogenous probability p_{ij}. We characterize the Nash networks and we find that the set of Nash networks are either singletons with no links formed or separated blocks or components with mixed blocks or connected.

Chapter 1

Introduction

A social network is a social structure made up of individuals called "nodes", which are tied (connected) by one or more specific types of interdependency, such as friendship, kinship, common interest, financial exchange, dislike, sexual relationships, or relationships of beliefs, knowledge or prestige. Social networks analysis in economics have great importance in investigating how the information transmits, health outcomes; labor market, crime, education outcomes and financial transactions. More specifically, social network analysis explain important questions such as how diseases spread, which products we buy, which languages we speak, how we vote, as well as whether or not we decide to become criminals, how much education we obtain, and our likelihood of succeeding professionally. The countless ways in which network structures affect our well-being make it critical to understand how networks are formed. My research is motivated by these important questions and focuses on network formation. My dissertation focuses on the identifying the important players in team environment, conflicts between stable and socially optimal outcomes and the link reliability in social network where identity is introduced as a single dimension variable to capture the different characteristics of the players.

Chapter 2: Key Players and Key Groups in Teams

In the recent years, understanding which node plays a more important role in the network configuration has become a main theme for many of the research papers. There are two main approaches in the literature to determine the key player problem. The first approach is based

on Graph Theory. This approach involves fragmentation and minimum distance principles. The fragmentation type measures count the number of fragments if we remove one player from the game at a time. Second approach is to utilize centrality measures. Centrality measures are generally used for determining the importance of a node in a network. To determine the importance of each node, various centrality measures have been developed. Then these measures are used to determine which node plays a more important role in a network.

The second approach relies on the centrality measures. There are three classes of centrality measures that are mostly developed in sociology such as: closeness like, betweeness like and degree like measures. Closeness like measures such as Freeman (1979) considers the geodesic distance from a given node to the other nodes. Betweeness type measures are based on the number of walks or flows that pass through a given node. The measures that are proposed in Katz (1953), Freeman (1979), Hubbell (1965), Bonacich (1972) and Sade (1989) are involved in degree like measures. Generally, this type of measures calculates the number of walks originating or terminating at a given node.

The seminal paper of Ballester, Calvo-Armengol and Zenou (2006, henceforth BCZ) provides the basic framework to identify key players using centrality measures in networks. BCZ (2006) provide microfoundation for the key player problem in a fixed network. Their model considers two vital ingredients: individual actions and interaction between players. In the Nash equilibrium of the game each player chooses their individual action taking both components into account. The key player is the one whose removal leads to the highest overall reduction in effort. Thus, their approach builds strategic behavior into the network, and combines both negative and positive aspects of the problem. My first chapter makes a contribution to this literature by modifying the original model of BCZ (2006) to capture team situations by providing the network members a common goal as well and illustrates the approach by using unique data from soccer matches. Team like situations dominate many social and economic environments. Firms and organizations are usually made up of

smaller groups or teams. Recommendation letters often mention person's ability to be a team player. An applicant's ability to be a team player is also tested in many interviews. Work environments like R&D groups, special task forces and even academia to a certain extent function as teams. Teamwork is an important feature of many games like soccer, basketball and volleyball. This makes understanding the contribution of individual members to a team very useful exercise. It can help design better teams and compensation packages. Identifying the key players in teams is also very useful for retention issues.

This chapter has two contributions to the key player literature. The first contribution is providing an empirical illustration of the approach using a team sport, namely soccer. We observe the passing effort of international soccer players to proxy the amount of interaction between players in UEFA European Championship 2008 and identify the key players and key groups in the network. It is important to note that we are not seeking the best player on the field. Rather, we are looking for the player whose contribution to his team is maximal. Finally, we show that players who have higher interactions (passings and receivings) and abilities have significantly higher ratings from experts and market values. We adapt the asymmetries in the interactions which provides a more detailed analysis. The second contribution is extending the BCZ (2006) model and introducing a team (or network) outcome component into the analysis to rank players according to their contributions to their teammates. The Nash equilibrium of the model provides the optimal amount of individual efforts of each player. It implies that if the player has a higher return for his individual actions or a higher ability parameter, then she will have more incentives to perform individual actions. Determining a key group instead of a key player is also an interesting aspect since more than one player may have equivalent level of contribution to their teammates. In addition to that, it is important to identify which combination of players has more importance within the network. This information is crucial for the team managers who wish to form a team with individuals who provide different inputs to their teammates. It is important to note that the members of the of key groups are not the best working peers but they are the ones

whose joint contribution to their team is maximal. The approach in the chapter can be extended and replicated to more general team situations as well as for other sports where players work in teams.

Chapter 3: Efficient Networks in Models of Player and Partner Heterogeneity

A substantial literature focuses on individual optimization through Nash networks, where no agent can make herself better off by deviating from her current strategy, given the strategies of the other players. A central theme in the literature on network formation is the conflict between the set of stable and efficient networks. Even though Nash networks are very well studied in broader manners, there has been few studies about efficiency. This chapter fills the gap in the literature of network formation by describing the efficient network architectures under heterogeneous agents. In this chapter, we address the issue of efficiency in the form of maximizing aggregate utility of players in the network and compare the architectures of efficient networks with Nash networks.

The seminal papers on network formation are Jackson and Wolinsky (1996) and Bala and Goyal (2000). Jackson and Wolinsky (1996) provide two sided link formation, where the cost of forming links is shared by the participants and introduces pairwise stability. Bala and Goyal (2000) provide a theoretical framework to address network formation in a non-cooperative setting with homogenous players, where the cost of forming links is on one side. Bala and Goyal (2000) discuss two different types of flow. In the two way flow model, the network is undirected, so both players participating in a link access can access benefits from each other. However, in the one way flow model the network is directed; hence, only the initiating player can access the benefit of the link. This chapter follows the Bala and Goyal (2000) framework, where link costs are levied only by the person initiating the link. Galeotti et al. (2006) relaxes the homogeneity assumption in the two way flow model, so the benefits from a link and the cost of sponsoring a link are player dependent. A more recent paper (Billand et al. 2010) introduces partner heterogeneity in the two way flow model, where the benefit and cost of making a link is partner heterogeneous, meaning that it only depends

on whom is being accessed in terms of the benefits and costs. Although Nash networks are clearly identified under heterogeneity, not much has been done about efficient networks under heterogeneity. Specifically, in this chapter, we illustrate situations where Nash and efficient networks do not coincide. Our approach is to start with a general payoff specification in a two way flow model satisfying common assumptions in the literature. Without imposing any restrictions, efficient networks can have maximal diameter and it is not possible to characterize the architecture. Once we allow for heterogeneity between players, the efficient network architectures depend on four factors: the value of players (benefit obtained by linking to each player), the number of minimum cost players, the difference in cost between the minimum cost player(s) and the other players and finally the functional form of payoffs. In this chapter, we provide the architecture, as well as, the diameter to identify the efficient networks. The architecture provides information about how the efficient networks look, and the diameter helps to determine the maximum distance between any two players in the network.

Our results indicate significant differences in the architecture and diameter between Nash and efficient networks in two-way flow models. We find that the crucial difference between player and partner heterogeneity models is in the change of the player who sponsors the links. There are certain differences in the network formation model and results under player or partner heterogeneous players. Finally, for one way flow models, there is no difference between player and partner heterogeneity models and the Nash and efficient architectures coincide.

Chapter 4: Identity and Link Reliability in Networks

Many empirical observations suggest that one of the key determinants in the network formation is the similarities in the characteristics of the players. In this chapter, we consider a network formation game where the identity characteristics are introduced to capture the similarities or differences between the players. We consider a framework where the links between players are not fully reliable i.e, the success of a link is probabilistic. We assume

that identity characteristics are assigned exogenously to the players and each player decides about how much to commit to her current characteristic. For example, exogenously given characteristics can be listed as race, nationality and culture and in the network formation game players choose how much to commit to their given characteristic as well as their linking strategies.

Esteban and Ray (1994), Akerlof and Kranton (2000), Chen et al. (2000), Fryer and Jackson (2002), Currarini, Jackson, and Pin (2008), Marti and Zenou (2009), Dev (2009) and Dev (2010) study the importance of identity dimension in the network formation. The standard network formation models used is the related to the literature on non-cooperative network formation models pioneered by Bala and Goyal (2000a) as well as Bala and Goyal (2000b) with related work by Galeotti, Goyal, and Kamphorst (2003), Hojman and Szeidl (2006), Billand et al. (2006), Galeotti (2006). The other strand in this literature follows from Jackson and Wolinsky (1996). Bala and Goyal (2000b), Haller and Sarangi (2003) explore the effects of link readability on the network formation. The important deviation of this chapter from this literature is to combine the non-cooperative network formation models with identity and probabilistic link reliability. In our model, the links will be formed based on identity characteristics and each link or player can fail with a certain probability. Probabilistic links provide many incentives form different architectures and yields results that corresponds to the empirical observations.

We consider a non-cooperative network game, players how much to commit their identities and their linking strategies. Player's commitment decision affects the cost of forming a link. By choosing her commitment level, a player reveals which type of players she can easily form a link. We assume that players with same identities and higher commitment levels can make links easier. Cost of link formation with different identities and higher commitment levels becomes very costly. Players access to information by making links with others. We assume two way flow and undirected network without decay which means that both players access to the same benefit of the link through direct or indirect connections.

We consider two cases of costly link formation. In the first case, the cost of the link is paid by the initiator. In the second case, the players who are involved in the link share the cost depending on their initial link offers. Hence, the second case enables us to investigate the network formation under mutual consent. For simplicity of the analysis, we assume that link reliability probabilities within the same identity group are equal to each other. We show that costly link formation between different identities can lead to fragmented architectures. However, with fully heterogeneous link failures, it is possible to have more integrated groups with different identities which may include many components however this only occurs if the link between the different identity groups is more reliable than a link between the players when both players involved in the link share the common identity characteristic. The actual determination of reliability probability ranges is fairly complicated and we demonstrate the intuition with an example for some cases.

Our model combines the results of Bala and Goyal (2000b) and Dev (2010). Compared to Bala and Goyal (2000b) model, the cost of link formation is heterogenous and its a function of players' identity characteristics and commitment levels. This allows us to study the effect of identity on their model. Compared to Dev (2010) model, our model introduces probabilistic link reliability and the Nash networks may include different identity characteristics in a single component.

Chapter 2

Key Players and Key Groups in Teams

2.1 Introduction

In recent years the literature in economics on networks has begun to focus on centrality measures. The seminal paper of Ballester et al. (2006) provides the basic framework to identify key players using centrality measures in networks. However, to the best of our knowledge there is no paper that studies centrality measures using actual data. This chapter makes a contribution to this literature - it modifies the original model of Ballester et al. (2006) to capture team situations by providing the network members a common goal as well. After obtaining the necessary theoretical results we apply it to a unique data set collected from soccer games.

Team like situations dominate many social and economic environments. Firms and organizations are usually made up of smaller groups or teams. Recommendation letters often mention person's ability to be a team player. An applicant's ability to be a team player is also tested in many interviews. Work environments like R&D groups, special task forces and even academia to a certain extent function as teams. Teamwork is an important feature of many games like soccer, basketball and volleyball. This makes understanding the contribution of individual members to a team very useful exercise. It can help design better teams and compensation packages. Identifying the key players in teams is also very useful

for retention issues. In this chapter, we develop a method for identifying key players and key groups in teams.

There is a substantial literature in graph theory on identifying the key node in a network. To determine the importance of each node, various centrality measures have been developed. These may be degree based measures that take into account the number of links that emanate and end at a node (see for instance Katz (1953), Freeman (1979), Hubbell (1965), Bonacich (1987) and Sade (1989)). Closeness measure like those developed by Sabidussi (1966) and Freeman (1979) use some type of topological distance in the network to identify the key players. Another measure called betweeness measure (see for instance Freeman (1979)) uses the number of paths going through a node to determine its importance. Borgatti and Everett (2006) develop a unified framework to measure the importance of a node. Borgatti (2006) identifies two types of key player problems (KPP). He argues that in KPP-positive situation key players are those who can optimally diffuse something in the network. In a KPP negative situation key players are individuals whose removal leads to maximal disruption in the network.

Ballester, Calvo-Armengol and Zenou (2006, henceforth BCZ) provide microfoundation for the key player problem in fixed networks. Their model considers two vital ingredients: individual actions and interaction between players. In the Nash equilibrium of the game each player chooses their individual action taking both components into account. The key player is the one whose removal leads to the highest overall reduction in effort. Thus, their approach builds strategic behavior into the network, and combines both negative and positive aspects of the problem. Calvo-Armengol, Patacchini and Zenou (2009, henceforth CPZ) extend BCZ (2006) and propose a peer effects model to study educational outcomes.

We develop a Team Game based on the individual actions and interactions between players. Additionally, each player gains utility when the team achieves its desired outcome. This team outcome depends on individual effort and an ability term for each player. Another interesting feature is that following BCZ (2006) we define key player problem from a social

planner's perspective. In context of teams, team leader or head coaches can be regarded as the social planners. We then develop two new intercentrality measures that take into account two different criteria for the social planner. The first intercentrality measure is derived form the reduction in aggregate Nash equilibrium effort levels whereas the second intercentrality measure is derived using the externality a player gets from her teammates.

This chapter has two contributions to the key player literature. The first contribution is providing an empirical illustration of the approach using a team sport, namely soccer. We observe the passing effort of international soccer players to proxy the amount of interaction between players in UEFA European Championship 2008 and identify the key players and key groups in the network. It is important to note that we are not seeking the best player on the field. Rather, we are looking for the player whose contribution to his team is maximal. Finally, we show that players who have higher interactions (passings and receivings) and abilities have significantly higher ratings from experts and market values. We adapt the asymmetries in the interactions which provides a more detailed analysis. Our approach is different from CPZ (2009) in two aspects. CPZ (2009) focuses on peer effects in a student environment whereas we are directly interested in determining key players and ranking the individuals according to their contribution to their teams. The second contribution is extending the BCZ (2006) model and introducing a team (or network) outcome component into the analysis to rank players according to their contributions to their teammates. The Nash equilibrium of the model provides the optimal amount of individual efforts' of each player. It implies that if the player has a higher return for his individual actions or a higher ability parameter, then she will have more incentives to perform individual actions.

Determining a key group instead of a key player is also an interesting aspect since more than one player may have equivalent level of contribution to their teammates. In addition to that, it is important to identify which combination of players have more importance within the network. This information is crucial for the team managers who wish to form a team with individuals who provide different inputs to their teammates. It is important to note

that the members of the of key groups are not the best working peers but they are the ones whose joint contribution to their team is maximal. Temurshoev (2008) extends BCZ (2006) paper by introducing the key group dimension. Temurshoev (2008) searches for the key group, whose members are, in general, different from the players with highest individual intercentralities. We apply Temurshoev's (2008) approach to determine the key groups of players.

The approach in this chapter can be extended and replicated to more general team situations as well as for other sports where players work in teams. However, we provide an empirical example using international soccer matches. Taking this approach has some advantages. First, since soccer is a team sport and the payoff of players depends on the team outcome. Second, interactions within soccer teams are observable and passing effort of players is a good metric to identify these interactions. We create a unique passing data from UEFA European Championship 2008 and identify the key players and key groups of teams which played in the Quarter Final, Semi-Final and Final stage of the tournament.[1]

Subsequent sections of the chapter are organized as follows: Section 2 defines the team game and various centrality measures. It also identifies the Nash equilibrium and our team intercentrality measures. Section 3 motivates the use of soccer data as an empirical application. Section 4 identifies the empirical methodology. The chapter concludes with discussions and possible extensions.

2.2   Team Game

In this section, we first define the team game and interpret the model for soccer. Then, we introduce the various centrality measures and find the Nash equilibrium of the game. Finally, we provide the relationship between the Nash equilibrium and the intercentrality

---

[1]Fifty European national teams played qualifying stages and only 16 of them were qualified for the UEFA Euro 2008. So, it is reasonable to expect that the quality of the players in the national tournaments are similar. Thus, interaction between players plays a crucial role in determining the outcome of the matches making our results more important.

measure(s) considering two different scenarios.

We begin by introducing the team game. We define the individual player's payoff function using the notation of BCZ (2006) as far as possible. We also interpret the model variables.

$$U_i(x_1, ..., x_n) = \alpha_i x_i + \frac{1}{2}\sigma_{ii} x_i^2 + \sum_{j \neq i} \sigma_{ij} x_i x_j + \theta Z. \tag{2.1}$$

The first two terms form a standard quadratic utility function where $x_i \geq 0$ is defined as the individual effort of player i. $\alpha_i > 0$ stands for the coefficient of individual actions and $\sigma_{ii} < 0$, the coefficient of the second term, defines concavity in own effort i.e., $\partial^2 U_i/\partial x_i^2 = \sigma_{ii} < 0$. For simplicity we assume that these coefficients are identical for all players and we drop the subscript.[2] The third term captures the bilateral influences between players with σ_{ij} being the coefficient of this term. Let $\mathbf{\Sigma} = [\sigma_{ij}]$ be the matrix of these coefficients. Note that σ_{ij} could be positive or negative. The last expression is the team outcome term denoting the desired team goal. It represents how an individual's utility depends on team outcome. We assume that the team outcome, Z is a linear function of each player's effort and ability parameter. The coefficient θ is a scale parameter that can be used to capture the importance of the game.

For soccer, x_i term can be interpreted as the attributes such as creativity, distance traveled, attention, speed or shots on goal. α_i measures the returns from individual actions and σ_{ii} introduces the concavity in effort in the sense that as players perform actions, they spend stamina and it becomes costly. We utilize passing behavior to infer the interactions between players. Thus, player i's utility from interacting with player j is weighted by how often he passes to j. σ_{ij} can be interpreted as the complementary action of player i on player j. For the case of soccer, σ_{ij} indicates the number of (discounted) successful passes from player i to j minus the number of (discounted) unsuccessful passes from player i to j.[3] This

---

[2] We relax this assumption and consider the cases when α_i and σ_{ii} can be different for every player in Proposition 1 (a)-(b).

[3] We acknowledge that there are other complementarity actions other than passing behavior; however, these factors are very difficult to measure. Taking passing as a metric for complementarity action simplifies the empirical model and enables us to quantify.
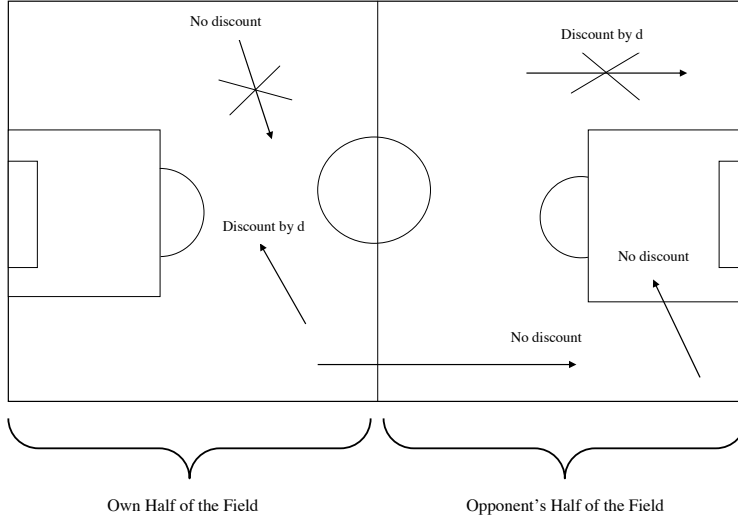
Figure 2.1: Unsuccessful passing attempts are shown by cross overs.

means that if σ_{ij} is positive, number of successful passes from player i to player j exceeds the number of unsuccessful passes from player i to j.

While measuring the complementarity in players' effort, we introduce a discounting parameter, d in constructing the Σ matrix. Passes that are made far from the opponent's goal have little influence on creating a goal scoring opportunity. Therefore, we discount the passes that are made in own half of the field by a factor $0 < d < 1$. On the other hand, if player i successfully passes the ball to player j, and if player j is the opponent's half, then we do not discount that pass. Unsuccessful passes are discounted in the opposite way. If a player i losses the ball while trying to pass to player j, we look at the position of j. If player j is in the opponent's half, then we discount that loss by d. Similarly, if player i losses the ball while trying to pass to player j who is in his own half, then we do not discount that loss. Basically, if player i losses the ball near his own goal then that is a serious loss for the team. The intuition for not discounting the unsuccessful passes made in the own half is that players have to run back which hurts the team's play and may create an opportunity for the opponent to start an attack from an advantageous position. The below figure shows an example of discounting.

13

For the empirical model, the ability parameter, δ_i is defined as the scoring probability of player i where δ_i = Number of goals scored by player i / Number of total shots on goal of player i. Alternatively, Z can also be defined as the outcome of the match. Specifically, Z can be assumed to be taking values of $\{1, 0, -1\}$ where $Z = 1$ implies that the team wins the game, $Z = 0$ implies that the match ended in a draw, and $Z = -1$ implies that the team lost the match. With the above definition of Z, the Nash equilibrium of the Team Game is identical to the BCZ (2006) and allows us to use the ICM provided by the authors in Remark 5 (pg. 1412). For a soccer game this could be winning the game or scoring more goals. This term contains a the same set of variables for all players since they all share the same outcome. For simplicity, let $Z = \sum_{i=1}^{n} \delta_i x_i$ where δ_i defines each individual's ability to help achieve the team's goal. The parameter θ is a scale factor that could be used to capture the importance of different events for the team.[4]

Our team game differs from that of BCZ (2006) model in the last term. This allows us to consider the n players acting together towards a common objective. While alternative formulations of this are possible, we believe our framework has certain advantages. First, it allows for explicit comparison with BCZ (2006). Second, while all effort by player provides a utility, the effort adjusted by the ability parameter is important for achieving the team outcome. This can be useful for empirical illustration since it may not be possible to obtain data on α_i and σ_{ii}. The ability parameter δ_i on the other hand could be obtained from available data.

In order to proceed following BCZ (2006), we let $\underline{\sigma} = \min\ (\sigma_{ij}|i \neq j)$ and $\bar{\sigma} = \max\ (\sigma_{ij}|i \neq j)$. We assume that $\sigma < \min\ (\underline{\sigma}, 0)$. Let $\gamma = -\min\ \underline{\sigma}, 0 \geq 0$. If efforts are strategic substitutes for some pair of players, then $\underline{\sigma} < 0$ and $\gamma > 0$; otherwise, $\underline{\sigma} \geq 0$ and $\gamma = 0$. Let $\lambda = \bar{\sigma} + \gamma \geq 0$. We assume that $\lambda > 0$. Define $g_{ij} = (\sigma_{ij} + \gamma)/\lambda$. Note that, the g_{ij}'s are weighted and directed allowing us to obtain relative complementarity measures. Consequently, the elements g_{ij} of the weighted adjacency matrix lie between 0 and 1. If we do

---

[4]In principle, one could define θ_Z to capture the importance of the level of achievement in the team's objective. Here, for simplicity we assume it to be θ.

not use a weighted $\mathbf{G}$ matrix then it contains only 0s and 1s as its elements. This will imply that the additional weight for having more connections with the same player is zero. So, when $g_{ij} = 1$ then there is a connection and if $g_{ij} = 0$ then there is no connection between player i and player j. However, it is very important to identify the relative interaction between players rather than just considering if there is a connection between player i and j. Thus, using a weighted $\mathbf{G}$ matrix is important to illustrate team environments.

The adjacency matrix $\mathbf{G} = [g_{ij}]$ is defined as a zero diagonal nonnegative square matrix. The zero diagonal property assures that no player is connected to themselves (i.e, there are no direct loops from player i to i.) Then, $\mathbf{\Sigma}$ matrix which captures the cross effects can be decomposed into the following expression:

$$\mathbf{\Sigma} = -\beta\mathbf{I} - \gamma\mathbf{U} + \lambda\mathbf{G} \tag{2.2}$$

where $-\beta\mathbf{I}$ shows the concavity of the payoffs in terms of own actions, $-\gamma\mathbf{U}$ shows the global interaction effect, and $\lambda\mathbf{G}$ shows the complementarity in players' efforts. Using the above decomposition, Equation (1) becomes:

$$U_i(x_1, ..., x_n) = \alpha x_i - \frac{1}{2}(\beta - \gamma)x_i^2 - \gamma \sum_{j=1}^{n} x_i x_j + \lambda \sum_{j=1}^{n} g_{ij} x_i x_j + \theta_z Z \tag{2.3}$$

for all players $i = \{1, ..., n\}$.

2.2.1 Centrality Measures

Here, we define the centrality measures needed to identify the key player. Let $\mathbf{M}$ be a matrix defined as follows:

$$\mathbf{M}(\mathbf{g}, a) = [\mathbf{I} - a\mathbf{G}]^{-1} = \sum_{k=0}^{\infty} a^k \mathbf{G}^k. \tag{2.4}$$

The above matrix keeps track of the number of paths that start from player i and end at player j with a decay factor, a and a given adjacency matrix $\mathbf{G}$. Note that players can also

15

contribute to their teammates through indirect connections, but these have lower weights.

Following BCZ (2006), we define the Bonacich centrality measure as:

$$\mathbf{b}(\mathbf{g}, a) = [\mathbf{I} - a\mathbf{G}]^{-1} \cdot \mathbf{1} \tag{2.5}$$

where $\mathbf{1}$ is a $n \times 1$ vector of ones, n is number of players in the team and $\mathbf{I}$ is a $n \times n$ identity matrix. The Bonacich centrality measure counts the total number of paths that originates from player i. Note that b_i is the row sum of the $\mathbf{M}$ matrix. Equivalently, the Bonacich centrality measure is $b_i(\mathbf{g}, a) = m_{ii}(\mathbf{g}, a) + \sum_{i \neq j} mij(\mathbf{g}, a)$. Next, we define a weighted Bonacich centrality measure with the ability parameter, δ_i as the weight:

$$\mathbf{b}_\delta(\mathbf{g}, a) = [\mathbf{I} - a\mathbf{G}]^{-1} \cdot \boldsymbol{\delta} \tag{2.6}$$

We define another centrality measure which accounts for the weighted receivings of the players where the weights are given by δ_i:

$$r_i(\mathbf{g}, a) = \sum_{j=1}^{n} m_{ji}(\mathbf{g}, a) \times \delta_i \tag{2.7}$$

This (receiving) centrality measure takes into account the paths that end in player i weighted by the ability parameter of the player. This measure captures the externality a player gets from her teammates and weights it according to the ability of the player.

The BCZ (2006) intercentrality measure (ICM) for an asymmetric $\mathbf{G}$[5] is given by:

$$\tilde{c}_i(\mathbf{g}, a) = b_i(\mathbf{g}, a) \times \frac{\sum_{j=1}^{n} m_{ji}(\mathbf{g}, a)}{m_{ii}(\mathbf{g}, a)} \tag{2.8}$$

Unlike the Bonacich centrality measure, ICM takes into account both the connections that player i sends to her teammates and the number of connections that player i receives.

---

[5]Note that, in the context of teams, the $\boldsymbol{\Sigma}$ matrix is unlikely to be symmetric since the number of paths from player i to player j will be different for at least one pair. Hence, an asymmetric $\boldsymbol{\Sigma}$ matrix will lead to an asymmetric $\mathbf{G}$ matrix.

Throughout the chapter, we define two intercentrality measures to take into account possible different objectives of the social planner while identifying the key players. We provide two alternative objectives for the social planner. In the first case, the social planner determines the player whose removal leads to the highest amount of reduction in the aggregate Nash Equilibrium effort. We derive the intercentrality measure for this objective and call this measure as team intercentrality measure, $TICM$. For an asymmetric $\mathbf{G}$ matrix, we define $TICM$ as:

$$\bar{c}_i(\mathbf{g}, a) = b_i(\mathbf{g}, \lambda^*) \times \left(\frac{\sum_{j=1}^n m_{ji}(\mathbf{g}, \lambda^*) + \sum_{j=1}^n m_{ji}(\mathbf{g}, \lambda^*)\delta_j}{m_{ii}(\mathbf{g}, \lambda^*)} \right) \tag{2.9}$$

$TICM$ measures player i's contribution to the interaction matrix as well as her contribution to the team outcome. The difference between ICM and $TICM$ is in the last term in the parentheses which captures player i's importance in creating the team outcome.

Alternatively, one can argue that the social planner is equally interested in the interaction between players as well as the externality term in the payoff of each player. While the $TICM$ measure above takes the first effect into account it does not take the second effect into account. Therefore, we introduce a second measure of the importance of a player in the game by taking the interaction into account as well as how the contribution of other players affects the performance of each player weighted by their ability. We derive the intercentrality measure for this objective and call this measure as team intercentrality measure with externality, $(TICM^e)$. For an asymmetric $\mathbf{G}$ matrix, we define $TICM^e$ as:

$$\hat{c}_i(\mathbf{g}, a) = b_i(\mathbf{g}, a) \times \frac{\sum_{j=1}^n m_{ji}(\mathbf{g}, a)}{m_{ii}(\mathbf{g}, a)} + \sum_{j=1}^n m_{ji}(\mathbf{g}, a) \times \delta_i \tag{2.10}$$

The primary difference between ICM and $TICM^e$ is in the last term which measures the externality player i receives from her teammates weighted by the ability parameter of the player.

2.2.2 Nash Equilibrium of the Team Game

In this section, we show that the Team Game has a unique interior Nash equilibrium by the following theorem. Nash equilibrium shows that the aggregate equilibrium effort increases with higher centrality measures.

Theorem 2.1 Consider a matrix of cross-effects which can be decomposed into (3). Suppose $\sigma_{ij} \neq \sigma_{ji}$ for at least one $j \neq i$, $\beta/\lambda > (\rho(\mathbf{G}))$ and a small enough θ. Define $\lambda^* = \lambda/\beta$. Then, there exists a unique, interior Nash Equilibrium of the team game given by:

$$\hat{x}^*(\mathbf{\Sigma}) = \frac{\alpha \mathbf{b}(\mathbf{g}, \lambda^*) + \theta \mathbf{b}_\delta(\mathbf{g}, \lambda^*)}{\beta + \gamma \hat{b}(\mathbf{g}, \lambda^*)}$$

where $\hat{b}(\mathbf{g}, \lambda^*) = \sum_{i=1}^{n} b_i(\mathbf{g}, \lambda^*)$ and $\hat{x}^* = \sum_{i=1}^{n} x_i$.

Proof: The proof is an adaptation of BCZ (2006) and can be found in Appendix.

The Nash equilibrium of the Team Game has some interesting implications. First, it identifies the optimal effort of individuals in the network based on the given interactions between players. It also explains why some players provide higher individual effort by indicating that players who have higher ability parameter or who make more interactions with their teammates, will have higher individual effort.

A unique interior Nash equilibrium exists even when players have heterogeneity in returns (α_i) and concavity (σ_{ii}) in individual actions are proved in Proposition 1 (a) and (b).

Proposition 1 (a): Suppose $\alpha_i \neq \alpha_j$ and θ is small enough then Nash equilibrium of the Team Game is:

$$\hat{x}^*(\mathbf{\Sigma}) = \frac{\mathbf{b}_\alpha(\mathbf{g}, \lambda^*) + \theta \mathbf{b}_\delta(\mathbf{g}, \lambda^*)}{\beta + \gamma \mathbf{b}(\mathbf{g}, \lambda^*)}$$

(b): Suppose $\alpha_i \neq \alpha_j$, $\sigma_{ii} \neq \sigma_{jj}$ for at least one player and θ is small enough, then Nash equilibrium of the team game is:

$$\hat{x}^*(\mathbf{\Sigma}) = \frac{\mathbf{b}_{\widetilde{\alpha}}(\mathbf{g}, \widetilde{\lambda}^*) + \theta \widetilde{\mathbf{b}_\delta}(\mathbf{g}, \widetilde{\lambda}^*)}{\widetilde{\beta} + \widetilde{\gamma} \hat{\mathbf{b}}(\mathbf{g}, \widetilde{\lambda}^*)}$$

Proof: See Appendix.

2.2.3 Key Player

In this subsection, we develop two alternative measures of identifying the key player in the teams by considering different criteria of the social planner. In the first case, social planer is interested in finding the player whose removal causes highest amount of reduction in the aggregate Nash Equilibrium. In this approach, we determine the key player as in BCZ (2006) which is to minimize sum of efforts after removal. We denote by $\mathbf{G}^{-i}$ (resp. $\mathbf{\Sigma}^{-i}$) the new adjacency matrix (resp. matrix of cross-effects), obtained from $\mathbf{G}$ (resp. from $\mathbf{\Sigma}$) by setting all of its ith row and column coefficients to zero. The resulting network is $\boldsymbol{g}^{-i}$. The social planner's objective is to reduce $\mathbf{x}^*(\mathbf{\Sigma})$ optimally by picking the appropriate player from the population. Formally, she solves $max \ \{\mathbf{x}^*(\mathbf{\Sigma}) - \mathbf{x}^*(\mathbf{\Sigma}^{-i}) | i = 1, ..., n\}$. This is a finite optimization problem and has at least one solution. Theorem 2.2 shows that the player who has the highest amount of $TICM$ will be the solution of this problem.

Theorem 2.2 Let $\beta > \lambda \rho_1(\mathbf{G})$. The key player of the Team Game, i^* solves $max\{\mathbf{x}^*(\mathbf{\Sigma}) - \mathbf{x}^*(\mathbf{\Sigma}^{-i}) | i = 1, ..., n\}$ and has the highest team intercentrality measure $(TICM)$ in $\mathbf{g}$, that is $\bar{c}_{i^*}(\mathbf{g}, \lambda^*) > \bar{c}_i(\mathbf{g}, \lambda^*)$ for all $i = 1, ..., n$.

Proof: See Appendix

Alternatively, social planner may consider the externalities players get from their teammates which is not included in the Nash Equilibrium of the team game. It might be the case that social planner is interested in considering each player's effect on the interaction matrix as well as taking the externality into account. BCZ (2006) provides the effect of player i's removal to the interaction matrix under Remark 5 for the asymmetric case. We define the externality player i receives from her teammates $r_i(\mathbf{g}, \lambda^*) = \sum_{j \neq i} m_{ji}(\mathbf{g}, \lambda^*)$ and we weight that with the ability parameter of the player. Theorem 2.3 indicates that the player who has the highest amount of $TICM^e$ will be the solution of this case.

Theorem 2.3 Let $\beta > \lambda \rho_1(\mathbf{G})$. The key player of the Team Game with externalities, i^* solves $b(\mathbf{g}, \lambda^*) - b(\mathbf{g}^{-i}, \lambda^*) + r_\delta^i(\mathbf{g}, \lambda^*)$ and has the highest $TICM^e$ in $\mathbf{g}$, that is $\hat{c}_{i^*}(\mathbf{g}, \lambda^*) > \hat{c}_i(\mathbf{g}, \lambda^*)$ for all $i = 1, ..., n$.

Proof: See Appendix

2.3   Soccer: A Team Game and The Role of Passing

Modern soccer is very much a team game. The performance of players depends crucially on each other's actions and interaction between players forms a vital component of the game. Soccer coaches, training books and authorities emphasize the team aspect of the game. As the great Brazilian soccer player Pele said in a press conference in Singapore in November 2006, "I think the problem with Brazil was lack of teamwork because everybody used to say Brazil will be in the final." Pele added that Brazil had the best individual players against France, but they lost the game because they could not play as a team.[6] On November 29, 2007, Gerard Houllier, the famous technical director of the French Football Federation, speaking at the 9th UEFA Elite Youth Football Conference summed this up as "Teamwork is the crux of everything."[7]

One important aspect of soccer that makes it a team game is the fact that passing is a very crucial part of the game. In the early days of soccer, the game was based on individual skills such as tackling and dribbling. In 1870s, the Scots invented the passing game and everyone soon realized that it is easier to move the ball than players since the ball travels faster than humans. Since then passing and receiving have become a key part of a soccer team's strategies. A soccer training manual by Luxbacher (2005) emphasizes the importance of passing in the following "Passing and receiving skills form the vital thread that allows 11 individuals to play as one - that is the whole to perform greater than the sum of its parts." Similarly, Miller and Wingert (1975) addresses the importance of passing in soccer by stating

---

[6]See http :// findarticles.com/p/articles/mi kmafp/is 200611/ai n16939060.
[7]See http://www.uefa.com/uefa/keytopics/kind=1024/newsid=629284.html
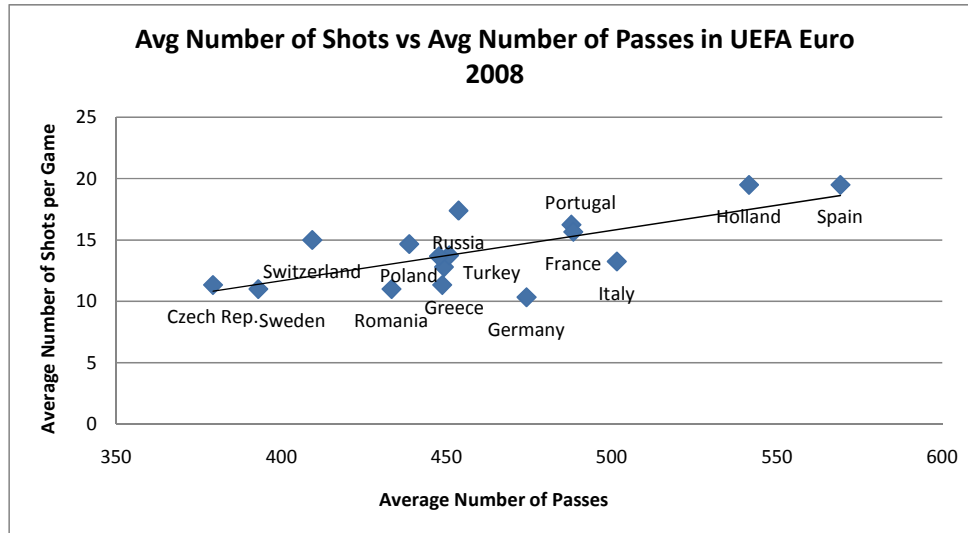
Figure 2.2: The relationship between average number of shots per game and average number of passes per game in UEFA Euro 2008.

that "There are no more crucial skills than passing in soccer because soccer is a team sport. The most effective set plays involve accurately passing and receiving the ball."

Luhtanen et al. (2001) report that successful passes at the team level are important for explaining the success in the UEFA European Championship 2000. Specifically, Luhtanen et al. (2001) document that there is one to one relationship between the ranking of the team in Euro 2000 and the ranking of the team in terms of successful passing and receivings. This provides evidence that in the competition, having more interactions in the team leads to better results. It seems reasonable that passing is a good metric for identifying the interactions between players.

Figure 2 displays the relationship between average number of shots per game and average number of passes per game of the national teams in the UEFA European Championship 2008.[8] The correlation coefficient between these variables is 0.7. The regression coefficient obtained from regressing average number of shots on goal per game on average number of passes per game indicates that on the average 27 passes created 1 additional shot on goal for

[8]This data was accessed from the following website http://www1.uefa.com/tournament/statistics/teams. It is available from the authors upon request.

the team in Euro 2008. This is consistent with the idea that teams need ball possession to create goal scoring opportunities which directly affects the outcome of the match. Clearly, passing is an important interaction variable in our dataset.

There are some advantages to using passing to infer player interactions. First, it is pairwise and both the originator and receiver must be successful to complete the action. The pairwise aspect of passing enables us to utilize the network theory to understand the contribution of each player to the team. Second, passing as a measure of interaction is observable and easily quantifiable. Data for other aspects of the soccer such as tackling, dribbling or off the ball movement of players are very hard to observe. In addition, often identifying the quality of these actions require subjective judgement. Finally, even if we had data about these aspects, it would be still difficult to quantify those variables exactly.

2.4 Empirical Methodology

This section illustrates our methodology for identifying the key player and key groups in soccer teams. First, we describe our data collection process. Next, we calculate the ICM and $TICM^e$ by using the corresponding definitions and provide our results for the key players and key groups. Finally, we conduct sensitivity checks for the model parameters which are used for identifying key players.

2.4.1 Data and Results

Our data consists of all the matches from the Quarter Final onwards for the UEFA European Championship 2008. All the data that is used in the chapter is available from the authors on request. Unfortunately, official passing data from UEFA's website is not adequate for our study due to a number of reasons. First, UEFA provides data only on the successful passes between player i and j and excludes the unsuccessful passes. Second, UEFA statistics do not provide the passing position of the players which is important for assessing the quality

of passing. Hence, we created a unique passing data set ourselves by watching the matches from DVD's. This was done by freezing the frame at the time of the passing attempt and recording the player making the pass and the receiver in a matrix by noting the position of the receiver. We also discounted the passes using the method described in the previous section.[9] The net discounted passes are used to determine the $\sigma'_{ij}s$ in $\mathbf{\Sigma}$ matrix. As expected, the $\mathbf{\Sigma}$ and $\mathbf{G}$ matrices are both asymmetric.

In order to facilitate comparisons across matches, we define a tournament wide λ and γ which are the same for every team. First we obtain the highest amount of positive and negative interaction between each pair of players throughout the tournament. Using this, the tournament wide γ and λ parameters are chosen as 5 and 20 respectively. This allows us to compare the same player's intercentrality measure from different matches as well as to compare the intercentrality measure of different players from different matches.[10]

Data for creating the tournament wide scoring probabilities of each player was obtained from ESPN's website.[11] Ideally, the life time scoring probability of a player would be δ_i in the model. However, this data is not available and we use the tournament wide measure as a proxy for this. Next, we calculate the $\mathbf{M}$ matrix, centrality vector ($\mathbf{b}$) and (team) intercentrality vectors $\mathbf{c}$ and $\hat{\mathbf{c}}$ by using the definitions provided in Section 2. Note that assigning a value to a is crucial for obtaining a pure and interior Nash equilibrium. BCZ (2006) note that for the case of asymmetric $\mathbf{\Sigma}$ and $\mathbf{G}$ matrices, a should be less than the spectral radius of $\mathbf{G}$, which is inverse of the norm of the highest eigenvalue of $\mathbf{G}$. The greatest eigenvalue of $\mathbf{G}$ matrices for the teams in the sample is 7.07 and hence following the above rule, the decay factor, a, is set to 0.125 for all matches. Since we did not have any guide lines for discount factor, $d \in [0,1]$ we assume that $d = 0.5$ for all matches. Using all

---

[9] According to our passing data on average 972.5 successful passes occurred in a match and it is a tedious exercise to record every passing attempt. Note that we also take into account the unsuccessful passing effort which are 272.3 on average in each match.

[10] Note that the Netherlands vs Russia, Spain vs Italy and Croatia vs Turkey matches went into the extra time. Therefore, comparing the players in the these games with those ended in 90 minutes is not possible. The cross comparisons are valid for match lengths of the same duration. We discuss this issue in more detail in the next subsection.

[11] http://soccernet.espn.go.com/euro2008/stats

of these parameters we then compute ICM and $TICM^e$ of each player.

The corresponding calculations for the Final, Semi Final and Quarter Final games for Euro 2008 are reported for each team in Tables 2.1-2.7.

Table 2.1: Spain vs Germany Final Game, a=0.125 and d=0.5

Name	Position	$\hat{c}_i$	c_i	Name	Position	$\hat{c}_i$	c_i
Xavi	M	4.31	3.97	Lahm	D	4.38	3.73
Fabregas	M	3.97	3.35	Schweinsteiger	M	4.37	3.57
Senna	M	3.66	3.66	Frings	M	4.02	4.02
Ramos	D	3.53	3.53	Podolski	M	4.02	3.45
Capdevila	D	3.49	3.49	Metzelder	D	3.88	3.88
Puyol	D	3.48	3.48	Mertesacker	D	3.77	3.77
Silva	M	3.47	3.36	Ballack	M	3.73	3.38
Guiza*	F	3.46	3.02	Klose	F	3.57	3.14
Marchena	D	3.46	3.46	Hitzlsperger	M	3.57	3.57
Iniesta	M	3.45	3.45	Friedrich	D	3.47	3.47
Torres	F	3.18	2.98	Lehmann	G	3.34	3.34
Xabi Alonso*	M	3.14	3.14	Jansen*	M	3.25	3.25
Cazorla*	M	3.11	3.11	Gomez*	F	3.12	3.12
Casillas	G	3.07	3.07	Kuranyi*	F	2.99	2.99

Table 2.1: In the above table, the first 4 columns are for Spain and the remaining ones are for Germany. $\hat{c}_i$ represents $TICM^e$ and c_i represents ICM. * indicates that player is a substitute.

Table 2.2: Spain vs Russia Semi-Final Game, a=0.125 and d=0.5

Name	Position	$\hat{c}_i$	c_i	Name	Position	$\hat{c}_i$	c_i
Fabregas*	M	6.04	5.24	Zyryanov	M	4.87	4.56
Silva	M	5.50	5.35	Semak	M	4.37	4.37
Xavi	M	5.48	5.09	Zhirkov	D	4.12	4.12
Ramos	D	5.37	5.37	Anyukov	D	4.10	4.10
Iniesta	D	4.97	4.97	Arshavin	F	4.06	3.70
Senna	M	4.82	4.82	Ignashevich	D	3.84	3.84
Torres	F	4.30	4.06	Berezutski	D	3.83	3.83
Capdevila	D	4.23	4.23	Saenko	M	3.63	3.63
Xabi Alonso*	M	4.11	4.11	Semshov	M	3.57	3.57
Villa	F	4.08	3.59	Pavlyuchenko	F	3.52	3.30
Guiza*	F	4.06	3.58	Sychev*	F	3.46	3.46
Marchena	D	4.02	4.02	Akinfeev	G	3.36	3.36
Puyol	D	3.75	3.75	Biyaletdinov*	M	3.28	3.28
Casillas	G	3.69	3.69				

Table 2.2: In the above table, the first 4 columns are for Spain and the remaining ones are for Russia. $\hat{c}_i$ represents $TICM^e$ and c_i represents ICM. * indicates that player is a substitute.

Table 2.3: Germany vs Turkey Semi-Final Game, a=0.125 and d=0.5

Name	Position	$\hat{c}_i$	c_i	Name	Position	$\hat{c}_i$	c_i
Schweinsteiger	M	5.01	4.15	Ugur	M	4.95	4.23
Lahm	D	4.48	3.82	Hamit	M	4.73	4.73
Podolski	M	4.12	3.55	Semih	F	4.69	3.85
Mertesacker	D	3.85	3.85	Ayhan	M	4.65	4.65
Friedrich	D	3.75	3.75	Sabri	D	4.64	4.64
Hitzlsperger	M	3.75	3.75	Hakan	D	4.32	4.32
Frings*	M	3.68	3.68	Kazim	M	4.23	4.23
Metzelder	D	3.67	3.67	Aurelio	M	4.15	4.15
Ballack	M	3.63	3.29	Gokhan	D	4.14	4.14
Klose	F	3.51	3.07	Mehmet	M	4.11	4.11
Rolfes	M	3.22	3.22	Gokdeniz*	M	3.45	3.45
Lehmann	G	2.98	2.98	Rustu	G	3.44	3.44
Jansen*	M	2.91	2.91	Mevlut	F	3.43	3.43
				Tumer*	M	3.34	3.34

Table 2.3: In the above table, the first 4 columns are for Germany and the remaining ones are for Turkey. $\hat{c}_i$ represents $TICM^e$ and c_i represents ICM. * indicates that player is a substitute.

Table 2.4: Netherlands vs Russia Quarter Final Game, a=0.125 and d=0.5

Name	Position	$\hat{c}_i$	c_i	Name	Position	$\hat{c}_i$	c_i
Van Bronchorst	D	5.40	4.34	Arshavin	F	5.04	4.61
Van Der Vaart	M	5.08	5.08	Zhirkov	D	4.56	4.56
Nistelrooy	F	5.08	4.63	Pavlyuchenko	F	4.47	4.24
Sneijder	M	4.51	4.30	Zyryanov	M	4.45	4.16
Van Persie*	F	4.51	4.16	Semak	M	4.22	4.22
Heitinga*	D	4.43	4.43	Torbinski*	M	4.16	3.68
Boulahrouz	D	4.39	4.39	Anyukov	D	4.01	4.01
Oojer	D	4.35	4.35	Semshov	M	3.91	3.91
De Jong	M	4.24	4.24	Saenko	M	3.90	3.90
Kuyt	M	4.24	3.73	Kolodin	D	3.88	3.88
Afellay*	M	4.16	4.16	Ignashevich	D	3.76	3.76
Van Der Sarr	G	4.09	4.09	Bilyaletdinov*	M	3.68	3.68
Englaar	M	4.03	4.03	Akinfeev	G	3.52	3.52
Mathijsen	D	4.00	4.00	Sychev*	F	3.40	3.40

Table 2.4: In the above table, the first 4 columns are for Netherlands and the remaining ones are for Russia. $\hat{c}_i$ represents $TICM^e$ and c_i represents ICM. * indicates that player is a substitute.

Table 2.5: Germany vs Portugal Quarter Final Game, a=0.125 and d=0.5

Name	Position	$\hat{c}_i$	c_i	Name	Position	$\hat{c}_i$	c_i
Schweinsteiger	M	5.47	4.58	Deco	M	6.44	5.81
Podolski	M	5.20	4.55	Ronaldo	M	5.47	5.31
Ballack	M	5.10	4.69	Simao	M	5.39	5.39
Klose	F	5.01	3.96	Bosingwa	D	5.18	5.18
Lahm	D	4.95	4.25	Pepe	D	5.15	4.63
Rolfes	M	4.38	4.38	Meireles*	M	4.91	4.48
Hitzlsperger	M	4.05	4.05	Ferreira	D	4.70	4.70
Friedrich	D	3.87	3.87	Petit	M	4.51	4.51
Lehmann	G	3.60	3.60	Carvalho	D	4.39	4.39
Mertesacker	D	3.58	3.58	Nuno Gomes	F	4.39	3.98
Metzelder	D	3.57	3.57	Postiga*	F	4.26	3.85
Fritz*	M	3.39	3.39	Moutinho	M	4.02	4.02
Borowski*	M	3.32	3.32	Nani*	M	4.02	4.02
Jansen*	M	3.31	3.31	Ricardo	G	3.81	3.81

Table 2.5: In the above table, the first 4 columns are for Germany and the remaining ones are for Portugal. $\hat{c}_i$ represents $TICM^e$ and c_i represents ICM. * indicates that player is a substitute.

Table 2.6: Spain vs Italy Quarter Final Game, a=0.125 and d=0.5

Name	Position	$\hat{c}_i$	c_i	Name	Position	$\hat{c}_i$	c_i
Fabregas*	M	8.43	7.49	Grosso	D	5.74	5.74
Silva	M	8.15	7.97	De Rossi	M	5.13	5.13
Xavi	M	7.59	7.13	Panucci	D	4.99	4.29
Capdevila	D	7.53	7.53	Ambrossini	M	4.93	4.93
Senna	M	7.23	7.23	Aquilani	M	4.73	4.73
Villa	F	6.66	6.00	Zambrotta	D	4.65	4.65
Ramos	D	6.52	6.52	Camoranesi*	M	4.65	4.65
Marchena	D	6.14	6.14	Chiellini	D	4.56	4.56
Iniesta	M	5.67	5.67	Toni	F	4.38	4.38
Puyol	D	5.63	5.63	Cassano	F	4.28	4.28
Torres	F	5.56	5.29	Buffon	G	4.05	4.05
Guiza*	F	5.33	4.77	Di Natale*	F	4.05	4.05
Cazorla*	M	5.18	5.18	Perrotta	M	3.94	3.94
Casillas	G	4.72	4.72	Del Piero*	F	3.59	3.59

Table 2.6: In the above table, the first 4 columns are for Spain and the remaining ones are for Italy. $\hat{c}_i$ represents $TICM^e$ and c_i represents ICM. * indicates that player is a substitute.

Table 2.7: Croatia vs Turkey Quarter Final Game, a=0.125 and d=0.5

Name	Position	$\hat{c}_i$	c_i	Name	Position	$\hat{c}_i$	c_i
Modric	M	5.69	4.58	Arda	M	7.67	5.26
Pranjic	D	4.47	4.47	Hamit	M	5.24	5.24
Rakitic	M	4.18	4.18	Tuncay	M	5.24	5.24
Srna	M	4.06	3.83	Hakan	D	5.11	5.11
Klasnic*	F	4.05	3.31	Nihat	F	5.03	4.40
N. Kovac	M	3.97	3.97	Semih*	F	4.89	3.99
Simunic	D	3.90	3.90	Ugur*	M	4.48	3.82
Corluka	D	3.77	3.77	Sabri	D	4.24	4.24
R. Kovac	D	3.77	3.77	Gokhan	D	4.14	4.14
Kranjcar	M	3.66	3.66	Emre	D	4.13	4.13
Olic	F	3.64	3.37	Mehmet	M	4.09	4.09
Petric*	F	3.32	3.32	Kazim	M	3.91	3.91
Pletikosa	G	3.21	3.21	Rustu	G	3.69	3.69
				Gokdeniz*	M	3.43	3.43

Table 2.7: In the above table, the first 4 columns are for Croatia and the remaining ones are for Turkey. $\hat{c}_i$ represents $TICM^e$ and c_i represents ICM. * indicates that player is a substitute.

In the above tables, $\hat{\boldsymbol{c}}$ refers to $TICM^e$ and $\boldsymbol{c}$ indicates ICM of BCZ (2006). We find that the results obtained by using $TICM^e$ are generally better at capturing the players who have a direct influence on the outcome of the matches since it also incorporates the scoring probabilities. The highest value of $TICM^e$ is observed in the Spain vs Italy Quarter Final game for Fabregas of Spain who has a value of 8.43. Note that this match ended in extra time. In all of the matches which ended in normal time, the highest value of $TICM^e$ is observed in the Germany vs Portugal Quarter Final game for Deco of Portugal who has a value of 6.44. The highest ICM reported as 7.97 in Spain vs Italy match for David Silva of Spain. Note that this match ended in extra time. The next highest ICM is in Germany vs Portugal match (which ended in normal time) for Deco 5.81. Unlike the conventional belief that midfielders would always be key players due to their field position, there are examples in our data that proves otherwise. For instance, in Netherlands vs Russia match, Russian key player turns out to be a Arshavin, a forward player. Also, the major difference $TICM^e$ and ICM is that forward and midfield players appear in the higher ranks according to $TICM^e$.

Table 2.8: Spain vs Germany Final Game, $TICM$, $TICM^e$ and ICM a=0.125 and d=0.5

Name	Position	$\bar{c}_i$	$\hat{c}_i$	$\tilde{c}_i$	Name	Position	$\bar{c}_i$	$\hat{c}_i$	c_i
Xavi	M	4.42	4.31	3.97	Lahm	D	4.53	4.38	3.73
Fabregas	M	4.04	3.97	3.35	Schweinsteiger	M	4.45	4.31	3.57
Senna	M	3.78	3.66	3.66	Frings	M	4.25	4.02	4.02
Ramos	D	3.64	3.53	3.53	Podolski	F	4.15	4.01	3.45
Capdevila	D	3.60	3.49	3.49	Metzelder	D	4.07	3.88	3.88
Puyol	D	3.59	3.48	3.48	Mertesacker	D	3.95	3.77	3.77
Marchena	D	3.57	3.46	3.46	Ballack	M	3.88	3.71	3.38
Silva	M	3.57	3.47	3.36	Hitzspelger	M	3.77	3.58	3.58
Iniesta	M	3.57	3.45	3.45	Klose	F	3.71	3.59	3.13
Guiza	F	3.53	3.46	3.02	Friedrich	D	3.64	3.47	3.47
Torres	F	3.26	3.18	2.98	Lehmann	G	3.50	3.34	3.34
Xabi	M	3.24	3.14	3.14	Jansen	D	3.41	3.25	3.25
Santi	M	3.21	3.11	3.11	Gomez	F	3.27	3.12	3.12
Casillas	G	3.16	3.07	3.07	Kuranyi	F	3.12	2.99	2.99

Table 2.8: In the above table, the first 5 columns are for Spain and the remaining ones are for Germany. $\bar{c}_i$ represents $TICM$, $\hat{c}_i$ represents $TICM^e$, and c_i represents ICM. * indicates that player is a substitute.

Table 2.9: Netherlands vs Russia Quarter Final Game, $TICM$ and $TICM^e$ a=0.125 and d=0.5

Name	Position	$\bar{c}_i$	$\hat{c}_i$	$\tilde{c}_i$	Name	Position	$\bar{c}_i$	$\hat{c}_i$	c_i
Gio	D	5.50	5.32	4.34	Arshavin	F	5.16	4.98	4.61
Vaart	M	5.31	5.08	5.08	Zhirkov	D	4.68	4.56	4.59
Nistelrooy	F	5.29	5.05	4.63	Pavlyuchenko	F	4.61	4.47	4.24
Sneijder	M	4.68	4.51	4.30	Zyryanov	M	4.53	4.46	4.16
Van Persie	F	4.68	4.50	4.16	Semak	M	4.33	4.22	4.22
Boulahrouz	D	4.64	4.39	4.39	Torbinski	M	4.22	4.17	3.68
Heitinga	D	4.63	4.43	4.43	Anyukov	D	4.11	4.01	4.01
Ooijer	D	4.53	4.35	4.35	Semshov	M	4.01	3.91	3.91
De Jong	M	4.41	4.24	4.24	Saenko	M	4.01	3.90	3.90
Afellay	M	4.38	4.16	4.16	Kolodin	D	3.97	3.88	3.88
Kuyt	F	4.37	4.21	3.73	Ignashevich	D	3.85	3.76	3.76
Van der Sarr	G	4.25	4.09	4.09	Bilyaletdinov	M	3.77	3.68	3.68
Engelaar	M	4.21	4.03	4.03	Akinfeev	G	3.60	3.52	3.52
Mathijsen	D	4.19	4.00	4.00	Sychev	F	3.48	3.40	3.40

Table 2.9: In the above table, the first 5 columns are for Netherlands and the remaining ones are for Russia. $\bar{c}_i$ represents $TICM$, $\hat{c}_i$ represents $TICM^e$, and c_i represents ICM. * indicates that player is a substitute.

In Tables 2.8 and 2.9, we provide all the intercentrality measures($TICM$, $TICM^e$ and ICM) for Spain vs Germany Final and Netherlands vs Russia Quarter Final matches. The difference between $TICM$ and ICM provides the externality player i receives from his teammates. The difference between $TICM$ and ICM yields the contribution of the player to the team outcome. It can be seen in all tables that if $\delta_i = 0$, then $TICM^e$ and ICM are equal to each other. However, $TICM$ can still be different since it includes δ'_js. We can conclude that both measures are reasonably close and the main results do not change.

An important fact to mention is that since the data on scoring probabilities of players is not life time scoring probabilities, $TIMC$ and $TICM^e$ cause players who have very few shots in the tournament but scored a goal to have a high measures in some matches. Therefore, we report ICM results as a sensitivity check.[12] Some teams in our data are observed more than once and yet the key player in the same team differs in different matches. This might be due to the fluctuations in the performance of players as well as the different playing style of the players in different matches.

For the case of soccer, it is mostly the case that the head coach of the national team as the social planner would not be very interested in the reduction in aggregate Nash equilibrium. There is no guarantee that teams having more effort or playing better will get higher ranks in the tournament. Hence, we will use the externality scenario since it includes the quality of effort of the players directly and it captures the externality a player gets from the social planner's perspective, not through the player optimization through the Nash Equilibrium. We provide some calculations of $TICM$ in the next section to make some comparisons however, through the empirical part, we mainly focus on $TICM^e$ and ICM.

---

[12]According to our data there are a few players such as Van Bronchorst of Netherlands and Lahm of Germany who has only one or two shots on goal through the tournament however scoring a goal in those attempts. That makes their scoring probability relatively higher and substantially increases their $TICM^e$.

2.4.2 Sensitivity Checks

There is a concern that determination of the key player may depend on the our chosen values of the decay factor, a and discount factor, d. In fact, by means of an example BCZ (2006) show that the key player may be different for different values of a. Similarly, the key player may change depending the value of discount factor, d. Hence, in order to check the robustness of our results, we conduct a simulation analysis by changing the values of those parameters. We allow a to vary from 0 to 0.125 in increments of 0.001. Simultaneously, we use the same increment and increase the value of d from 0 to 1. Since we perform the simulations for all matches and all teams, this gives us $14 \times 125{,}000 = 1.75$ million simulations. We find that the key player identified by ICM changes about 15 percent of the simulations. On the other hand, the identified key players by using $TICM^e$ change 40 percent of the simulations. There is a greater variability in $TICM^e$ results since the scoring probability of the players are specific to the Euro 2008 tournament. Since the scoring probability itself shows great variability, it makes the $TICM^e$ measure more idiosyncratic. The passing game on the other hand is more stable and therefore the ICM results have smaller variation.

2.4.3 Key Group

In this section, we determine the key groups of players. The idea of searching for the key group was initiated by BCZ (2004). However, in this chapter we prefer to follow Temurshoev's (2008) approach for computational convenience. Key groups of players in the matches provide information about the joint performance of players in the group. This is a valuable information for the soccer clubs, managers and coaches who wish the form their teams with individuals that provide different adjacencies to their teammates.[13] In order to identify key groups of size k in a team, we take every possible combination of k players from the team and determine the reduction in the interaction matrix as well as the externalities. The key

---

[13]The identified key groups do not reveal best working individuals, but it reveals which combination of players have higher importance in the interactions and externalities.

group consists of players whose joint removal leads to the largest reduction.

We use Temurshoev's (2008) approach to compute the TICM of a group of k players. Removing players from the game causes a reduction in the interaction between players in addition to the reduction to the loss of those players ability. Therefore, we derive the group intercentrality measure for $TICM^e$ as:

$$\hat{\boldsymbol{c}}_{\boldsymbol{g}} = \mathbf{b}'\mathbf{E}(\mathbf{E}'\mathbf{ME})^{-1}\mathbf{E}'\mathbf{b} + (\mathbf{1}'\mathbf{ME})(\mathbf{E}'\boldsymbol{\delta}) \tag{2.11}$$

where $\mathbf{E}$ is the $n \times k$ matrix defined as $\mathbf{E} = (e_{i1}, ..., e_{ik})$ with e_{ir} being the i_r^{th} column of the identity matrix, k being the number of players in the group and $1 \leq k \leq n$.

The first term captures the effect of the removal of a group of players in $\mathbf{g}$ and the second term captures the effect of reduction in the desired outcome of the team. It can be readily checked that for $k = 1$, the above expression boils down to the team intercentrality measure with externality ($TICM^e$) of a player which is given in Equation (10). Note that the key group is not always comprised of the individuals having the highest intercentrality measure. As described in Borgatti (2006) and Temurshoev (2008), according to the *redundancy principle* key group involves players who provide different adjacency to their teammates.

We choose key group sizes of $k = 2$ and $k = 3$ and calculate every possible group's intercentrality measure using Equation (11). The key group results for all the countries and matches in the sample are provided in Tables 2.10-2.14. In these tables, we report the top two (the best and the next best) key groups. In the key group tables, the column player position identifies the field position of the player. These positions are D (Defense), M (Midfield) and F (Forward). The rank in the $\hat{\boldsymbol{c}}$ column identifies the player's rank according to (TICM).

Table 2.10: Key Group of Spain, $TICM^e$, a=0.125, d=0.5

Match	Group Size	Player Position	Rank in $\hat{c}$	Player Names	$\hat{c}_g$
Final	2	M, M	1,2	Xavi, Fabregas	7.78
Final	2	M, F	1,8	Xavi, Guiza	7.36
Final	3	M, M, F	1,2,8	Xavi, Fabregas, Guiza	10.58
Final	3	M, M, M	1,2,3	Xavi, Fabregas, Senna	10.49
Semi-Final	2	M, M	1,3	Fabregas, Xavi	10.48
Semi-Final	2	M, M	1,2	Fabregas, Silva	10.40
Semi-Final	3	M, M, M	1,2,3	Fabregas, Silva, Xavi	14.03
Semi-Final	3	M, M, D	1,2,4	Fabregas, Silva, Ramos	13.76
Quarter Final	2	M, M	1,3	Fabregas, Xavi	14.49
Quarter Final	2	M, M	1,2	Fabregas, Silva	14.15
Quarter Final	3	M, M, M	1,2,3	Fabregas, Silva, Xavi	18.87
Quarter Final	3	M, M, F	1,3,6	Fabregas, Xavi, Vila	18.66

Table 2.10: In the above table, key group and the second best key group results of Spain are provided for group sizes 2 and 3.

Table 2.11: Key Group of Germany, $TICM^e$, a=0.125, d=0.5

Match	Group Size	Player Position	Rank in $\hat{c}$	Player Names	$\hat{c}_g$
Final	2	D, M	1,2	Lahm, Schweinsteiger	8.28
Final	2	M, M	2,4	Schweinsteiger, Podolski	7.97
Final	3	D, M, M	1,2,4	Lahm, Schweinsteiger, Podolski	11.44
Final	3	M, M, M	2,3,4	Schweinsteiger, Frings, Podolski	11.25
Semi-Final	2	M, D	1,2	Schweinsteiger, Lahm	8.92
Semi-Final	2	M, M	1,3	Schweinsteiger, Podolski	8.52
Semi-Final	3	M, D, M	1,2,3	Schweinsteiger, Lahm, Podolski	11.93
Semi-Final	3	M, D, M	1,2,9	Schweinsteiger, Lahm, Ballack	11.69
Quarter Final	2	M, M	1,2	Schweinsteiger, Podolski	9.86
Quarter Final	2	M, F	1,4	Schweinsteiger, Klose	9.85
Quarter Final	3	M, M, F	1,2,4	Schweinsteiger, Podolski, Klose	13.71
Quarter Final	3	M, M, D	1,4,5	Schweinsteiger, Klose, Lahm	13.69

Table 2.11: In the above table, key group and the second best key group results of Germany are provided for group sizes 2 and 3.

Table 2.12: Key Group of Russia, $TICM^e$, a=0.125, d=0.5

Match	Group Size	Player Position	Rank in $\hat{c}$	Player Names	$\hat{c}_g$
Semi-Final	2	M, M	1,2	Zyryanov, Semak	8.36
Semi-Final	2	M, D	1, 4	Zyryanov, Anyukov	8.19
Semi-Final	3	M, M, F	1,2,5	Zyryanov, Semak, Arshavin	11.23
Semi-Final	3	M, D, F	1,4,5	Zyryanov, Anyukov, Arshavin	11.11
Quarter Final	2	F, M	1,4	Arshavin, Zyryanov	8.80
Quarter Final	2	F, F	1,3	Arshavin, Pavlyuchenko	8.79
Quarter Final	3	F, F, M	1,3,6	Arshavin, Pavlyuchenko, Torbinski	12.08
Quarter Final	3	F, D, F	1,2,3	Arshavin, Zhirkov, Pavlyuchenko	11.95

Table 2.12: In the above table key group and the second best key group results of Russia are provided for group sizes 2 and 3.

Table 2.13: Key Group of Turkey, $TICM^e$, a=0.125, d=0.5

Match	Group Size	Player Position	Rank in $\hat{c}$	Player Names	$\hat{c}_g$
Semi-Final	2	M, F	1,2	Ugur, Semih	9.18
Semi-Final	2	F, D	2,5	Semih, Sabri	8.96
Semi-Final	3	M, F, D	1,2,5	Ugur, Semih, Sabri	12.60
Semi-Final	3	M, F, M	1,2,3	Ugur, Semih, Hamit	12.54
Quarter Final	2	M, F	1,5	Arda, Nihat	11.89
Quarter Final	2	M, F	1, 6	Arda, Semih	11.85
Quarter Final	3	M, F, F	1,5,6	Arda, Nihat, Semih	12.31
Quarter Final	3	M, M, F	1,2,6	Arda, Hamit, Semih	12.18

Table 2.13: In the above table, key group and the second best key group results of Turkey are provided for group sizes 2 and 3.

Table 2.14: Key Group of Other Countries, $TICM^e$, a=0.125, d=0.5

Match	Group Size	Player Position	Rank in $\hat{c}$	Player Names	$\hat{c}_g$
Netherlands					
Quarter Final	2	D, F	1,3	Bronckhorst, Nistelrooy	9.81
Quarter Final	2	D, M	1,2	Bronckhorst, Vaart	9.64
Quarter Final	3	D, M, F	1,2,3	Bronckhorst, Vaart, Nistelrooy	13.27
Quarter Final	3	D, F, F	1,3,5	Bronckhorst, Nistelrooy, Persie	13.15
Portugal					
Quarter Final	2	M, D	1,5	Deco, Pepe	10.73
Quarter Final	2	M, M	1,2	Deco, Ronaldo	10.60
Quarter Final	3	M, M, D	1,2,5	Deco, Ronaldo, Pepe	14.35
Quarter Final	3	M, D, M	1,5,6	Deco, Pepe, Meireles	14.15
Italy					
Quarter Final	2	D, D	1,3	Grosso, Panucci	9.91
Quarter Final	2	D, M	1,2	Grosso, De Rossi	9.78
Quarter Final	3	D, M, D	1,2,3	Grosso, De Rossi, Panucci	13.47
Quarter Final	3	D, D, D	1,3,7	Grosso, Panucci, Zambrotta	13.25
Croatia					
Quarter Final	2	F, M	1,5	Modric, Klasnic	9.21
Quarter Final	2	D, M	1,2	Modric, Pranjic	9.17
Quarter Final	3	D, F, M	1,2,5	Modric, Pranjic, Klasnic	12.31
Quarter Final	3	D, D, M	1,3,5	Modric, Rakitic, Klasnic	12.18

Table 2.14: In the above table, key group and the second best key group results of Netherlands, Portugal, Italy and Croatia are provided for group sizes 2 and 3.

For an interesting comparison, we also provide the ICM key group results of Spain in Table 2.15.[14] Generally, the key groups obtained by using $TICM^e$ include more forwards and midfielders. According to Table 15, there are no forward players and several defenders in key groups. However, according to Table 2.10, key groups according to $TICM^e$ have some forwards and more midfielders.

Table 2.15: Key Group of Spain, ICM, a=0.125, d=0.5

Match	Group Size	Player Position	Rank in $\tilde{c}$	Player Names	$\tilde{c}_g$
Final	2	M, M	1, 2	Xavi, Senna	7.02
Final	2	M, D	1, 3	Xavi, Ramos	6.97
Final	3	M, D, D	1, 3, 4	Xavi, Ramos, Capdevila	9.61
Final	3	M, D, D	1, 3, 5	Xavi, Ramos, Puyol	9.60
Semi-Final	2	D, M	1,2	Ramos, Silva	9.52
Semi-Final	2	M, M	2,3	Silva, Fabregas	9.45
Semi-Final	3	M, D, M	5,1,3	Iniesta, Ramos, Fabregas	12.82
Semi-Final	3	M, M, M	5,4,8	Iniesta, Xavi, Xabi	12.73
Quarter Final	2	D, M	2,1	Capdevila, Silva	13.51
Quarter Final	2	M, M	4,1	Senna, Silva	13.13
Quarter Final	3	D, M, M	2,4,1	Capdevila, Senna, Silva	17.48
Quarter Final	3	D, M, M	2,4,3	Capdevila, Senna, Fabregas	17.39

Table 2.15: In the above table, key group and the second best key group results of Spain are provided for group sizes 2 and 3 by using the BCZ (2006) ICM key group measure.

2.4.4 Player Ratings, Market Value and (Team) Intercentrality

In this subsection, we discuss the effect of the ICM and $TICM^e$ on player ratings and market values of the players in our sample. The *transfermarkt.de* website provides information about other observable characteristics of the players such as: Date of birth, club, nation, position, and number of international appearances, number of international goals, preferred foot and captaincy. We use the Club UEFA points and Nation UEFA points which are available from UEFA's website in order to capture quality and reputation of the players. Club and Nation points are announced by UEFA yearly. These points are earned for being successful in UEFA club or national tournaments. The points that are provided by UEFA for the

---

[14]The key groups according to ICM for other countries are available upon request.

year 2008 are composed of the points earned in 2003-2008 period. We merge the available data from *transfermarkt.de* with the (team) intercentrality measure, Club and Nation Rank measured by the UEFA points in 2008. The descriptive statistics about the data set are provided in Table 16.

Table 2.16: Descriptive Statistics

Variable	Obs	Mean	Std. Dev	Min	Max
Average Rating	113	6.17	0.85	3.83	8.17
ICM	113	4.21	0.61	2.99	5.81
$TICM^e$	113	4.38	0.74	2.99	7.67
Log Market Value	112	2.25	0.84	-0.22	4.32
Age	113	28.29	3.55	22	36
Height(cm)	112	181.27	6.85	168	198
Club UEFA Points	112	66.95	33.33	0	124.99
Nation UEFA Points	112	44.27	17.05	11.62	75.27
International Caps	112	46.52	21.53	10	98
International Goals	112	7.73	9.56	0.00	48
Captaincy	112	0.16	0.37	0.00	1.00
Right-footed	112	0.57	0.50	0.00	1.00
Left-footed	112	0.20	0.40	0.00	1.00
Both	112	0.23	0.42	0.00	1.00
Defender	113	0.27	0.45	0.00	1.00
Midfielder	113	0.47	0.50	0.00	1.00
Forward	113	0.25	0.43	0.00	1.00
Expiring	112	0.14	0.35	0.00	1.00
Move	112	0.38	0.49	0.00	1.00

Table 2.16: Average ratings for the players are obtained by taking the average of the player ratings available through Goal.com ESPN Soccer and Skysports. Log of the market value of players are obtained from *transfermarkt.de* along with the player characteristics. Club and Nation UEFA points are available from UEFA's website. We use 2007-2008 points, which is earned in 2003-2008 period by clubs or nations in UEFA tournaments. Captaincy is equal to 1 if the player is a captain in his club or his national team and 0 oth. D, M and F are dummy variables to indicate the field position of the players. They represent Defender, Midfielder and Forward positions respectively. Goalkeepers are excluded from the sample. Expiring is equal to 1 if the players current contract with his club is expiring at the end of 2009-2010 season 0 oth. Move is equal to 1 if the player transferred to another after Euro 2008 and 0 oth.

2.4.4.1 Player Ratings

We consider player ratings by experts after each game to show that the individual performance of the players can be explained by the (team) intercentrality measures. We obtain player ratings from three sources: Goal.com, ESPN and SkySports. We create a variable called rating for each player which is obtained by taking of the average of these ratings.[15] These sources are used since they use the same scale and also provide ratings for the substitute and substituted players in the matches. Also, these sources are outside the competing countries in UEFA Euro 2008 which eliminates potential country bias in the ratings.

In order to analyze the relationship between player ratings and ICM and $TICM^e$, we consider the following base model:

$$\text{Rating}_{it} = \alpha_1 + \beta_1 \ ICM_{it}(TICM_{it}^e) + \gamma_1 \ Age_i + \theta_1 \ Age_i^2 + \lambda_1 \ Position_i$$
$$+ \psi_1 \ ClubRank_i + \phi_1 \ NationRank_i + \epsilon_{it}.$$

In the above regression model, the i subscript represents the player i and the t subscript represents the match t. Rating is the dependent variable and represents for the average of the player ratings obtained from the three sources. ICM stands for the intercentrality measure of BCZ (2006). $TICM^e$ represents the team intercentrality measure with externalities from Equation (10). Position is a dummy variable that identifies the field position of the player. We consider three different field positions: Defense (D), Midfield (M) and Forward(F).[16]

The estimation results for the relationship between average player ratings, ICM and $TICM^e$ are provided in Tables 2.17 and 2.18. In the pooled OLS estimation, we estimate a linear regression model where the time variable, t which is used to identify each match.[17]

Table 2.18 reports the results from a GLS estimation with bootstrapped robust standard

---

[15]The correlation coefficient of ratings from the above sources are 0.7 thus we prefer to take the average of these ratings rather than using them one by one.

[16]Goalkeepers are excluded from the regression analysis. Niko Kovac (Croatia) retired from professional soccer before 2010 and are also excluded.

[17]Ideally, we would prefer to run a random effects model, but there is not enough idiosyncratic variance in the data to allow for this.

errors. We take the average of the ratings and (team) intercentrality measures and have only one observation for each player and we report robust standard errors. Both Tables 2.17 and 2.18 show that there is a strong relationship between the $TICM^e$ and the average ratings. Specifically, players who have higher $TICM^e$ performed better than their teammates according to the experts. In the pooled OLS estimation, the estimated coefficient for the $TICM^e$ is significant at 5 % significance level while the coefficient of ICM is significant at 10 % significance level. In the GLS estimation, the estimated coefficient of $TICM^e$ is significant at 1 % significance level while the estimated coefficient of ICM is significant at 5 % significance level. As a sensitivity check, we only include the players who played longer than 30 minutes in the matches. This reduces the number of observations by 30, but the results are robust. Another important factor to control for is whether or not the match ended in normal time. We define a dummy variable ET which is equal to 1 if the match ended in extra time and 0 otherwise. With the inclusion of this variable, the estimated coefficient of $TICM^e$ is significant at 5 % significance level whereas the estimated coefficient of ICM is not statistically different from zero. Therefore, we conclude that $TICM^e$ and ICM explain the expert ratings.

In order to investigate whether the experts regard different importance to the ICM and $TICM^e$ according to their field position of the players, we interact the ICM and $TICM^e$ of players with their position dummies. None of the estimated coefficients are statistically significant. Therefore, we conclude that the ICM and $TICM^e$ are equally important regardless of the field positions of the players. (i.e, the effect of the intercentrality measures on ratings is homogenous in the sample with respect to players' positions on the field.) In addition to the control variables in base model, we run regressions with a broader set of control variables including international appearances, international goals, captaincy, height and preferred foot. The estimated coefficients and their significance are very similar.

Table 2.17: Average Ratings, ICM and $TICM^e$ Pooled OLS Estimation

Variable	I	II	III	IV	V	VI
ICM	0.254***	0.248***	0.226			
	(0.093)	(0.094)	(0.146)			
$TICM^e$				0.308***	0.296***	0.262**
				(0.090)	(0.086)	(0.119)
ET	0.327*	0.375*	0.361*	0.303	0.345*	0.327
	(0.189)	(0.209)	(0.217)	(0.193)	(0.207)	(0.217)
Age	-0.198	-0.104	-0.135	-0.127	-0.036	-0.108
	(0.262)	(0.273)	(0.267)	(0.268)	(0.271)	(0.277)
Age Squared	0.003	0.001	0.002	0.002	0.000	0.002
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Club UEFA pts	0.001	0.001	0.001	0.000	0.000	0.000
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Nation UEFA pts	0.007*	0.008*	0.008*	0.008*	0.008*	0.008*
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
D		-0.272*	-0.119		-0.233	0.052
		(0.149)	(0.935)		(0.145)	(0.836)
F		-0.066	-0.987		-0.099	-1.483
		(0.189)	(1.225)		(0.184)	(1.122)
DxI			-0.035			
			(0.213)			
FxI			0.236			
			(0.305)			
DxTI						-0.065
						(0.184)
FxTI						0.328
						(0.264)
Constant	7.886**	6.576*	7.122*	6.602*	5.374	6.564
	(3.807)	(3.928)	(3.864)	(3.895)	(3.957)	(4.083)
Observations	178	178	178	178	178	178
Number of Players	111	111	111	111	111	111
R-squared	0.119	0.131	0.137	0.153	0.162	0.176
Wald Chi Sq statistic	18.96	18.59	18.86	23.37	24.77	24.12

Table 2.17: The dependent variable is average ratings and the pooled OLS coefficients are reported in the above regressions. Bootstrapped robust standard errors are given in parentheses. ***, **, * indicate 1, 5 and 10 percent significance levels respectively. ET is a dummy variable which takes the value of 1 if the player played more than 90 minutes in any of the matches and 0 otherwise. Goalkeepers are excluded from the sample. DxICM, FxICM, Dx$TICM^e$ and Fx$TICM^e$ are interaction variables obtained by interacting the (team) intercentrality measure with the position dummy.

Table 2.18: Average Ratings and (Team) Intercentrality GLS Estimation

Variable	I	II	III	IV	V	VI
(ICM)	0.557***	0.564***	0.495**			
	(0.155)	(0.154)	(0.197)			
$(TICM^e)$				0.541***	0.534***	0.434***
				(0.146)	(0.135)	(0.165)
Age	-0.070	-0.066	-0.084	-0.034	0.058	-0.029
	(0.301)	(0.314)	(0.323)	(0.299)	(0.322)	(0.307)
Age Squared	0.001	0.001	0.001	-0.001	-0.001	0.000
	(0.005)	(0.005)	(0.006)	(0.005)	(0.006)	(0.005)
Club UEFA pts	-0.003	-0.003	-0.002	-0.004	-0.004	-0.004
	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)
Nation UEFA pts	-0.001	0.001	0.001	0.001	0.001	0.001
	(0.005)	(0.006)	(0.007)	(0.005)	(0.005)	(0.005)
Defender		-0.066	-0.375		-0.001	-0.439
		(0.166)	(1.158)		(0.175)	(1.024)
Forward		0.015*	-1.067		-0.065	-2.189*
		(0.212)	(1.675)		(0.199)	(1.247)
DxICM			0.072			
			(0.280)			
FxICM			0.272			
			(0.418)			
Dx$TICM^e$						0.098
						(0.237)
Fx$TICM^e$						0.504*
						(0.296)
Constant	5.399	5.305	5.875	3.797	3.510	5.253
	(4.431)	(4.626)	(4.889)	(4.342)	(4.647)	(4.530)
Observations	112	112	112	112	112	112
R square	0.147	0.148	0.153	0.206	0.207	0.231
R square adj	0.106	0.091	0.078	0.169	0.154	0.163
Wald Chi Sq Statistic	14.29	14.83	18.18	15.51	16.76	26.63

Table 2.18: The dependent variable is average ratings and the GLS estimation results are reported in the above regressions. Bootstrapped robust standard errors with 1000 replications are given in parentheses. ***, **, * indicate 1, 5 and 10 percent significance levels respectively. Goalkeepers are excluded from the sample. DxICM, FxICM, Dx$TICM^e$ and Fx$TICM^e$ are interaction variables obtained by interacting the (team) intercentrality measure with the position dummy.

2.4.4.2 Market Values

Next, we investigate whether having a higher ICM or $TICM^e$ in Euro 2008 affects the market values of the players. Investigating the effect of intercentrality on salaries would be more interesting, but the club salaries of soccer players in Europe are not publicly available. Hence, we consider market values instead of salaries. Frick (2007) and Battre et al. (2008) regard the estimated market value of the soccer players obtained from http://www.transfermarkt.de [18] as a good and reliable source to proxy the undisclosed salary of players. Battre et al.(2008) points out that there is a strong relationship between the market value of the players and their salaries for the players in Bundesliga, German First Division.[19] So, the estimated market value of the players are obtained for the year 2010 may be regarded as a proxy for the salaries of the players in our sample.

We use the following base model to investigate the relationship between the market values of soccer players and their ICM or $TICM^e$:

$$
\begin{aligned}
LogMV_i \; = \;\; & \alpha_2 + \; \beta_2 \; ICM_i(TICM_i^e) + \; \gamma_2 \; Age_i + \; \theta_2 \; Age_i^2 + \; \lambda_2 \; Position_i \\
& + \; \psi_2 \; ClubRank_i + \; \phi_2 \; NationRank_i + \; u_i.
\end{aligned}
$$

In the above model, $LogMV$ is the dependent variable obtained from *transfermarkt.de* and represents the log of the market value of the players in million euros. Another important factor affecting the market values of players might be the contract length of the players because of the Bosman Rules in European football. It is likely that players whose contracts are about to expire have lower market values.[20] Using the contract duration information

---

[18]transfermarkt.de does not allow user to track the past market values. We saved the data about the players at March, 19 2010.

[19]Battre et al.(2008) obtains estimated market values of soccer players from a German sports magazine Kicker. However, Kicker only provides the market values of the players who only play at Bundesliga. They conduct a sensitivity check with transfermarkt.de data and they state that the correlation between those two sources are high.

[20]Bosman Rules is an important factor affecting the free movement of labor and had a profound effect on the transfers of football players within the European Union (EU). It allows professional football players in the EU to move freely to another club at the end of their contract with their present team.

available from *transfermarkt.de* we identify the players whose contracts' are expiring at the end of 2009-2010 season. Inclusion or exclusion of those players do not affect our results. However, we find evidence that players whose contracts are expiring in 2009-2010 season have lower market values. Another factor to control for is whether or not the player transferred to another club between 2008 and 2010. We define a dummy variable called move and it is equal to 1 if the player has complected a transfer and 0 otherwise. The estimated coefficient and significance of ICM and $TICM^e$ is robust.

Some players are observed more than once in the tournament and they have different average ratings and (team) intercentrality measures in different matches. However, we have only one observation for the market value of the players and the other control variables are time independent with the current setup. Thus, the above model cannot be estimated by panel data methods. In order to deal with this issue, we take the average of the (team) intercentrality measures and use GLS estimation with bootstrapped robust standard errors. We also run sensitivity regressions with clustered errors according to players, the results are very similar.

The estimation results investigating the relationship between the estimated market value and (team) intercentrality measures in Euro 2008 is provided in Table 2.19. We again report the results for both ICM and $TICM^e$. The standard errors are bootstrapped with 1000 replications.[21]

---

[21]Since we have only one market value observation for the players, we lose significant amount of observations. To deal with this issue, we bootstrap the standard errors.

Table 2.19: Market Values and (Team) Intercentrality GLS Estimation

Variable	I	II	III	IV	V	VI
ICM	0.256**	0.311***	0.458***			
	(0.104)	(0.104)	(0.137)			
$TICM^e$				0.240***	0.247***	0.333***
				(0.082)	(0.079)	(0.103)
Age	0.430*	0.398	0.408*	0.474*	0.452*	0.505**
	(0.256)	(0.244)	(0.228)	(0.259)	(0.248)	(0.247)
Age squared	-0.010**	-0.009**	-0.009**	-0.010**	-0.010**	-0.011**
	(0.004)	(0.004)	(0.004)	(0.005)	(0.004)	(0.004)
Club UEFA pts	0.005**	0.005**	0.005**	0.005**	0.005**	0.006**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Nation UEFA pts	0.013***	0.012***	0.013***	0.014***	0.014***	0.014***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Defender		-0.187	1.254		-0.166	0.597
		(0.135)	(1.028)		(0.133)	(0.992)
Forward		0.181	1.746		0.121	1.629*
		(0.153)	(1.053)		(0.147)	(0.931)
DxICM			-0.335			
			(0.233)			
FxICM			-0.387			
			(0.258)			
Dx$TICM^e$						-0.174
						(0.221)
Fx$TICM^e$						-0.357*
						(0.216)
Constant	-4.183	-3.982	-4.807	-4.825	-4.583	-5.779
	(3.739)	(3.520)	(3.317)	(3.750)	(3.604)	(3.570)
Observations	112	112	112	112	112	112
R square	0.512	0.536	0.551	0.523	0.509	0.552
R square adj	0.489	0.505	0.512	0.500	0.507	0.512
Wald Chi Sq Statistic	120.55	137.33	139.61	124.69	139.67	148.88

Table 2.19: The dependent variable is the natural log of 2010 market value of the players obtained from transfermarkt.de. Goalkeepers are excluded from the sample. DxICM, FxICM, Dx$TICM^e$ and Fx$TICM^e$ are variables obtained by interacting the (team) intercentrality measures with the position dummy of the player. Bootstrapped robust standard errors with 1000 replications are reported in parentheses. ***, **, * indicate 1, 5 and 10 percent significance levels respectively.

According to the estimation results, intercentrality measure in UEFA Euro 2008 explains the 2010 market values of the players. One standard deviation increase in ICM creates on the average 15.62 percent increase in the market values of the players. On the other hand, one standard deviation increase in $TICM^e$ yields on the average 18.22 percent increase in the market values of the players. It might be the case that, intercentrality measures are important for only a certain group of players who play in a certain position of the field (say midfielders). To test this hypothesis, we interact the ICM and $TICM^e$ of players with their position dummies. The findings suggest that ICM and $TICM^e$ are equally important at 5 % significance level. (i.e, the effect of the intercentrality measures on the market values is homogenous in the sample with respect to players' positions on the field.) In addition to the control variables in the above model, we regress the same dependent variable on a broader set of control variables including national team dummies, international appearances, international goals, captaincy, height and preferred foot. The estimates are close and the coefficient of ICM and $TICM^e$ variables are still significant. Since we have a small sample size, we prefer to use and report the results for the base models.

Note that the regression models use the intercentrality measures (ICM and $TICM^e$) which are calculated for specific parameters of $a = 0.125$ and $d = 0.5$. As a sensitivity check, we calculated those intercentrality measures for the parameter sets $a = 0.1$ and $d = 0.4, 0.5, 0.6$ and $a = 0.125$ and $d = 0.4, 0.6$. The estimated coefficients and their significance are very similar.[22]

2.5 Conclusion and Discussion

In this chapter, we introduce a Team Game and develop a measure for identifying the key player in teams. Our work extends the intercentrality measure of BCZ (2006) to include an additional term which captures the team outcome expression in the utility functions of play-

---

[22]We do not report the estimates obtained by using the above parameters but they are available upon request.

ers. This term suggests that a player gets utility when her team achieves its desired outcome. To identify the contribution of players to their teammates, we develop two intercentrality measures which derives from possible considerations of the social planner. $TICM$ considers the effect of a player's removal on the aggregate Nash Equilibrium effort levels. $TICM^e$ identifies the externality each player gets from her teammates and weights it according to the ability of the player.

Our measures also have some common features with intercentrality measure (ICM) of BCZ (2006). We can say that a key player does not need to have the highest amount of individual payoff. In addition, a key player does not need to have the highest amount of individual action. It is important to note that both BCZ (2006) and our framework are not seeking the best players in the network. The identified key players and key groups are the ones that have the highest contribution to the corresponding aggregate Nash equilibrium effort levels or according to the externality scenario the key players are the ones who get the highest amount of externality from their teammates which is weighted by the ability parameter of each player.

In the empirical part of the chapter, first we illustrate how to utilize the intercentrality measures. Then, we show that there is a positive relationship between the average ratings and $TICM^e$ and ICM in the sample. This fact reflects that soccer players having more interactions with their teammates get more credit in performance by the experts. Moreover, the market value of the soccer players increase with both $TICM^e$ and ICM which is assumed to be reflected in their salaries. This effect is homogenous in the sample, it doesn't depend on the position of the player on the field.

One interesting extension of the approach in the chapter might be considering the effort variable to be a vector and allowing different types of individual actions. This will require a new set of theocratical results. Depending on the availability of data this model then can be empirically tested. In soccer, for instance one could include distance traveled, tackling and dribbling data. Given the relationship between passing and scoring opportunities, this

way will not alter our primary results, but will provide us a more precise way to identify key players and key groups.

An interesting extension to our model would be to investigate key player problem as a network design game. The planner is the head coaches who have to announce the national squads. There are qualities, δ_i's and possible interaction possibilities between players. This can be modeled as an expected utility maximization problem with a two stage team game. At the first stage, squads are announced and at the second stage players optimize their effort with given interactions.

2.6 Appendix

Proof of **Theorem 2.1**

The condition for a well defined interior Nash equilibrium of the Team Game is that the $[\beta\mathbf{I} - \lambda\mathbf{G}]^{-1}$ matrix must be invertible. We can rewrite the $[\beta\mathbf{I} - \lambda\mathbf{G}]^{-1}$ matrix as

$$\lambda[\frac{\beta}{\lambda}\mathbf{I} - \mathbf{G}]^{-1}$$

Let $(\rho_1(\mathbf{G}))$ be the spectral radius of $\mathbf{G}$ matrix.[23] Then, $\beta > \lambda(\rho_1(\mathbf{G}))$ ensures that Equation (9) is invertible by Theorem III of Debreu and Herstein (1953, pg.601). Once the condition is verified, an interior Nash equilibrium in pure strategies $\mathbf{x}^* \in R_+^n$ satisfies:

$$\frac{\partial U_i}{\partial x_i}(x_i^*) = 0 \quad \text{and } x_i^* > 0 \quad \text{for all i=1, 2,...,n}$$

---

[23]Spectral radius of $\mathbf{G}$ matrix is defined as the inverse of the norm of the highest eigenvalue of $\mathbf{G}$ matrix.

Hence, maximizing U_i with respect to x_i yields:

$$\frac{\partial U_i}{\partial x_i} = \alpha_i + \sigma_{ii}x_i + \sum_{j \neq i} \sigma_{ij}x_j + \theta\delta_i = \mathbf{0}$$

$$= \alpha - \beta x_i - \gamma \sum_{j \neq i}^{n} x_j + \lambda \sum_{i=1}^{n} g_{ij}x_j + \theta\delta_i$$

In vector notation:

$$\frac{\partial \mathbf{U}}{\partial \mathbf{x}} = \alpha\mathbf{1} - (\beta\mathbf{I} - \gamma\mathbf{U} - \lambda\mathbf{G})\mathbf{x} + \theta\boldsymbol{\delta} = \mathbf{0} \tag{2.12}$$

The above equation can be rewritten as:

$$\beta(\mathbf{I} - \lambda/\beta\mathbf{G})\mathbf{x}^* = \alpha\mathbf{1} - \gamma\mathbf{U}\mathbf{x}^* + \theta\boldsymbol{\delta}$$

Let $\mathbf{x}^*(\boldsymbol{\Sigma})$ be the solution to the above equation. By using $\mathbf{U}\mathbf{x}^* = \hat{x}^*\mathbf{1}$ where $\hat{x}^* = \sum_{i=1}^{n} x_i^*$ and rearranging terms we obtain:

$$\beta(\mathbf{I} - \lambda^*\mathbf{G})\mathbf{x}^* = (\alpha - \gamma\hat{x}^*)\mathbf{1} + \theta\boldsymbol{\delta}$$

Multiplying both sides with $(\mathbf{I} - \lambda^*\mathbf{G})^{-1}$ yields:

$$\beta\hat{x}^* = (\alpha - \gamma\hat{x}^*)(\mathbf{I} - \lambda^*\mathbf{G})^{-1}\mathbf{1} + \theta(\mathbf{I} - \lambda^*\mathbf{G})^{-1}\boldsymbol{\delta}$$

By using the definitions $\mathbf{b}(\mathbf{g}, \lambda^*) = (\mathbf{I} - \lambda^*\mathbf{G})^{-1}\mathbf{1}$ and $\mathbf{b}_\delta(\mathbf{g}, \lambda^*) = (\mathbf{I} - \lambda^*\mathbf{G})^{-1}\boldsymbol{\delta}$ the above expression becomes:

$$\beta hat x^* = (\alpha - \gamma\hat{x}^*)\mathbf{b}(\mathbf{g}, \lambda^*) + \theta\mathbf{b}_\delta(\mathbf{g}, \lambda^*)$$

$$= \alpha\mathbf{b}(\mathbf{g}, \lambda^*) - \gamma\hat{x}^*\mathbf{b}(\mathbf{g}, \lambda^*) + \theta\mathbf{b}_\delta(\mathbf{g}, \lambda^*)$$

which is equivalent to:

$$\hat{x}^* = \alpha \mathbf{b}(\mathbf{g}, \lambda^*) - \gamma \mathbf{x}^* \hat{b}(\mathbf{g}, \lambda^*) + \theta \mathbf{b}_\delta(\mathbf{g}, \lambda^*)$$

where $\hat{b}(\mathbf{g}, \lambda^*) = \sum_{i=1}^n b_i(\mathbf{g}, \lambda^*)$. By rearranging terms we get the following[24]:

$$\hat{x}^*(\mathbf{\Sigma}) = \frac{\alpha \mathbf{b}(\mathbf{g}, \lambda^*) + \theta \mathbf{b}_\delta(\mathbf{g}, \lambda^*)}{\beta + \gamma \hat{b}(\mathbf{g}, \lambda^*)}$$

Given that $\alpha + \theta \delta > 0$ and $b_i(\mathbf{g}, \lambda^*) + b_\delta(\mathbf{g}, \lambda^*) \geq 1$ for all $i = 1, ..., n$, there is only one critical point and $\frac{\partial^2 U_i}{\partial x_i^2} = \sigma_{ii} < 0$ is always concave. This argument ensures that x^* is interior. Now, we establish uniqueness by dealing with the corner solutions.

Let $\beta(\mathbf{\Sigma}), \gamma(\mathbf{\Sigma}), \lambda(\mathbf{\Sigma})$ and $\mathbf{G}(\mathbf{\Sigma})$ be the elements of the decomposition of $\mathbf{\Sigma}$. For all matrices $\mathbf{Y}$, vector $\mathbf{y}$ and set $S \subset 1, 2, ..., n$, $\mathbf{Y}_s$ is a submatrix of Y with s rows and columns and $\mathbf{y}_s$ is the subvector of $\mathbf{y}$ with rows in s. Then, $\gamma(\mathbf{\Sigma_s}) \leq \gamma(\mathbf{\Sigma})$, $\beta(\mathbf{\Sigma_s}) \geq \beta(\mathbf{\Sigma})$ and $\lambda(\mathbf{\Sigma_s}) \leq \lambda(\mathbf{\Sigma})$. Also, $\lambda(\mathbf{G}) = \mathbf{\Sigma} + \gamma(\mathbf{U} - \mathbf{I}) - \sigma_{ii}\mathbf{I} - \theta_z Z$ and the coefficients in $\lambda \mathbf{G}$ (s rows and columns) are at least as high as the coefficients in $\lambda(\mathbf{\Sigma_s})\mathbf{G}_s$. From Theorem I of Debreu and Herstein (1953, pg.600), $\rho_1(\lambda(\mathbf{\Sigma_s})\mathbf{G}_s) \leq \rho_1(\lambda(\mathbf{\Sigma})\mathbf{G})$. Therefore, $\beta(\mathbf{\Sigma}) > \lambda(\mathbf{\Sigma})\rho_1(\mathbf{G})$ implies that $\beta(\mathbf{\Sigma_s}) > \lambda(\mathbf{\Sigma_s})\rho_1(\mathbf{G}_s)$.

Let $\boldsymbol{y}^*$ be a non interior Nash equilibrium of the Team Game. Let $S \subset 1, 2, ..., n$ such that $y_i^* = 0$ if and only if $i \in N \setminus S$. Thus, $y_i^* > 0$ for all $i \in S$.

$$\frac{\partial U_i}{\partial x_i} = \alpha - \beta x_i - \gamma \sum_{j \neq i}^n x_j + \lambda \sum_{i=1}^n g_{ij} x_j + \theta \delta_i$$

$$\frac{\partial U_i}{\partial x_i}(0) = \alpha_i + \theta \delta_i$$

---

[24]The last step of writing the Nash Equilibrium follows from the simple algebra stated in BCZ (2006) page 1414.

and 0 cannot be a Nash Equilibrium. Then,

$$-\mathbf{\Sigma}_s \boldsymbol{y}_s^* = (\beta\mathbf{I}_s + \gamma\mathbf{U}_s - \lambda\mathbf{G}_s)\boldsymbol{y}_s^* = \boldsymbol{\alpha} + \theta\boldsymbol{\delta}$$

$$\beta\boldsymbol{y}_s^* + \gamma\mathbf{U}_s\boldsymbol{y}_s^* - \lambda\mathbf{G}_s\boldsymbol{y}_s^* = \alpha + \theta\boldsymbol{\delta}_s$$

$$\beta[\mathbf{I}_s - \lambda^*\mathbf{G}_s]\boldsymbol{y}_s^* = \alpha + \theta\boldsymbol{\delta}_s - \gamma\hat{y}_s^* \cdot \mathbf{1}_s$$

where the last step utilizes $\mathbf{U}_s \boldsymbol{y}_s^* = \hat{y}_s^* \cdot \mathbf{1}_s$ and $\lambda^* = \lambda/\beta$. Pre-multiplying both sides by $[\mathbf{I}_s - \lambda^*\mathbf{G}_s]^{-1}$ yields:

$$\beta\boldsymbol{y}_s^* = [\mathbf{I} - \lambda^*\mathbf{G}_s]^{-1}\alpha + \theta[\mathbf{I} - \lambda^*\mathbf{G}_s]^{-1}\boldsymbol{\delta}_s - \gamma\hat{y}_s^*[\mathbf{I}_s - \lambda^*\mathbf{G}_s]^{-1} \cdot \mathbf{1}_s \qquad (2.13)$$

$$\boldsymbol{y}_s^* = \frac{(\alpha - \gamma\hat{y}_s^*)\mathbf{b}_s(\mathbf{g}, \lambda^*) + \theta\mathbf{b}_\delta^s(\mathbf{g}, \lambda^*)}{\beta} \qquad (2.14)$$

Every player $i \in N \setminus S$ is best responding with $y_i^* = 0$ so that y_j^* is the action of the subset S of players.

$$\frac{\partial U_i}{\partial x_i}(y_i^*) = \alpha - \sum_{j\in S}\sigma_{ij}y_j^* + \theta\delta_i$$

$$\frac{\partial U_i}{\partial x_i}(y_i^*) = \alpha - \gamma\hat{y}_s^* + \lambda\sum_{j\in S}g_{ij}y_j^* + \theta\delta_i \leq 0$$

for all $i \in N \setminus S$. Now substitute y_s^* instead of y_j^* in the above equation:

$$\frac{\partial U_i}{\partial x_i}(\mathbf{y}^*) = \alpha - \gamma\hat{y}_s^* + \lambda\sum_{j\in S}g_{ij}\big(\frac{(\alpha - \gamma\hat{y}_s^*)b_j(\mathbf{g}, \lambda^*) + \theta\mathbf{b}_\delta^j(\mathbf{g}, \lambda^*)}{\beta}\big) \leq 0$$

$$\frac{\partial U_i}{\partial x_i}(\mathbf{y}^*) = (\alpha - \gamma\hat{y}_s^*)[1 + \lambda^*\sum_{j\in S}g_{ij}b_j(\mathbf{g}, \lambda^*)] + \theta\lambda^*\sum_{j\in S}\mathbf{b}_\delta^j(\mathbf{g}, \lambda^*) \leq 0$$

If $\theta \leq |\alpha - \gamma\hat{y}_s^*|$ then $y_i^* \leq 0$ using Equations (10) and (11), which is a contradiction. Note that to reestablish uniqueness θ has to be small enough such that $\theta \leq |\alpha - \gamma\hat{y}_s^*|$.

Proof of **Proposition 1.a**:

Note that equation (9) still holds for this case. $\mathbf{\Sigma}$ matrix is substituted in equation (10) to

obtain:

$$(\beta \mathbf{I} + \gamma \mathbf{U} - \lambda \mathbf{G}) x^* = \boldsymbol{\alpha} + \theta \boldsymbol{\delta}$$

where $\boldsymbol{\alpha}$ is now a $n \times 1$ column vector and its elements shows the returns to individual actions. Now, substitute $\mathbf{x} \cdot 1$ instead of $\mathbf{U} \cdot x^*$:

$$[\beta \mathbf{I} - \lambda \mathbf{G}] x^* = \boldsymbol{\alpha} - \gamma x^* \cdot \mathbf{1} + \theta \boldsymbol{\delta} \longrightarrow \beta [\mathbf{I} - \lambda^* \mathbf{G}] x^* = \boldsymbol{\alpha} - \gamma x^* \cdot \mathbf{1} + \theta \boldsymbol{\delta}$$

And pre-multiply both sides by $[\mathbf{I} - \lambda^* \mathbf{G}]^{-1}$ matrix to obtain:

$$\begin{aligned}
\beta x^* &= [\mathbf{I} - \lambda^* \mathbf{G}]^{-1} (\boldsymbol{\alpha} + \theta \boldsymbol{\delta}) - \gamma x^* [\mathbf{I} - \lambda^* \mathbf{G}]^{-1} \cdot \mathbf{1} \\
\beta x^* &= \mathbf{b}_\alpha(\mathbf{g}, \lambda^*) + \theta \mathbf{b}_\delta + \gamma x^* \mathbf{b}(\mathbf{g}, \lambda^*) \\
x^*(\Sigma) &= \frac{\mathbf{b}_\alpha(\mathbf{g}, \lambda^*) + \theta \mathbf{b}_\delta(\mathbf{g}, \lambda^*)}{\beta + \gamma \mathbf{b}(\mathbf{g}, \lambda^*)}
\end{aligned}$$

Proof of **Proposition 1.b**:

Define:

$$\widetilde{\alpha}_i = \frac{\alpha_i}{\sigma_{ii}} \quad , \quad \widetilde{\sigma_{ij}} = \frac{\sigma_{ij}}{\sigma_{ii}} \quad , \quad \widetilde{\delta}_i = \frac{\delta_i}{\sigma_{ii}}$$

Now, rewrite the payoff function by using the above definition such that:

$$\begin{aligned}
U_i &= \widetilde{\alpha} x_i + \frac{1}{2} \widetilde{\sigma}_{ii} x_i^2 + \sum_{j \neq i} \widetilde{\sigma}_{ij} x_i x_j + \theta \widetilde{Z} \\
U_i &= \frac{\sigma_i}{|\sigma_{ii}|} x_i + \frac{1}{2} \frac{\sigma_{ii}}{|\sigma_{ii}|} x_{ii}^2 + \sum_{j \neq i} \frac{\sigma_{ij}}{|\sigma_{ii}|} x_i x_j + \theta \widetilde{Z} \\
\frac{\partial U_i}{\partial x_i} &= \frac{\alpha_i}{|\sigma_{ii}|} + \frac{\sigma_{ii}}{|\sigma_{ii}|} x_i + \sum_{j \neq i} \frac{\sigma_{ij}}{|\sigma_{ii}|} x_j + \theta \frac{\delta_i}{|\sigma_{ii}|} = 0 \\
\frac{\partial U_i}{\partial x_i} &= \frac{1}{|\sigma_{ii}|} (\alpha_i + \sigma_{ii} x_i + \sum_{j \neq i} x_j + \theta \delta_i) = 0 \\
\frac{1}{|\sigma_{ii}|} &(\alpha_i + \sigma_{i1} x_1 + \sigma_{i2}, x_2 + \ldots + \sigma_{in} x_n + \theta \delta_i) = 0 \quad \forall\ i = 1, \ldots, n
\end{aligned}$$

Let $\widetilde{\Sigma}$ be the following matrix:

$$\begin{bmatrix} \frac{1}{\sigma_{11}} & 0 & 0 & \cdots & 0 \\ 0 & \frac{1}{\sigma_{22}} & 0 & \cdots & 0 \\ 0 & 0 & \frac{1}{\sigma_{33}} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \frac{1}{\sigma_{nn}} \end{bmatrix}$$

Then,

$$\widetilde{\Sigma}(\boldsymbol{\alpha} + \Sigma\mathbf{x} + \theta\boldsymbol{\delta}) = \mathbf{0}$$

In the above equation, $\widetilde{\Sigma}$ is not zero matrix (since the diagonal elements are not equal to 0), then the second term must be equal to a zero vector. By using this, if the equation is solved for $\mathbf{x}$, then:

$$\mathbf{x}^*(\Sigma) = \frac{\mathbf{b}_{\widetilde{\alpha}}(\mathbf{g}, \widetilde{\lambda}^*) + \theta\widetilde{\mathbf{b}_\delta}(\mathbf{g}, \lambda^*)}{\widetilde{\beta} + \widetilde{\gamma}\hat{b}(\mathbf{g}, \widetilde{\lambda}^*)} \tag{2.15}$$

Proof of **Theorem 2.2**

Aggregate Nash equilibrium in the Team Game depends on the Bonacich centrality and the Bonacich centrality weighted by the ability parameter of the player. Note that $\rho_1(\mathbf{G}) > \rho_1(\mathbf{G}^{-i})$. Thus, when $\mathbf{M}(\mathbf{g}, \lambda^*)$ is well defined and nonnegative then so is $\mathbf{M}(\mathbf{g}^{-i}, \lambda^*)$ for all $i = 1, ..., n$.

Let $b_{ji}(\mathbf{g}, \lambda^*) = b_j(\mathbf{g}, \lambda^*) - b_j(\mathbf{g}^{-i}, \lambda^*)$ and $b_{ji}^\delta(\mathbf{g}, \lambda^*) = b_j^\delta(\mathbf{g}, \lambda^*) - b_j^\delta(\mathbf{g}^{-i}, \lambda^*)$ for $j \neq i$ which is the contribution of player i to player j's Bonacich centrality and Bonacich centrality weighted with the ability parameter respectively. The removal of player i from the network has the following effect:

$$\mathbf{b}(\mathbf{g}, \lambda^*) - \mathbf{b}(\mathbf{g}^{-i}, \lambda^*) + \mathbf{b}^\delta(\mathbf{g}, \lambda^*) - \mathbf{b}^\delta(\mathbf{g}^{-i}, \lambda^*) \equiv d_i(\mathbf{g}, \lambda^*)$$

where d_i is the loss function when player i is removed from the network. Our goal is to find i^{th} player whose removal will result in the highest d_i such that $d_{i^*}(\mathbf{g}, \lambda^*) \geq d_i(\mathbf{g}, \lambda^*)$ for all $i = 1, ..., n$. The solution of the first two terms is given by BCZ (2006) on page 1412 under Remark 5, so we focus on the last two terms coming from the additional terms in the team game.

$$
\begin{aligned}
\mathbf{b}^\delta(\mathbf{g}, \lambda^*) - \mathbf{b}^\delta(\mathbf{g}^{-i}, \lambda^*) &= b_i^\delta(\mathbf{g}, \lambda^*) + \sum_{j \neq i}(b_j^\delta(\mathbf{g}, \lambda^*) - b_j^\delta(\mathbf{g}^{-i}, \lambda^*)) \\
&= b_i^\delta(\mathbf{g}, \lambda^*) + \sum_{j \neq i} \sum_{k=1}^n (m_{jk}(\mathbf{g}, \lambda^*) - m_{jk}(\mathbf{g}, \lambda^*))\delta_j
\end{aligned}
$$

Lemma 1: Let $\mathbf{M} = [\mathbf{I} - a\mathbf{G}]^{-1}$ matrix be well defined and nonnegative. Then $m_{ji}(\mathbf{g}, a)m_{ik}(\mathbf{g}, a) = m_{ii}(\mathbf{g}, a)[m_{jk}(\mathbf{g}, a) - m_{jk}(\mathbf{g}^{-i}, a)]$ for all $k \neq i \neq j$.

From Lemma 1 it follows that

$$
\mathbf{b}^\delta(\mathbf{g}, \lambda^*) - \mathbf{b}^\delta(\mathbf{g}^{-i}, \lambda^*) = b_i^\delta(\mathbf{g}, \lambda^*) \sum_{j \neq i} \sum_{k=1}^n \frac{m_{ji}(\mathbf{g}, \lambda^*)m_{ik}(\mathbf{g}, \lambda^*)}{m_{ii}(\mathbf{g}, \lambda^*)}\delta_j
$$

which can be rewritten as:

$$
\begin{aligned}
\mathbf{b}^\delta(\mathbf{g}, \lambda^*) - \mathbf{b}^\delta(\mathbf{g}^{-i}, \lambda^*) &= \sum_{k=1}^n m_{ik}(\mathbf{g}, \lambda^*)\delta_i + \sum_{k=1}^n \sum_{j \neq i} \frac{m_{ji}(\mathbf{g}, \lambda^*)m_{ik}(\mathbf{g}, \lambda^*)}{m_{ii}(\mathbf{g}, \lambda^*)}\delta_j \\
&= \sum_{k=1}^n (m_{ik}(\mathbf{g}, \lambda^*)\delta_i + \sum_{j \neq i} \frac{m_{ji}(\mathbf{g}, \lambda^*)m_{ik}(\mathbf{g}, \lambda^*)}{m_{ii}(\mathbf{g}, \lambda^*)}\delta_j) \\
&= \sum_{k=1}^n (m_{ik}(\mathbf{g}, \lambda^*)\delta_i + \frac{m_{ik}(\mathbf{g}, \lambda^*)}{m_{ii}(\mathbf{g}, \lambda^*)} \sum_{j \neq i} m_{ji}\delta_j)
\end{aligned}
$$

By rearranging terms we obtain:

$$
\mathbf{b}^\delta(\mathbf{g}, \lambda^*) - \mathbf{b}^\delta(\mathbf{g}^{-i}, \lambda^*) = \sum_{k=1}^n m_{ik}(\mathbf{g}, \lambda^*)(\delta_i + \frac{\sum_{j \neq i} m_{ji}(\mathbf{g}, \lambda^*)\delta_j}{m_{ii}(\mathbf{g}, \lambda^*)})
$$

Using the definition $b_i(\mathbf{g}, \lambda^*) = \sum_{k=1}^{n} m_{ik}(\mathbf{g}, \lambda^*)$, we obtain:

$$\mathbf{b}^{\delta}(\mathbf{g}, \lambda^*) - \mathbf{b}^{\delta}(\mathbf{g}^{-i}, \lambda^*) = b_i(\mathbf{g}, \lambda^*)(\delta_i + \frac{\sum_{j \neq i} m_{ji}(\mathbf{g}, \lambda^*)\delta_j}{m_{ii}(\mathbf{g}, \lambda^*)})$$

The above expression measures the effect of player i's removal Bonacich centrality weighted with ability parameter. Combining the above expression with the ICM to get the full effect of player i's removal ($d_i(\mathbf{g}, \lambda^*)$) yields:

$$b_i(\mathbf{g}, \lambda^*) \times \frac{\sum_{j=1}^{n} m_{ji}(\mathbf{g}, \lambda^*)}{m_{ii}(\mathbf{g}, \lambda^*)} + b_i(\mathbf{g}, \lambda^*) \times (\delta_i + \frac{\sum_{j \neq i} m_{ji}(\mathbf{g}, \lambda^*)\delta_j}{m_{ii}(\mathbf{g}, \lambda^*)})$$

By taking into $b_i(\mathbf{g}, \lambda^*)$ parentheses:

$$b_i(\mathbf{g}, \lambda^*) \times (\frac{\sum_{j=1}^{n} m_{ji}(\mathbf{g}, \lambda^*) + \sum_{j \neq i} m_{ji}(\mathbf{g}, \lambda^*)\delta_j}{m_{ii}(\mathbf{g}, \lambda^*)} + \delta_i)$$

The above expression can be further simplified as:

$$\bar{c}_i(\mathbf{g}, a) = b_i(\mathbf{g}, \lambda^*) \times (\frac{\sum_{j=1}^{n} m_{ji}(\mathbf{g}, \lambda^*) + \sum_{j=1}^{n} m_{ji}(\mathbf{g}, \lambda^*)\delta_j}{m_{ii}(\mathbf{g}, \lambda^*)})$$

Proof of **Theorem 2.3**

Nash Equilibrium of the team game does not take into account the externality that a player gets from her teammates. It might be the case that social planner is interested in taking the externality into account as well as considering the each player's effect on the interaction matrix. Therefore, the loss function (the effect of player i's removal) becomes:

$$b(\mathbf{g}, \lambda^*) - b(\mathbf{g}^{-i}, \lambda^*) + r_{\delta}^i(\mathbf{g}, \lambda^*) = b_i(\mathbf{g}, \lambda^*) + \sum_{j \neq i} b_{ji}(\mathbf{g}, \lambda^*) + r_{\delta}^i(\mathbf{g}, \lambda^*) \equiv e_i(\mathbf{g}, \lambda^*)$$

where e_i is the loss function when player i is removed from the network. Our goal is to find i^{th} player whose removal will result in the highest e_i such that $e_{i^*}(\mathbf{g}, \lambda^*) \geq e_i(\mathbf{g}, \lambda^*)$ for all

$i = 1, ..., n$.. By following this approach, we come up with $TICM^e$ which is the following:

$$\hat{c}_i(\mathbf{g}, a) = b_i(\mathbf{g}, a) \times \frac{\sum_{j=1}^{n} m_{ji}(\mathbf{g}, a)}{m_{ii}(\mathbf{g}, a)} + \sum_{j=1}^{n} m_{ji}(\mathbf{g}, a) \times \delta_i$$

Chapter 3

Efficient Networks in Models of Player and Partner Heterogeneity

3.1 Introduction

A growing literature on social and economic networks addresses the dominant effect of networks on various important outcomes such as labor markets, the spread of diseases, education, and crime. Hence, vast research not only in economics, but also in the other disciplines has been conducted to understand how the networks emerge and how they evolve over time. A central theme in the literature on network formation explores the conflict between the set of stable and socially optimal (efficient) networks. Individual optimization happens through Nash networks, where no agent can make herself better off by deviating from her current strategy, given the strategies of the other players. Even though Nash networks are very well studied in broader manners, there has been few studies about efficiency. This chapter fills the gap in the literature of network formation by exploiting the efficient network architectures under heterogenous agents. In particular, 2 types for which Nash networks are known. I address the issue of efficiency in the form of maximizing aggregate utility of players in the network and compare the architectures of efficient networks with Nash networks.

The seminal papers on network formation are Jackson and Wolinsky (1996) and Bala and Goyal (2000). Jackson and Wolinsky (1996) provide two sided link formation, where the cost of forming links is shared by the participants and introduces pairwise stability. Bala

and Goyal (2000) provide a theocratical framework to address network formation in a non-cooperative setting with homogenous players, where the cost of forming links is on one side. Bala and Goyal (2000) discuss two different types of flow. In the two way flow model, the network is undirected, so both players participating in a link access can access benefits from each other. However, in the one way flow model the network is directed; hence, only the initiating player can access the benefit of the link. This chapter follows the Bala and Goyal (2000) framework, where link costs are levied only by the person initiating the link.

Galeotti et al. (2006) relaxes the homogeneity assumption in the two way flow model, so the benefits from a link and the cost of sponsoring a link are player dependent. A more recent paper (Billand et al. 2010) introduces partner heterogeneity in the two way flow model, where the benefit and cost of making a link is partner heterogenous, meaning that it only depends on whom is being accessed in terms of the benefits and costs. Although Nash networks are clearly identified under heterogeneity, not much has been done about efficient networks under heterogeneity. Galeotti et al. (2006) conclude that Nash and efficient networks coincide under linear payoff situations; however, I show that with more general payoff function specifications this result does not always hold. Specifically, I illustrate situations where Nash and efficient networks do not coincide.

Galeotti (2006) and Billand et al (2011) extend the one way flow model by allowing benefits and costs to be player and partner heterogenous respectively. In a one way flow model, only the node who sponsors a link accesses the benefit. This fact yields a wheel type efficient architecture, that coincides with strict Nash networks architecture.

I start with a general payoff in a two way flow model satisfying common assumptions in the literature. Without imposing any restrictions, efficient networks can have maximal diameter and it is not possible to characterize the architecture. Once we allow for heterogeneity between players, the efficient network architectures depend on four factors: *(i)* the value of players (benefit obtained by linking to each player), *(ii)* the number of minimum cost players, *(iii)* the difference in cost between the minimum cost player(s) and the other players and *(iv)*

the functional form of payoffs. The first factor is controlled by a restriction which ensures that all links are profitable. I do not impose any restrictions to control the second factor, since having more than one minimum cost player does not change the results qualitatively. To deal with the third factor, I introduce the widely used linearity, strict concavity, and convexity assumptions on the payoffs. The fourth factor is accounted for with the restrictions that are already imposed, but I introduce an additional condition in the two way flow player heterogeneity model.

I provide the architecture, as well as, the diameter to identify the efficient networks. The architecture provides information about how the efficient networks look, and the diameter helps to determine the maximum distance between any two players in the network. The crucial difference between player and partner heterogeneity models is in the change of the player who sponsors the links. In addition, there are some differences in the network formation model and results under player or partner heterogenous players. Finally, for one way flow models, there is no difference between player and partner heterogeneity models and the Nash and efficient architectures coincide.

Rest of the chapter is organized as follows: Section 2 presents the basics of the model setup and some important definitions. Section 3 identifies the efficient networks in a player heterogeneity model. Section 4 demonstrates the efficient networks in the partner heterogeneity model. Section 5 discusses one way flow models. The chapter concludes with a discussion and comparison of Nash and efficient networks.

3.2 Model Setup

$N = \{1, \ldots, n\}$ denotes the set of players. A directed network $\mathbf{g} = (N, A)$ is a pair of sets: the set N of players and the set $A \subset N \times N$ of links. $A(\mathbf{g})$ denotes the set of links of network $\mathbf{g}$. The undirected counterpart of $\mathbf{g}$, $\bar{\mathbf{g}}$, is obtained by ignoring the orientation of arcs of $\mathbf{g}$ and treating links between the same players as a single link.

Each player i chooses a strategy $\mathbf{g}_i = (g_{i,1}, \ldots, g_{i,i-1}, g_{i,i+1}, \ldots, g_{i,n})$ where $g_{i,j} \in \{0,1\}$ for all $j \in N \setminus \{i\}$. The interpretation of $g_{i,j} = 1$ is that player i forms an arc with player $j \neq i$, and the interpretation of $g_{i,j} = 0$ is that i forms no arc with player j. I assume that player i cannot form an arc with herself. I only consider pure strategies. Let $\mathcal{G}_i$ be the set of all strategies of player $i \in N$. Network relations among players are formally represented by directed networks whose nodes are identified by the players. I assume that if $g_{i,j} = 1$, then $ji \in A(\mathbf{g})$. An arc ji is shown by an arrow from j to i. Thus, if i chooses to link with j, the arc will be directed from j to i and it also means that player i sponsors the cost of the link formation. Hence, it is assumed that cost of link formation is always on one side of the participants of the link.

For a directed network, $\mathbf{g}$, a path from player k to player j, $j \neq k$, is a finite sequence $j_0, j_1, \ldots, j_m$ of distinct players such that $j_0 = j$, $j_m = k$ and $\mathbf{g}_{j_\ell, j_{\ell+1}} = 1$ for $\ell = 0, \ldots, m-1$. A chain exists between player k and player j, $j \neq k$ by replacing $\mathbf{g}_{j_\ell, j_{\ell+1}} = 1$ by $\max\{g_{j_\ell, j_{\ell+1}}, g_{j_{\ell+1}, j_\ell}\} = 1$.

Given a network $\mathbf{g}$, I define a component, $D(\mathbf{g})$, as a set of players such that there is a chain between any two players who belong to $D(\mathbf{g})$, and there does not exist a chain between a player in $D(\mathbf{g})$ and a player who does not belong to $D(\mathbf{g})$. A network $\mathbf{g}$ is said to be connected if it contains one component. A network is said to be minimal if any removal of a current links lead to an increase in the number of components. Finally, a network is minimally connected if it is not possible to preserve its connectivity whenever a link is removed.

A network $\mathbf{g}$ is a star if there is a player i such that $\max g_{i,j}, g_{j,i}\} = 1$ for all $j \in N \setminus \{i\}$ and $g_{\ell,j} = 0$ for all $\ell \in N \setminus \{i\}$ and $j \in N \setminus \{i, \ell\}$. The network $\mathbf{g}$ is an inward pointing star or center sponsored star if it is a star and for the center player i, we have $g_{j,i} = 0$ for all $j \in N \setminus \{i\}$. The network $\mathbf{g}$ is an outward pointing star or periphery sponsored star if it is a star and for the center player i, we have $g_{i,j} = 0$ for all $j \in N \setminus \{i\}$.

A network in which each group constitutes a star and a single player i of group l forms a

link with the central player j of group $l^{`}$ where $l \neq l^{`}$, then it is referred to as an interlinked star network. If each star is center-sponsored (periphery-sponsored), the network is said to be an interlinked center-sponsored (periphery-sponsored) star.

Define $N_i(\mathbf{g}) = \{i\} \cup \{j \in N \setminus \{i\} \mid$ there is a chain between i and j in $\mathbf{g}\}$ as the set of players who are observed by player i with the convention that player i always "observes" herself. I assume that player i obtains no additional resources from herself by forming arcs. However, player i can obtain her own resources even if she forms no arcs and there is no network.

I will now define two classes of models that assumes the benefit and cost of a link can be different among players. Precisely, in the player heterogeneity model proposed by (Galeotti et al. 2006), each player i obtains $V_i > 0$ from each player $j \in N_i(\mathbf{g}) \setminus \{i\}$, and incurs a cost $c_i > 0$ when she forms a link with player $j \neq i$. In the partner heterogeneity model introduced by (Billand et al. 2009), each player i obtains $V_j > 0$ from each player $j \in N_i(\mathbf{g}) \setminus \{i\}$, and incurs a cost $c_j > 0$ when she forms a link with player $j \neq i$. In these models it is assumed that each player in the network contains a value of information and whenever a link is formed between player i and j the players obtain player or partner specific benefits. For example, in the player heterogeneity model, player i obtains $V_i > 0$ from each link formed. So, the benefits only depend on the player i's characteristics. On the other hand, in the partner heterogeneity model, player i obtains V_j for $i \neq j$ form each link formed. Therefore, the benefit player i obtains from each link depends on to whom player i is connected. This implies that benefits obtained by player i form each link depends on the characteristics of player j.

Another class of models considers the direction of information flow. In the two way flow model, both players sharing a link access each other; hence, the information flow is symmetric (undirected). On the other hand, the one way flow model assumes that the network is directed; therefore, only the sponsor of a link can access benefit of the link.

Let $\pi : \mathbf{g} \to \mathbb{R}$ be the payoff of player i where $\phi : \mathbb{R}^2 \to \mathbb{R}$ and $\phi(x, y)$ is strictly increasing

in x and strictly decreasing in y.

$$\pi_i(\mathbf{g}) = \phi \left(\sum_{j \in N_i(\mathbf{g}) \setminus \{i\}} V, \quad \sum_{j \in N \setminus \{i\}} \mathbf{g}_{i,j} c \right) \tag{3.1}$$

Given the properties of the function ϕ, the first term, x can be interpreted as the "benefits" that agent i receives from her links, while the second term, y measures the "costs" associated with forming them. Observe that this is the model proposed by Bala and Goyal (2000). In the player heterogeneity model, V_i replaces V and c_i replaces c while in the partner heterogeneity model, V_j replaces V and c_j replaces c in the payoff function. This representation breaks the player uniformity in two variations, namely player and partner heterogenous players.

Given a network $\mathbf{g}$, the aggregate payoff is denoted by: $W(\mathbf{g}) = \sum_{i=1}^n \pi_i(V, c)$. A network is said to be efficient if $W(\mathbf{g}) \geq W(\mathbf{g}')$ for any $\mathbf{g}' \neq \mathbf{g}$ where $\mathbf{g}'$ represents all other possible networks.

3.3 Player Heterogeneity Model

3.3.1 General Payoff Function

In this section, I study the efficient network architectures with player heterogenous agents in two-way flow information. Player heterogeneity allows each player to obtain player specific value of information and cost of link formation. Hence, the value of information and cost of link formation only depends on the individual attributes and abilities of the players. For example, suppose player i and j have a link which allows them to access information from each other. Without loss of generality suppose that player i is sponsoring the link. Then, the value of information player i receives from this link is represented by V_i and the cost of link formation is represented by c_i. On the other hand, the value of information player j receives is V_j and it is not necessarily equal to V_i.

Let $\pi : \mathbf{g} \to \mathbb{R}$ and $\phi : \mathbb{R}^2 \to \mathbb{R}$ be such that $\phi(x, y)$ is strictly increasing in x and strictly

decreasing in y. So, the payoff function for player i is:

$$\pi_i(\mathbf{g}) = \phi \left(\sum_{j \in N_i(\mathbf{g}) \backslash \{i\}} V_i, \quad \sum_{j \in N \backslash \{i\}} g_{i,j} c_i \right) \tag{3.2}$$

The above representation does not assume any particular functional form and it does not impose any restriction on the levels of value of information and cost of link formation. One immediate result for the efficient networks is that it is not possible to characterize the architecture. In particular, there might be some players who are not connected to the other players; hence, the diameter of the efficient network goes to infinity. Even though we cannot characterize the diameter or architecture of the network, the following proposition shows that we still obtain minimal networks as the efficient networks.

Proposition 1: Suppose the payoff function is given by (2) and information flow is two-way without decay. Then, any efficient network, $\mathbf{g}$ is minimal.

Proof: Suppose $\mathbf{g}$ is a minimal network and $\mathbf{g}'$ represents all the other possible non-minimal networks where the number of observed agents are the same in $\mathbf{g}$ and $\mathbf{g}'$. Then, there exists two players i and j in $\mathbf{g}'$ such that there exists more than one link between the players i and j. Then, it is possible to maintain the total benefits but decrease the total cost of link formation by deleting the extra link between player i and j since we assume there is no decay and information flow is symmetric. We assume that ϕ is a strictly decreasing function in cost, then it follows that $W(\mathbf{g}) > W(\mathbf{g}')$. Hence, $\mathbf{g}'$ cannot be an efficient network.

The above proposition shows that the players access to each other with minimum number of links in the efficient networks. This indicates that if player i and j are connected in the efficient network, there will be only one path from player i to j. This means that any redundant links are eliminated to maximize the aggregate payoff of the players. It follows that the efficient networks are minimal in two way flow player heterogeneity model. Note that this result also holds for two way flow partner heterogeneity model.

To ensure connectivity in the efficient network, I impose the following assumption, A1. A1 eliminates the efficient networks which may include singletons or disconnected components. Once A1 is satisfied, efficient networks are connected.

Assumption 1 (A1): $\phi(x + V_i, y + g_{ij}c_i) \geq \phi(x, y)$, for all $i \in N$.

Lemma 1: Suppose the payoff function is given by (2) and A1 is satisfied. Then, any efficient network, $\mathbf{g}$ is minimally connected.

Proof: Suppose not. Then there exists an efficient network, $\mathbf{g}'$ with two components $D_1(\mathbf{g}')$ and $D_2(\mathbf{g}')$. From Proposition 1 we know that $D_1(\mathbf{g}')$ and $D_2(\mathbf{g}')$ are internally minimal but they can still be disconnected from each other. By A1, we know that making links to an unobserved player provides positive net payoffs. Now consider two players k and m such that $k \in D_1(\mathbf{g}')$ and $m \in D_2(\mathbf{g}')$. Let $\mathbf{g}$ be a network such that $\mathbf{g} = \mathbf{g}' + km$. Then, by A1, it follows that $W(\mathbf{g}) > W(\mathbf{g}')$. Therefore $\mathbf{g}'$ cannot be an efficient network and the proof follows. From the above contradiction, we conclude that if A1 is satisfied, then the efficient networks do not have any singletons or disconnected components. As shown in Proposition 1, to maximize the aggregate payoff any redundant links in the network are eliminated. Combining these results, we obtain that the efficient networks are minimally connected.

Note that Lemma 1 does not directly impose a condition on the functional form and holds for both player and partner heterogeneity models. It says that values obtained by connecting to other players and costs of link formation vary such that every link is profitable, which implies that efficient networks have one component. Even though this assumption leads to analyzing only connected networks, it makes it possible to identify and characterize the efficient network structure.

Efficient networks are expected to have a very low diameter and exhibit very high centrality. However, once the values and costs are allowed to be player (partner) specific, the efficient networks can have more decentralized architectures. The next proposition indicates that when the benefits and costs of link formation are allowed to vary freely, it is possible

to have decentralized architectures and even maximal diameter.

Proposition 2: Suppose the payoff function is given by (2) and satisfies A1. There exists parameter values V_i and c_i such that the efficient network, $\mathbf{g}$, has maximum diameter, $(n-1)$.

Formal proof is omitted. To illustrate an efficient network which has maximal diameter, I provide the following example which utilizes an additively separable form, where the benefits and costs of link formation can be expressed as separate terms.[1]

Example 1: Suppose the payoff function of player i is given by:

$$\pi_i(\mathbf{g}) = \Big(\sum_{j \in N_i(\mathbf{g}) \backslash \{i\}} V_i \Big) - \Big(\sum_{j \in N - \{i\}} g_{i,j} c_i \Big)^2$$

Suppose there is one minimum cost player represented by L and 3 high cost players represented by H. Let c_L and c_H be equal to 2 and 3 respectively. Assume that all players have the same $V_i = V = 100$.[2] Note that A1 holds for these parameter values. Now suppose that all links are sponsored by L, which can be represented by an inward pointing star. For this case, $\pi_L = \sqrt{400} - (3 \times 2)^2 = -16$ and $\pi_H = \sqrt{400}$ which implies $W(g') = \sum_i \pi_i(\mathbf{g'}) = 44$. Compare this with the efficient network $\mathbf{g}$, where each player sponsors a single link except one H type player. For this case, $\pi_L = 16$ and $\pi_H = 20$ for the H type player who did not sponsor any link and $\pi_H = 11$ for the ones who sponsored a single link. This yields $W(\mathbf{g}) = 58$. This example shows that the diameter can be as high as $(n-1)$ depending on the payoff function. Due to general payoff function specification, we cannot ensure that all the links will be sponsored by the minimum cost player in the efficient network. After sponsoring a single link, the marginal cost of sponsoring another link to the minimum cost player is $16 - 4 = 12$. However, the marginal cost of sponsoring a single link for a high cost player is $3^2 = 9$.

---

[1] I consider an additively separable form which has economic interpretation and easy to construct. The below function satisfies the assumptions of being strictly increasing in value and strictly decreasing in cost. In addition, it exhibits strict concavity in values and strict convexity in cost.

[2] For simplicity I introduce two types of players represented by L and H. However, it is possible to have any heterogeneity between players in the values and costs.

The above example provides the intuition to determine the factors that affect the architecture of the efficient networks. It also demonstrates that strict concavity or convexity of the payoff function in terms of cost plays a role in determining the diameter of the efficient network.[3] However, strict concavity or convexity may not be enough to characterize the efficient network since V_i and c_i also play a role. Without a restriction on the payoff specification, it is not possible to characterize the efficient networks. Therefore, I impose restrictions on the payoff function. I consider the following cases: linear payoffs, and strictly concave and convex payoffs in the cost argument.

3.3.2 Linear Payoffs

Suppose the payoff function of player i is given by:

$$\pi_i(\mathbf{g}) = \sum_{j \in N_i(\mathbf{g}) \setminus \{i\}} V_i - \sum_{j \in N \setminus \{i\}} g_{i,j} c_i \tag{3.3}$$

where $\forall i, j \in N$.

Proposition 3: Suppose the payoff function is given by (3) and $\sum_{i \in N} V_i \geq c$ where $c = argmin_{i \in N}\{c_i\}$. If $\mathbf{g}$ is an efficient network, then it is minimally connected and $2 \leq D(\mathbf{g}) \leq 2s^{[c_{i_0}]}$. Moreover, the efficient network is a center sponsored star or a center sponsored interlinked star.

Proof: I start with proving the diameter and the architecture follows. It is clear that if $\mathbf{g}$ is an efficient network and $\sum_{i \in N} V_i \geq c$ then $\mathbf{g}$ is minimally connected. If $n \geq 3$, it follows that $D(\mathbf{g}) \geq 2$. Moreover, if a player $i \in N$ forms a link with player k, then $i \in argmin_{j \in N}\{c_j\}$, otherwise it is possible to replace the link i, k by a link j, k where $j \in argmin_{j \in N}\{c_j\}$ to obtain a higher total utility. I now show that it is not possible to have an efficient network $\mathbf{g}$ with $D(\mathbf{g}) > 2s^{[c_{i_0}]}$. To introduce a contradiction, assume that $\mathbf{g}$ is an efficient network such

---

[3]In Example 1, if the cost term is taken as square root instead of square then the efficient network has diameter equal to 2.

that $D(\mathbf{g}) > 2s^{[c_{i_0}]}$. Then, there exists a chain $\mathcal{C}_{i,k}$ between two players say i and k such that $\ell(\mathcal{C}_{i,k}) > 2s^{[c_{i_0}]}$. In other words, there are more than $2s^{[c_{i_0}]}$ links between i and k. It follows that at least one link has been formed by $j \notin S^{[c_{i_0}]}$, a contradiction. Star or interlinked star architecture occurs since the high cost players will be gathered around the minimum cost player(s).

Observe that if there is a single minimum cost player, then the efficient network is a center sponsored star. If there is more than one minimal cost players, then the star architecture is still efficient. Galeotti et al. (2006) concludes that the efficient networks and strict Nash networks are the same when a linear payoff specification is assumed. However, if there are more than one minimum cost players in the network, then the center sponsored interlinked star, where the links are sponsored by the minimum cost players is also efficient. This points out a conflict between stability and efficiency.

Figure 1 illustrates the possible efficient networks associated with Proposition 2.[4]

(a) Center Sponsored Star

(b) Interlinked Star I

(c) Interlinked Star II
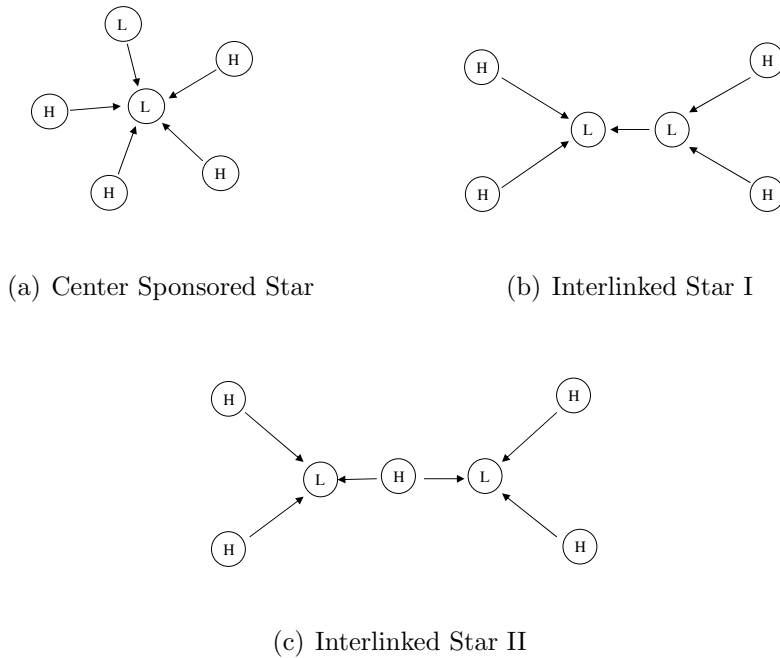
Figure 3.1: Examples of Efficient Networks

[4]In the below figures, L represents minimum cost players and H represents high cost players. H type players can have any cost $c + \epsilon_i$ where $\epsilon > 0$.

A linear payoff specification is the simplest case and it is possible to identify the diameter and architecture of the efficient network without A1.[5]

3.3.3 Strictly Concave and Convex Payoffs in Cost

With a linear payoff function, it is possible to determine the diameter and architecture without any strong restrictions. However, the interpretation of the model is limited. Strict concavity or convexity specifications are widely used in the literature to address this limitation. I define strictly concave and convex payoff functions[6] in terms of the cost argument in the following way:

Definition 1: Suppose $c_1, c_2 \in (0, \sum_{i \in N} c_i), c_1 < c_2$ $\kappa \in (0, c_1]$. If $\phi(V, c_1) - \phi(V, c_1 - \kappa) > \phi(V, c_2) - \phi(V, c_2 - \kappa)$ then the payoff function is strictly concave in the cost argument.

Definition 2: Suppose $c_1, c_2 \in (0, \sum_{i \in N} c_i), c_1 < c_2$ $\kappa \in (0, c_1]$. If $\phi(V, c_1) - \phi(V, c_1 - \kappa) < \phi(V, c_2) - \phi(V, c_2 - \kappa)$ then the payoff function is strictly convex in the cost argument.

Proposition 4: Suppose the payoff function is given by (2) and $\phi(x, y)$ satisfies strict concavity in cost and A1. If **g** is an efficient network, then it is minimally connected and the unique efficient architecture is a center sponsored star.

Proof: It is clear that if **g** is an efficient network and $\sum_{i \in N} V_i \geq c_i$ then Lemma 1 implies that **g** is minimally connected. Since $n \geq 3$, it follows that $D(\mathbf{g}) \geq 2$. Moreover, if a player $i \in N$ forms a link with player k, then $i \in argmin_{j \in N} c_j$. Otherwise, it is possible to replace the link i, k by a link j, k where $j \in argmin_{j \in N} c_j$ to obtain a higher total utility. From the property of strictly concave functions, $\phi(V, (k-1) \times c_i) < (k-1) \times \phi(V, c_i)$. Therefore, to maximize total utility, all the links must be sponsored by a single minimum cost player.

---

[5]Assuming A1 will provide the same result, however note that A1 is stronger than assuming $\sum_{i \in N} V_i \geq c$ where $c = argmin_{i \in N} \{c_i\}$.

[6]An important fact to mention is that all the concave (convex) functions are subadditive (superadditive); however, the reverse is not true. The subadditivity (superadditivity) property implies that the marginal cost of adding links for a minimum cost player decreases (increases) as the number of links sponsored increases. These properties enable us to track the marginal cost of sponsoring links and simplify the identification of efficient networks.

Since all the links are sponsored by a single minimum cost player, star is the only possible efficient architecture.

Note that under the strict concavity case, the number of minimum cost players does not play a role, the unique efficient network is a center sponsored star. For the convexity case, I will impose an additional condition to characterize the efficient network.

Assumption 2 (A2): Let c be $c = argmin_{c\{j \in n\}}$. Suppose $\phi(x, y)$ satisfies $\phi(V, k \times c) - \phi(V, (k-1) \times c) < \phi(V, k \times c_i) \forall c_i \neq c$ and $k = \{1, 2, ..., n\}$.

Note that the above assumption is not on the functional form of the payoffs. It can easily be shown through modifying Example 1 that the diameter can be as high as $(n-1)$ when the payoff function is strictly convex in cost. Condition 1 eliminates cases where the marginal cost of sponsoring a link for a high cost player is less than the marginal cost of sponsoring a link for a minimum cost player. Once Condition 1 is satisfied, all the links will be sponsored by the minimum cost player(s).

Proposition 5: Suppose the payoff function is given by (2). Also suppose $\phi(x, y)$ satisfies the convexity in the cost argument and satisfies A1 and A2 about the degree of convexity in cost. If $\mathbf{g}$ is an efficient network, then it is minimally connected and $2 \leq D(\mathbf{g}) \leq 2s^{[c_{i_0}]}$. The efficient network is a center sponsored star for $s^{[c_{i_0}]} = 1$ and an interlinked star for $s^{[c_{i_0} > 1]}$.

Proof: It is clear that if $\mathbf{g}$ is an efficient network and $\sum_{i \in N} V_i \geq c_i$ then $\mathbf{g}$ is minimally connected. Since $n \geq 3$, it follows that $D(\mathbf{g}) \geq 2$. Condition 1 ensures that even though ϕ is strictly convex in cost, to maximize total utility the minimum cost player still sponsors the links. That is, if a player $i \in N$ forms a link with player k, then $i \in argmin_{j \in N} c_j$, otherwise it is possible to replace the link i, k by a link j, k where $j \in argmin_{j \in N} c_j$. Condition 1 ensures that all the links will be sponsored by the minimum cost players. I obtain the interlinked star type architecture where a high cost player can stay in the center as a bridge player without sponsoring any links.

Note that if there is more than one minimum cost players in $\mathbf{g}$, then a center sponsored

interlinked star is the efficient architecture as opposed to the star architecture. The intuition for this case is that it becomes inefficient for a single minimum cost player to sponsor all the links since the payoff is strictly convex in cost.

The efficient networks in two way flow and player heterogeneity models generally exhibit high centrality and low diameter. Introducing the player heterogeneity increases the possibilities of efficient networks. Compared to Bala and Goyal (2000), in addition to the star architecture, interlinked stars can also be efficient, which occurs in linear specification and strictly convex in cost cases. Assuming strict convexity requires an additional condition that ensures all links are sponsored by the minimum cost player(s).

3.4   Partner Heterogeneity Model

Partner heterogeneity model introduced by Billand et al. (2011) considers a framework where benefits and costs of link formation vary according to whom is being accessed. With this type of heterogeneity, the set of strict Nash networks substantially increase and new architectures arise. In this section, I characterize the efficient network architectures with player heterogenous agents in two-way flow information. Partner heterogeneity allows each player to obtain partner specific value of information and cost of link formation. Hence, the value of information and cost of link formation only depends on the attributes and abilities of the player to whom players access. For example, suppose player i and j have a link which allows them to access information from each other. Without loss of generality suppose that player i is sponsoring the link. Then, the value of information player i receives from this link is represented by V_j and the cost of link formation is represented by c_j which is sponsored by player i. On the other hand, the value of information player j receives is V_i and it is not necessarily equal to V_j.

Let $\pi : \mathbf{g} \to \mathbb{R}$ and $\phi : \mathbb{R}^2 \to \mathbb{R}$ such that $\phi(x, y)$ is strictly increasing in x and strictly decreasing in y. Let the payoff function for player i be:

$$\pi_i(\mathbf{g}) = \phi \left(\sum_{j \in N - \{i\}} V_j, \sum_{j \in N \setminus \{i\}} g_{i,j} c_j \right) \tag{3.4}$$

Remark 1: Partner Heterogeneity with General Payoffs

As in the player heterogeneity case when the values and costs vary freely, it is possible to have an efficient network, $\mathbf{g}$ which has maximal diameter equal to $(n-1)$. This can easily be shown by modifying Example 1, where V_j replaces V_i and c_j replaces c_i. Therefore, I impose additional restrictions on the payoffs as in the player heterogeneity model. As before, I consider the linear payoffs, strictly concave, and convex payoffs in cost cases.

Remark 2: Partner Heterogeneity with Linear Payoffs

To have partner heterogeneity, V_j replaces V_i and c_j replaces c_i in (2). The diameter and efficient architecture for this case is essentially the same as the player heterogeneity with linear payoffs. The main difference between the player and partner heterogeneity model is who sponsors the links. In the player heterogeneity case, the minimal cost player sponsors the links. However, in the partner heterogeneity case, the links are sponsored by non-minimal cost players. Therefore, the efficient network has periphery sponsored architecture for the partner heterogeneity case.

Remark 3: Partner Heterogeneity with Strictly Concave Payoffs in Cost

The diameter and the architecture can be different from the player heterogeneity case. In fact, $2 \le D(\mathbf{g}) \le 4$. If the marginal cost of adding links for a high cost player becomes very little after sponsoring links with the minimum cost players, the efficient network can have a center sponsored star architecture. However, if the marginal cost of adding links for a high cost player does not decrease enough, it is possible that there is a core in the center where a high cost player sponsor links to the minimum cost players, and then other high cost players
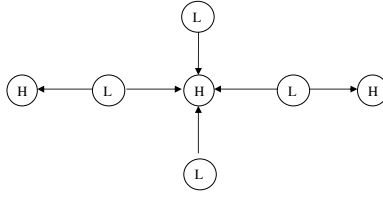
69

Figure 3.2: Core and Periphery

sponsor a link with the minimum cost players and form a periphery around the core. This architecture can be called mixed star and the diameter in that case can be at most 4. Figure 2 illustrates the possible architecture with diameter of 4.

Remark 4: Partner Heterogeneity with Strictly Convex Payoffs in Cost

For this case, $2 \leq D(\mathbf{g}) \leq 1 + s^{[c_{i_0}]}$. Note that this architecture is different from the player heterogeneity. In player heterogeneity, a non-minimal cost player can serve as a bridge player as seen in Figure 1. However, in the partner heterogeneity, efficient network does not have such architecture. Observe that to have a bridge the player in player heterogeneity model is not sponsoring any links. These links are sponsored by the minimal cost players around the bridge player. However, such a bridge player in the partner heterogeneity model requires sponsoring two links for that player, which is not efficient when payoff is strictly convex cost.

3.5 One-Way Flow Models

In this section, I identify the efficient networks in the one way flow models. In the one way flow model, only the player who sponsors the link accesses the benefit of the link. I consider Galeotti (2006) and Billand et al. (2009) models, which are extensions of Bala and Goyal (2000). Both of these papers establish strict Nash networks; however, they give little information about the efficient networks.[7] The next proposition identifies the efficient network for player heterogenous one way flow model, which is an adaption from Bala and

---

[7]Galeotti (2004) working paper discusses the efficient architecture and compares it with Nash networks very briefly.

Goyal (2000).

Let $\pi : \mathbf{g} \to \mathbb{R}$ such that $\phi(x, y)$ is strictly increasing in x and strictly decreasing in y. Let the payoff function for player i be:

$$\pi_i(\mathbf{g}) = \pi_i \left(B(\sum_{j \in N \ \{i\}} g_{i,j} \hat{V}_i), \ C(\sum_{j \in N} g_{i,j} c_i) \right) \tag{3.5}$$

Proposition 6: Suppose the payoff function is given by (5) and the condition in Lemma 1 is satisfied. If $\phi(\sum_{i=1}^n V_i, c_i) > \phi(V_i, 0)$ for all $i = \{1, ..., n\}$ then the unique efficient architecture is a wheel. If $\phi(\sum_{i=1}^n V_i, c_i) < \phi(V_i, 0)$ for all $i = \{1, ..., n\}$ then the empty network is efficient.

Proof: I will first consider the connected case. Let Γ be the set of values $(B(\mathbf{g}), C(\mathbf{g}))$. If $C(\mathbf{g}) = 0$, then $B(\mathbf{g}) = V_i$, while if $C(\mathbf{g}) \in \{1, ..., n-1\}$, then $C(\mathbf{g}) \in V_i, \sum_{i=1}^n V_i$. Thus, $\Gamma \subset \{1, ..., n\} x \{1, ..., n-1\} \cup \{(1, 0)\}$. Given $(x, y) \in \Gamma - (V_i, 0)$, then $\pi(\sum_i^n V_i, c_i) \geq \pi(\sum_i^n V_i, y) \geq \pi(x, y)$ since π is increasing in benefits and decreasing in costs. For the wheel network g^w, $B_i(g^w) = \sum_i^n V_i$ and $C_i(g^w) = c_i$. For any other network $g \neq g^w$ for each $i \in N$, if $C_i(g) \geq c_i$, then $B_i(g) \leq \sum_i^n$, while if $C_i(g) = 0$, then $V_i(g) = V_i$ since the player observes herself. In either case, $\pi_i(g^w) = \pi(\sum_{i=1}^n V_i, c_i) \geq \pi(B_i(g), C_i(g)) = \pi_i(g)$, $\pi(\sum_{i=1}^n V_i, c_i) > \pi(V_i, 0)$ for all $i = 1, ..., n$. It follows that $W(g^w) = \sum_{i \in n} \pi_i(\sum_{i=1}^n V_i, c_i) \geq \sum_{i \in n} \pi_i(B_i(g), C_i(g)) = W(g)$ as well. Therefore, g^w is an efficient architecture. To show uniqueness, note that the assumptions on π imply that $W(g^w > W(g)$ if $C_i(g) \neq c_i$ or if $B_i(g) < n$. Let $g \neq g^w$ be given; if $C_i(g) \neq 1$ for even one i, then $W(g^w) > W(g)$. On the other hand, suppose $C_i(g) = 1$ for all $i \in N$. As the wheel is the only connected network with n agents, and $g \neq g^w$, there must be an agent j such that $B_j(g) \leq \sum_{j=1}^n V_j$. Thus, $W(g^w) > W(g)$, proving uniqueness.

In a wheel architecture, sponsoring a single link allows all the agents to access each other. Note that the order of players or having more than one minimum cost player does not have

any effect on the efficient architecture.

Remark 5: Partner Heterogeneity in one way flow model

For this case only the condition to have a connected efficient architecture changes. If $\phi(\sum_{i=1}^{n} V_j, c_j) > \phi(V_j, 0)$ for all $i = 1, ..., n$ then the unique efficient architecture is a wheel. If $\phi(\sum_{i=1}^{n} V_i, c_i) < \phi(V_i, 0)$ for all $i = 1, ..., n$ then the empty network is efficient. Note that the partner heterogeneity setting yields the same result as the player heterogeneity model. Again, by sponsoring a single link and incurring the cost c_j, each player can access all players in the network.

3.6 Discussion and Conclusion

This chapter studies the efficient networks in the player and partner heterogeneity models with various functional forms. With a two way flow of information, I find that payoff function specification, the differences in cost between minimum cost player(s) and the other players and the number of minimum cost players play a crucial role in determining the architecture of the efficient networks. I deal with the first two determinants by imposing additional restrictions if necessary. Note that having more than one minimum cost players introduces some symmetries and additional efficient network architectures. However, the efficient network architectures in the case of having more minimum cost players can still be identified without imposing any restrictions unlike the other two determinants.

There are some notable differences between the efficient architectures under the player and partner heterogeneity models. The first difference is in the player heterogeneity model minimum cost player(s) sponsor all the links. However, in partner heterogeneity a high cost player can also sponsor some links. In the homogenous model discussed by Bala and Goyal (2000), efficient networks have star architecture. However, when we allow for heterogeneity in values and costs, a rather decentralized efficient architecture in the form of interlinked stars occurs in linear and strictly convex payoffs in cost.

In the one way flow model, the architecture of the efficient network with player and partner heterogeneity is identical. The only connected efficient architecture is a wheel for these cases, where each agent can access to all other agents by sponsoring a single link. I conclude that strict Nash networks in one way flow model discussed by Galleotti (2006) and Billand et al. (2011) are also efficient.

Nash and efficient networks coincide if the payoff function is linear in benefits and costs. This holds true for one way flow of information and some cases in the two way flow of information. However, we show that in two way flow of information this result is not robust when payoff function exhibits strict convexity in cost. Also, with the linear specification efficient networks can also be interlinked stars while only star type architectures are Nash networks. Note that not all of the payoff function specifications are comparable in terms of Nash and efficient networks since I impose additional requirements to identify link formation in some cases.

Chapter 4

Identity and Link Reliability in Networks

4.1 Introduction

Networks have an undoubtable effect on how we take our places in the society. A vast literature on network formation sheds light on how the networks form and take shape under different circumstances. Many empirical observations suggests that one of the key determinant in the network formation is the similarities in the characteristics of the players. In this chapter, we consider a network formation game where the identity characteristics are introduced to capture the similarities between players. We consider a framework where the links between players are not fully reliable i.e, the success of a link is probabilistic.

We introduce identity as a single dimension variable to capture the different characteristics of players. We assume that identity characteristics are assigned exogenously to the players and each player decides about how much to commit to her current characteristic. Nationality, race and culture are good examples of exogenously assigned identities.

Players access to information by making links with others. We assume two way flow and undirected network without decay which means that both players access to the same benefit of the link through direct or indirect connections. We consider two cases of costly link formation. In the first case, the cost of the link is paid by the initiator. In the second case, the players who are involved in the link share the cost depending on their initial link

offers. Hence, the second case enables us to investigate the network formation under mutual consent.

In a non-cooperative network game, players how much to commit their identities and their linking strategies. Player's commitment decision affect the cost of forming a link. By choosing her commitment level, a player reveals which type of players she can easily form a link. We assume that players with same identities and higher commitment levels can make links easier. Cost of link formation with different identities and higher commitment levels becomes very costly. In this setting, cost of link formation c becomes player and partner heterogenous, c_{ij}. For simplicity, we assume that the benefits from all the links are the same which is denoted by V.

We introduce link and node failures into the network formation model. Link and node reliability sheds light into network formation in a realistic setting. Even though a link exists between players, the transmission of information might fail without a further notice. We first introduce an exogenous and constant link failure probability, p. Then, we relax this assumption heterogenize the link failures and allow pairwise link failures, p_{ij}. For simplicity of the analysis, we assume that link reliability probabilities within the same identity group are equal to each other. We show that costly link formation between different identities can lead to fragmented architectures. However, with fully heterogenous link failures, it is possible to have more integrated groups with different identities which may include many components however this only occurs if the link between the different identity groups are more reliable than a link between the players when both players involved in the link share the common identity characteristic. The actual determination of reliability probability ranges are fairly complicated and we demonstrate the intuition with an example for some cases.

Next, we allow each player (or node) in the network fail with a constant probability, f. Player failure can happen if player is not reachable or if she simply goes beyond reach with a certain probability. The difference between link and player imperfection is that in the former each existing link is successful with a probability and in the latter all the links that player

i is involved might fail. We identify the Nash network(s) under these link and node failures and with the available scenarios for identity and commitment levels.

Akerlof and Kranton (2000), Chandra (2001), Chen et al. (2007), Currarini, Jackson, and Pin (2008), De Mari and Zenou (2009), Dev (2009) and Dev (2010) study the importance of identity dimension in the network formation. The standard network formation models used is the related to the literature on non-cooperative network formation models pioneered by Bala and Goyal (2000a) as well as Bala and Goyal (2000b) with related work by Galeotti, Goyal, and Kamphorst (2003), Hojman and Szeidl (2006), Billand et al. (2006), Galeotti (2006). The other strand in this literature is from Jackson and Wolinsky (1996). Bala and Goyal (2000b), Haller and Sarangi (2005) explore the effects of link readability on the network formation. The important deviation of this chapter from this literature is to combine the non-cooperative network formation models with identity and probabilistic link failures. In our model, the links will be formed based on identity characteristics and each link or player can fail with a certain probability. Probabilistic links provide many incentives form different architectures and yields results that corresponds to the empirical observations.

Rest of the chapter is organized as follows: Section 2 illustrates our model setup and provides some useful network definitions. Section 3 describes the Nash networks where identity is exogenously assigned to the players. The chapter concludes with a discussion and possible extensions.

4.2   Model

The set of all players is $N = \{1, 2, ..., n\}$, with generic members i and j. For ordered pairs $(i, j) \in N \times N$, we use the shorthand notation ij. For non-ordered pairs i, j, we use the notation $[ij]$. Throughout the chapter, we assume that $n \geq 3$.

Identity is defined as a single dimension variable which consists of a set of characteristics following Dev (2010) and (2011). Each player's identity, I_i consists of one of the characteris-

tics. For simplicity we assume that there is a single identity and there are two characteristics associated with the identity. For example, race is an identity and possible characteristics are only black and white. Relaxing this assumption yields fairly complicated probability ranges and left as a possible extension.

The identity profile of the population is represented by $n \times 1$ vector $\mathbf{I}$. Identity can be assigned exogenously or it can be a choice variable. For example, race is an exogenous identity characteristic whereas being democrat or republican is a choice characteristic. We define a block which is made up players who share the same type of characteristics. Each player has the following strategy profile:

- Identity (where applicable): player i chooses his identity such that $I_i = \{A, B\}$. We work on the each case separately where identity is exogenous and where it can be a variable of choice. Identity is a very powerful concept to include different aspects to the network formation game. As an illustrative example, in a student network players can be classified as nerds, jorks and burnouts. Or from a different point of view they cane be classified according to the race and culture.

- Commitment, $\theta \in [0, 1]$: Commitment level indicate the devotion of a player to her characteristic. In general, a higher commitment to any characteristic will make players with the same characteristic cheaper but make links more expensive with players who do not have this characteristics. $n \times 1$ vector $\boldsymbol{\Theta}$ represents the commitment profile of the population.

- Link offers: $l_i = \{l_{i1}, ..., l_{ii-1}, l_{ii} = 0, l_{ii+1}, ..., l_{in}\}$ where $l_{ij} \in [0, 1]$. A link between player i and player j is formed if $l_{ij} + l_{ji} \geq c_{ij}$. Let $\mathcal{L} = \{l_1, l_2, ..., l_n\}$. Let L_i be a n dimensional vector such that

$$
L_{ij} = \begin{cases} \frac{l_{ij}}{(l_{ij} + l_{ji})} & \text{if } l_{ij} + l_{ji} \geq 0 \\ 0 & \text{otherwise.} \end{cases}
$$

Let $(\mathbf{I}, \boldsymbol{\Theta}$ and $\mathcal{L})$ be n-dimensional vector that hold the strategy profile of the players in the

network.

- Cost of link formation: We consider two types of link formation. In the fist case, we assume that initiator of the link fully pays the cost of link formation. This scenario explains situations where there is a devoted recruiter in the network who pays all of the cost of link formation. In the second scenario, we assume that cost of link formation is shared between the participants in proportion of their link offers. For both of these scenarios, we define cost of link formation as a function of identity characteristics and commitment levels. So, cost of link formation is player and partner specific as assumed in some cases of Galeotti et al. (2006) but it is not fully heterogenous and defined as a function of identity characteristic and commitment levels. We assume that the players involved in the link share the same identity, it is cheaper to form a link. Also, as the players increase their commitment levels, the cost of a link between these players decreases even further. However, we assume that if the players involved in the link have different identity characteristics and very high commitment levels (ex. $\theta = 1$) to their identities, then the link formation is not profitable for any V or any functional form of the payoffs.

- Link failure: We allow an every existing link in the network to have a reliability probability, $p \in [0, 1]$. The failure of a link can be interpreted as link still exists but fails to transmit information. First, we assume that reliability probability is exogenous and constant. Then, we heterogenize the link reliability by defining a p_{ij}. We assume that $p_{ij} = p_{ji}$. For simplicity and to incorporate heterogenous link reliability probabilities, we introduce 3 different link failure probabilities that arise from the identity characteristics and commitment levels. If player i and $j \in N$ have the same identity characteristic $(I_i = I_j) = A$, we call it as the first case and we assume that all the other players represented by i and j who have the same identity and characteristics have same link reliability probability, p_1. Similarly, if player i and j have the same identity characteristic $(I_i = I_j) = B \neq A$ then we assume that their link failure probability is p_2. If player i and j have different identity characteristic $(I_i \neq I_j)$ then we assume that their link failure probability is p_3. We assume that all link failures are

independent from each other.

- Expected Benefits: We assume that each player has constant and exogenous information of value (V) to the other players. A player can get access to more information by forming links with other players. We assume two way flow (symmetric) and undirected network, which implies that both players participating in the link can access benefits of the link. And we assume that there is no decay. However, as discussed in the link and node failure parts, even though a link is formed, the functionality of a link to transmit information is probabilistic. Without further notice, a link may not function to transmit information. Hence, under the probabilistic link or node failure, the players have an expected benefits from their links. Let $\mu_i(\mathbf{g})$ be the set of players that player i is linked to directly or indirectly and let $\mu_i^d(\mathbf{g})$ be the set of players to whom player i has formed a link and $\mu^{si}(\mathbf{g})$ be the set of players who share the same identity characteristics with player i.

Definition 1: The closure of $\mathbf{g}$ is a non-directed network denoted by $h = cl(\mathbf{g})$ and defined as $cl(\mathbf{g}) = \{ij \in N \times N : i \neq j \text{ and } g_{ij} = 1 \text{ or } g_{ji} = 1\}$.

The benefits from network $\mathbf{g}$ are derived from its closure $h = cl(\mathbf{g})$. For two players $i \neq j$, the non-ordered pair $[ij]$ represents the undirected or both-way link between i and j, i.e. the simultaneous occurrence of ij and ji. If $h_{ij} = h_{ij} = 1$, then $[ij]$ succeeds with probability $p_{ij} \in (0, 1)$ and fails with probability $1 - p_{ij}$ where p_{ij} is not necessarily equal to p_{ik} for $j \neq k$. It is assumed, however, that $p_{ij} = p_{ji}$. Furthermore, the success or failure of direct links between different pairs of players are assumed to be independent events. Thus, h may be regarded as a random network with possibly different probabilities of realization for different edges. To simply these link failure probabilities, we classify them in 3 different categories depending on the identity characteristic and commitment levels of the participants as discussed in link failure above.

We call a non-directed network h' a realization of h (denoted by $h' \subset h$) if it satisfies $h'_{ij} \leq h_{ij}$ for all i, j with $i \neq j$. The notation $[ij] \in h'$ signifies that the undirected link $[ij]$

belongs to h', that is $h'_{ij} = h'_{ji} = 1$. At this point the concept of a path (in h') between two players proves useful.

Definition 2: For $h' \subset h$, a path of length m from an player i to a different player j is a finite sequence $i_0, i_1, ..., i_m$ of pairwise distinct players such that $i_0 = i, i_m = j$, and $h'_{ik,ik+1} = 1$ for $k = 0, ..., m - 1$.

We say that player i observes player j in the realization h', if there exists a path from i to j in h'. We assume that links can fail independently, the probability of the network h' being realized given h is given by:

$$\lambda(h'|h) = \Pi_{[ij \in h']} p_{ij} \Pi_{[ij] \notin h'} (1 - p_{ij}) \tag{4.1}$$

Let $\mu_i(h')$ be the number of players that player i observes in the realization h', i.e. the number of players to whom i is directly or indirectly linked in h'. Each observed player in a realization yields a benefit $V > 0$ to player i. Without loss of generality assume that $V = 1$. Given the strategy profile of players in **g** and link failures, player i's expected benefit from the random network h is given by the following benefit function $B_i(h)$:

$$B_i(h) = \sum_{h' \subset h} \lambda(h'|h) \mu_i(h') \tag{4.2}$$

where $h = cl(\mathbf{g})$. The probability that network h' is realized is $\lambda(h'|h)$, in which case player i gets access to the information of $\mu_i(h')$ players in total. Note that the benefit function is clearly non-decreasing in the number of links for all the players.

- Payoffs: We assume that each link formed by player i costs $c > 0$. Cost of link formation depends on the identity and characteristics and the commitment levels. When the cost of link formation is only on one side the payoff of player i is as follows:

$$\Pi_i(\mathbf{I}, \mathbf{\Theta}, \mathbf{G}) = (B_i(cl(\mathbf{g})) - \sum_{j \neq i \in \mu_i^d} c(I_i, I_j, \theta_i, \theta_j) \tag{4.3}$$

where the expected benefits of player i from the network is indicated by B_i.

Then, player i's expected payoff from the network when the formation of link depends on link offers and mutual consent is given by:

$$\Pi_i(\mathbf{I}, \boldsymbol{\Theta}, \pounds) = \pi(B_i(cl(\mathbf{g})) - c(L_i, \boldsymbol{\Theta}, \mathbf{I}) \tag{4.4}$$

If the composition of the players with the same identity characteristic affects payoffs then:

$$\Pi_i(\mathbf{I}, \boldsymbol{\Theta}, \pounds) = \pi(B_i(cl(\mathbf{g})) - c(L_i, \boldsymbol{\Theta}, \mathbf{I}) \tag{4.5}$$

Note that $L_i, \boldsymbol{\Theta}, \mathbf{I}$ determines the costs of forming links and the last term is introduced when identity characteristic has a direct impact on the payoffs (third scenario of identity). We assume that π_i is a strictly increasing function in N^{si} meaning that as the number o players who share the identity characteristic of player i increases, player i obtains higher payoff.

Given a network $\mathbf{g} \in G$, let g_{-i} denote the network that remains when all of player i's links have been removed. Clearly $g = g_i \oplus g_{-i}$ where the symbol $\oplus$ indicates that $\mathbf{g}$ is formed by the union of links in g_i and g_{-i}.

Assumptions on the Payoff Function

We assume that payoff of player i is linear in expected benefits and costs. We propose the following assumptions on the payoff function:

1. Assumption on the reliability probabilities:

A1) Link and node failure events are all independent from each other meaning that one link or node's reliability does not affect the reliability of other links or nodes.

2. Assumptions on the benefits and costs of link formation:

A2) Π_i is a strictly increasing function of the number of players connected directly or indirectly. Also, if the links are more reliable (i.e, higher p) the expected payoff of player i increases.

A3) $c(I_i, I_j, \theta_i, \theta_j)$ is strictly decreasing in θ_i, θ_j for each $j \in \mu_i$ if $I_i = Ij$. And $c(I_i, I_j, \theta_i, \theta_j)$ is strictly increasing in θ_i, θ_j for each $j \in \mu_i$ if $I_i \neq Ij$. Moreover, we assume that there cannot be a profitable link between player i and j if $I_i \neq I_j$ and $\theta_i = \theta_j = 1$.

3. Assumptions on the link offers (where applicable):

A4) Π_i is a strictly decreasing function of L_i which represents the total link offers made by player i.

A5) Cost sharing in the link formation implies that in order to have the link between player i and j, link ij should be profitable to both player i and player j.

Definition 3: A strategy g_i is said to be a best response of player i to g_{-i} if $\Pi_i(g_i \oplus g_{-i}) \geq \Pi_i(g_i' \oplus g_{-i})$ for all $g_i' \in G_i$. Let $BR_i(g_{-i})$ denote the set of player i's best responses to g_{-i}.

A network $\mathbf{g} = (g_1, ..., g_n)$ is said to be a Nash network if $g_i \in BR_i(g_{-i})$ for each player i, i.e., players are playing a Nash equilibrium. A strict Nash network is one where players are playing strict best responses. Formally,

Definition 4: The Nash Equilibrium is a set of strategies $\Pi(\mathbf{I}, \mathbf{\Theta}, \mathcal{L})$ which result in network $\mathbf{g}$, such that for each player $i \in N$

$$\Pi(\mathbf{I}, \mathbf{\Theta}, \mathcal{L}) \geq \Pi(l_i', \theta_i', I_i')$$

Definition 5: A set $C \subset N$ is called a component of $\mathbf{g}$ if there exists a path in $cl(\mathbf{g})$ between any two players i and j in C and there is no strict superset $C' of C$ for which this holds true.

Definition 6: A connected network $\mathbf{g}$ is said to be minimally connected, if it is no longer connected after the deletion of any link. A network $\mathbf{g}$ is called complete, if all links exist in $cl(\mathbf{g})$. A network with no links is called an empty network.

4.3 Exogenous Identity

In this section, we study the network formation game where identity is assigned exogenously to the players. Players are allowed to choose how much to commit to their identity characteristics. We will first investigate one sided link formation where the cost of link formation is on the initiator of the link. Then, we will explore mutual consent link formation through link offers. We introduce the link failures in two ways. Firstly, we assume that link failure is constant and same for all links even within different characteristic groups, $p \in (0, 1)$. Then, we heterogenize the link failures in three groups as discussed in the model setup section. We characterize the Nash networks under these link imperfections in the following lemmas and propositions.

Lemma 1: Suppose the payoff of player $i \in N$ is given by Equation 3. Then, in a Nash network each identity characteristic group is composed of players choosing $\theta = 1$ if they are directly connected to the players where the identity characteristic is the same and $\theta = 0$ if they are connected to the different identity characteristics.

The proof of the lemma directly follows from the assumptions on the payoff function. Players who choose to make links only within their characteristic will choose $\theta = 1$ while all player's who choose to make links with players outside their characteristic will choose $\theta = 0$.

The proposition below identifies Nash network architectures for an exogenous and constant link failure probability, p_0. If $p_0 = p_0(c, n) = c \times (n-1)^{-1}$ it never benefits player i to initiate a link from i to j, no matter how reliably player j is linked to other players and, therefore, $g_{ij} = 0$ in any Nash equilibrium.

Definition 7: Define the cost of link formation $c_1 = c(I_i = I_j, \theta_i = \theta_j = 1)$ if players involved in the link share the same identity characteristic and highest commitment levels, and $c_2 = c(I_i \neq I_j, \theta_i, \theta_j)$ if the players involved in the link do not have a common identity characteristic. By assumption A3, it follows that $c_2 > c_1$.

Proposition 1: Exogenous Identity and Homogenous Link Failure

Under the assumptions A1-A3, the Nash Network of the game, where players choose their commitment levels and link strategies with the given probabilities will have one of the following structures:

1) Empty if $p_0 < argmin\{c_{ij} \times (n-1)^{-1}\}$

2) Separated, where each identity characteristic forms a component if $p_0 < c_{ij} \times (n-1)^{-1}$

3) Connected if $p_0 \geq \{c_{ij} \times (n-1)^{-1}$

Proof: Suppose **g** is a Nash network and its neither empty or connected. Then, without loss of generality, there exists three players i, j and k such that player i and j belong to one component, $C_1(\mathbf{g})$ and player k belongs to another component, $C_2(\mathbf{g})$. Then either g_{ij} or $g_{ji} = 1$ but g_{ki} and $g_{kj} = 0$. Without loss of generality, assume $g_{ij} = 1$. Define the benefit player i receives by linking to j in **g** as b_1. Since the link ij is made, it must be profitable for i to sponsor the link, so $b_1 > c$. Let $\mathbf{g}'$ be a network where all the links of player i is deleted. Under this condition, player i is an isolated player and the expected benefit for player i by linking to j is $b_2 = p \times (1 + V_j)$ where V_j is the expected benefit player j receives from her closure. In $\mathbf{g}'$ player i has no other direct links and all her indirect links go through player j. Hence, it follows that $b_2 > b_1$. Consider a link from player k to j in $\mathbf{g}'$ the expected benefit of this link for player k is $b_3 = p \times (p + 1 + V_j)$. Notice that $b_3 > b_2$ if $p > c_{ij} \times (n-1)^{-1}$. Hence, both in **g** and $\mathbf{g}'$ its profitable for player k to initiate the link kj. This contradicts the initial assumption that **g** is Nash.

Player i's expected benefit from the link ij is at most $p_0 \times (n-1)$. If $p_1 < p_0$ then, it never benefits the players i and j to form a link ij since the expected benefit of link ij is not high enough to compensate the cost of link formation for any l_{ij} and l_{ji}. If $p_0 > maxc_{ij} \times (n-1)^{-1}$ holds then a Nash network, **g** can either be connected, or separated with identities or empty. Note that due to a general payoff functional form, having $p_0 > maxc_{ij} \times (n-1)^{-1}$ does not guarantee that Nash network will be non-empty. However, if $p_0 < maxc_{ij} \times (n-1)^{-1}$ then Nash network will be guaranteed to be empty.

Proposition 2: Let the payoff of player i be given by (2.3). Given $p \in (0,1)$ there exists $c(p) > 0$ such that in a Nash network, **g** there will be at least more than one path between player i and j if $I_i = I_j$.

Proof: Player i's expected benefit from the link ij is at most $p_0 \times (n-1)$. If $p_1 < p_0$ then, it never benefits the players i and j to form a link ij since the expected benefit of link ij is not high enough to compensate the cost of link formation for any l_{ij} and l_{ji}. If $p_0 > maxc_{ij} \times (n-1)^{-1}$ holds then Nash network non-empty. However, if $p_0 < maxc_{ij} \times (n-1)^{-1}$ then Nash network will be guaranteed to be empty. For a player to make a redundant link, the marginal benefit of the additional link with an observed player must exceed the cost of link formation. Following Bala and Goyal (2000b), if $p \times (1 - p^{n/2}) > c_1$ or c_2 then for player i sponsoring a link to an indirectly connected player j provides an increase in the payoff of player i.

Note that proposition 2 result will only hold if the player i and j belong to the same identity group. If player i and j belong to a different identity group, then in order this result to hold, the cost of link formation between different identity groups must be small enough. Since, we do not assume any particular functional form on the cost of link formation, we are not able to determine the $c(p)$. This limitation can be analyzed if a particular functional form is assumed.

Proposition 3: Let the payoff of player i be given by (2.3). If $p > c$ and $(1 - p) + (n - 2)(1 - p^2) < c$ are satisfied then in a Nash network, each identity group will form a mixed star architecture.

Proof: Suppose **g** is a mixed star network and let player k be the center of the star. Consider a player $i \neq k$ where without loss of generality $g_{ik} = 1$. From Equation (2.3), player i's payoff is $p + (n - 2)p^2 - c$. If player i does not form this link, then she obtains zero payoff from the architecture of a star network. However, since $p > c$ it is profitable for player i to form at least one link. If player i chooses to link another player $j \neq k$ then player i's expected payoff is $p + p^2 + (n - 3)p^3 - c$. Linking with player k dominates this payoff if

$(1-p)+(n-2)(1-p^2) < c$ is satisfied. Hence, player i's optimal strategy is to link with player k if the conditions in the proposition hold. If player i chooses to form more than one link then his payoff is bounded above by $(n-1)-nc$. Subtracting this payoff from the star, the incremental benefit of player i at most $(1-p)+(n-2)(1-p^2)-c$ and this is negative if $(1-p)+(n-2)(1-p^2) < c$. Therefore, player i optimal strategy is to form a single link with player k.

The above proposition states that each identity group will be a derivative of star architecture. However, for the entire network to be connected, $p > argmax\{c(I_i \neq I_j)\}$ must be satisfied. In this case, the Nash network will be an interlink star where one of the periphery player will form a link with opposite identity group.

Proposition 4: Exogenous Identity and Heterogenous Link Failure

Suppose the heterogenous link failures can be classified in 3 types (p_1, p_2 and p_3). Under the assumptions A1-A4, the Nash Network of the game, where players choose their commitment levels and link strategies with the given link failure probabilities will have one of the following structures:

- Connected if $p_1 \geq \frac{1}{1+c_2}p_2$ and $p_2 \geq \frac{1}{1+c_1/n_b^2}p_1$ and $p_3 \geq \frac{1}{1+c/(n_b+1)^2}p_2$ where $c = min\{c_1, c_2\}$.

- Separated if $p_3 < \frac{1}{1+c/(n_b+1)^2}p_2$ where $c = min\{c_1, c_2\}$:

  a) If $p_2 \geq \frac{1}{1+c_1/n_b^2}p_1$ and $p_1 \geq \frac{1}{1+c_2}p_2$ then each block is a connected component.

  b) If $p_2 < \frac{1}{1+c_1/n_b^2}p_1$ and $p_1 \geq \frac{1}{1+c_2}p_2$ then the player with $\theta < 1$ will be isolated in each block.

  c) If $p_2 \geq \frac{1}{1+c_1/n_b^2}p_1$ and $p_1 < \frac{1}{1+c_2}p_2$ then there is only one link in each block, and it is formed between one of the players with $\theta = 1$ and the player with $\theta < 1$. All the other players don't form a link.

- Empty if $p_1 < \frac{1}{1+c_2}p_2$ and $p_2 < \frac{1}{1+c_1/n_b^2}p_1$ and $p_3 < \frac{1}{1+c/(n_b+1)^2}p_2$.

Proof: We prove each case separately. In order to have a connected Nash network, two players with different identities will choose $\theta < 1$ and the others will choose $\theta = 1$ in the equilibrium and all players will access to each other through direct or indirect connections. Note that the level of link failures and the number of players present in each identity type are the main determinants of expected benefits from a generic link ij. The link success parameters, namely p_1, p_2 and p_3 are exogenously assigned for each identity group. Let $c = min\{c_1, c_2, c_3\}$ be minimum cost of link formation in the network.

1) Connected Nash Networks: Consider a Nash network, **g**. Suppose **g** is neither empty nor connected. Then, there exist three agents i, j, and k such that i and j belong to one connected component of $cl(g)$, C_1 and k belongs to a different connected component of $cl(g)$, C_2. Then $g_{ij} = 1$ or $g_{ji} = 1$ whereas $g_{mk} = g_{km} = 0$ for all $m \in C1$. Without loss of generality assume $g_{ij} = 1$. Then, the incremental benefit to i of having the link from i to j is $b_1 - c$. Let g' denote the network which one obtains, if in **g** all direct links with i as a vertex are severed. The incremental expected benefit to i and j of forming the link ij in g' is $b_2 > b_1 > c_{ij}$ and can be written as $b_2 = p_{ij}(1 + V_j)$ where V_j is j's expected benefit from all the links j has in addition to ij. Now consider a link from k to j, given $g' \oplus g_{ij}$. This link is worth $b_3 = p_{kj}(p_{ij} + 1 + V_j)$ to player j and k. A link from k to j, given g, is worth $b_4 > b_3$ to k. We claim that $b_3 > b_2$, i.e., $p_{kj} > p_{ij}\frac{(1+V_j)}{(1+V_j+p_{ij})}$. Since g is Nash and $g_{ij} = 1$, we know $p_{ij} > p_0$. By assumption, $p_{kj} > \frac{1}{1+c/n^2}p_{ij}$. This shows the claim that $b_4 > b_3 > b_2 > b_1 > c$. If $p_1 \geq \frac{1}{1+c_2}p_2$ and $p_2 \geq \frac{1}{1+c_1/n_b^2}p_1$ and $p_3 \geq \frac{1}{1+c/(n_b+1)^2}p_2$ where $c = min\{c_1, c_2\}$ then having the link kj is better for k and j than not having it, contradicting that g is Nash.

2) Separated Nash networks:

a) If $p_2 \geq \frac{1}{1+c_1/n_b^2}p_1$ and $p_1 \geq \frac{1}{1+c_2}p_2$ then each block is a connected component. Then, none of the players will form a link with the opposite identity group and the Nash network is separated.

b)If $p_2 < \frac{1}{1+c_1/n_b^2}p_1$ and $p_1 \geq \frac{1}{1+c_2}p_2$ then the player with $\theta > 1$ will be isolated in each block.

c)If $p_2 \geq \frac{1}{1+c_1/n_b^2}p_1$ and $p_1 < \frac{1}{1+c_2}p_2$ then there is only one link in each block, and it is formed between one of the players with $\theta = 1$ and the player with $\theta > 1$. All the other players don't form a link.

3) Empty Nash networks: Consider $p_1 < \frac{1}{1+c_2}p_2$ and $p_2 < \frac{1}{1+c_1/n_b^2}p_1$ and $p_3 < \frac{1}{1+c/(n_b+1)^2}p_2$.

Next, we study the Nash networks under mutual consent. Mutual consent requires investments from both of the participants of the link and the cost of link formation will be shared among them. The following lemmas indicate a few properties that holds in a Nash network.

Lemma 2: Link offers.

Under the definitions and assumptions of the model, in a Nash network sum of the link offers from player i and j must be exactly equal to the cost of link formation.

Proof: Suppose there are two players i and j such that $l_{ij} + l_{ji} > c_{ij}$. Without loss of generality, player i will be better off by decreasing her link offer until $l_{ij} + l_{ji} = c_{ij}$ so that the link will still be preserved and player i's payoff would strictly increase from A3. Hence, a link offer $l_{ij} + l_{ji} > c_{ij}$ is not a best response strategy.

Lemma 3: Commitment levels of identities in a block.

Under the definitions and assumptions of the model, a block includes players with maximum commitment, i.e. $\theta = 1$ unless expected benefit from connecting to the other identity characteristic exceeds the cost of link formation to the opposite identity group.

Proof: From the definition of the cost of link formation, if players with the same identity choose commitment level $\theta = 1$, the cost of link formation becomes minimal. So, suppose there exists two players i and $j \in N$ such that $\theta_i < 1$ but $\theta_j = 1$. Then, player i can strictly increase her payoff by increasing her commitment. If the Nash network is connected with players with different identities, again from the definition of cost of link formation, there must be player k and m such that $I_k \neq I_m$ and θ_k and $\theta_m < 1$ which enables a link to be formed with different identities. Note that since there is no decay in the model, there will

be exactly 2 players in this situation which makes the network connected. Also, there will be exactly two players i and j with $\theta_i = \theta_j = 1$ connecting to players k and m. However, to have a connection between different identities, the benefits from linking to different identities must be high enough to offset this increase in the cost of link formation.

When the cost of link formation is shared between the participants of the link, we assume that if a link is made between player i and j then it must be profitable for each player.

Remark 1: Exogenous Identity, Homogenous Link Failure and Cost Sharing

Under the assumptions A1-A4, the Nash Network of the game, where players choose their commitment levels and link strategies with the given probabilities will have one of the following structures:

1) Empty if $p_0 \leq maxc_{ij} \times (n-1)^{-1}$ 2) Separated, where each block will either form a component or each player of the block remains a singleton if

3) Connected if $p_0 \geq maxc_{ij} \times (n-1)^{-1}$

If we expand the heterogeneity of link reliability to the broader terms, calculations are fairly complicated. We can say that given $p_{ij} \neq p_{km}$ there exists cost values such that any network can be a Nash equilibrium. Intuitively, imagine for a group of players the reliability is lower than the cost of link formation. Then, we can have disconnected components in a Nash network. Furthermore, if the link reliability between different identity groups are higher than the cost of link formation it is possible to observe players belonging to different identity characteristic to link each other. Also, the minimality result in Nash networks may not hold. In short, the optimal strategy for player i becomes a function of number of players in each characteristic and how fast the cost of link formation increases. We leave this as a future work.

4.4 Discussion and Extensions

In this chapter, we introduce link reliability to a non-cooperative network formation game where identity characteristics are capturing the similarities or differences between the players in the network. Compared to Bala and Goyal (2000b) framework, our model is includes an identity dimension. If the link reliability is the same for all players in the network then Bala and Goyal (2000b) results generally hold. However, once the link reliability is different between the different identity groups, Nash networks can include more than one component if the cost of link formation is too high between the different identity characteristics. Depending on the link reliability and cost of link formation, it is possible to observe very decentralized architectures and connectivity of the network greatly depends on how the cost of link formation increases between different identity groups. Compared to Dev (2009) and (2010), we find significant differences in the Nash networks if link reliability between the different identity groups are high and cost of link formation is low enough. Dev (2010) shows how the Nash networks can be fragmented according to the identity characteristics. In this chapter, we show that players belonging to different identity groups can choose to link if the reliability within their own group is lower than the reliability of the opposite group. Another difference is that Dev (2010) minimality result may not hold as the number of players in the network increases and the link reliability is low enough.

Our results in the chapter are preliminary and a number of extensions can be considered for future work. One possible extension is to consider identity as a choice variable rather than exogenous to the model. In this scenario, the network formation can be modeled as a dynamic game with rounds. In the first round, an arbitrary player chooses his identity and commitment level and makes link offers. In the other rounds, each player decides whether or no to accept the link offer.

Another possible extension is to consider endogenic link reliability and make it as a function of the link investment under this scenario, players can increase the link success probability if they bear higher cost of link formation. This setting is important to explain

star type architectures since the links are critical for the players in the periphery.

It is possible to consider player reliability, f in the sense that each player can leave the network without any further notice. In this case, if one of the players leave the network, all the links she is involved in will be lost.

Chapter 5

Conclusion

My dissertation contributes to the social network literature by investigating interesting questions. My first chapter introduces a Team Game, develops a measure for identifying the key player in teams. My second chapter explores the socially optimal network architectures and investigates the differences between Nash and efficient networks. My final chapter considers a non-cooperative network formation game where identity is introduced to capture the similarities or the differences among the players and the link realisability is probabilistic.

My first chapter extends the intercentrality measure of BCZ (2006) to include an additional term which captures the team outcome expression in the utility functions of players. This term suggests that a player gets utility when her team achieves its desired outcome. To identify the contribution of players to their teammates, we develop two intercentrality measures which derives from possible considerations of the social planner. $TICM$ considers the effect of a player's removal on the aggregate Nash Equilibrium effort levels. $TICM^e$ identifies the externality each player gets from her teammates and weights it according to the ability of the player. In the empirical part of the chapter, first we illustrate how to utilize the intercentrality measures. Then, we show that there is a positive relationship between the average ratings and $TICM^e$ and ICM in the sample. This fact reflects that soccer players having more interactions with their teammates get more credit in performance by the experts. Moreover, the market value of the soccer players increase with both $TICM^e$ and ICM which is assumed to be reflected in their salaries. This effect is homogenous in the

sample, it doesn't depend on the position of the player on the field. One interesting extension of the approach in the chapter might be considering the effort variable to be a vector and allowing different types of individual actions. This will require a new set of theocratical results. Depending on the availability of data this model then can be empirically tested. In soccer, for instance one could include distance traveled, tackling and dribbling data. Given the relationship between passing and scoring opportunities, this way will not alter our primary results, but will provide us a more precise way to identify key players and key groups. Another interesting extension to our model would be to investigate key player problem as a network design game. The planner is the head coaches who have to announce the national squads. There are qualities, δ_i's and possible interaction possibilities between players. This can be modeled as an expected utility maximization problem with a two stage team game. At the first stage, squads are announced and at the second stage players optimize their effort with given interactions.

My second chapter studies the efficient networks in the player and partner heterogeneity models with various functional forms. With a two way flow of information, I find that payoff function specification, the differences in cost between minimum cost player(s) and the other players and the number of minimum cost players play a crucial role in determining the architecture of the efficient networks. I deal with the first two determinants by imposing additional restrictions if necessary. Note that having more than one minimum cost players introduces some symmetries and additional efficient network architectures. However, the efficient network architectures in the case of having more minimum cost players can still be identified without imposing any restrictions unlike the other two determinants. There are some notable differences between the efficient architectures under the player and partner heterogeneity models. The first difference is in the player heterogeneity model minimum cost player(s) sponsor all the links. However, in partner heterogeneity a high cost player can also sponsor some links. In the homogenous model discussed by Bala and Goyal (2000), efficient networks have star architecture. However, when we allow for heterogeneity in values

and costs, a rather decentralized efficient architecture in the form of interlinked stars occurs in linear and strictly convex payoffs in cost. In the one way flow model, the architecture of the efficient network with player and partner heterogeneity is identical. The only connected efficient architecture is a wheel for these cases, where each agent can access to all other agents by sponsoring a single link. I conclude that strict Nash networks in one way flow model discussed by Galleotti (2006) and Billand et al. (2011) are also efficient. I find that Nash and efficient networks coincide if the payoff function is linear in benefits and costs. This holds true for one way flow of information and some cases in the two way flow of information. However, we show that in two way flow of information this result is not robust when payoff function exhibits strict convexity in cost. Also, with the linear specification efficient networks can also be interlinked stars while only star type architectures are Nash networks. Note that not all of the payoff function specifications are comparable in terms of Nash and efficient networks since I impose additional requirements to identify link formation in some cases.

In my last chapter, we introduce link reliability to a non-cooperative network formation game where identity characteristics are capturing the similarities or differences between the players in the network. Compared to Bala and Goyal (2000b) framework, our model is includes an identity dimension. If the link reliability is the same for all players in the network then Bala and Goyal (2000b) results generally hold. However, once the link reliability is different between the different identity groups, Nash networks can include more than one component if the cost of link formation is too high between the different identity characteristics. Depending on the link reliability and cost of link formation, it is possible to observe very decentralized architectures and connectivity of the network greatly depends on how the cost of link formation increases between different identity groups. Compared to Dev (2009) and (2010), we find significant differences in the Nash networks if link reliability between the different identity groups are high and cost of link formation is low enough. Dev (2010) shows how the Nash networks can be fragmented according to the identity characteristics. In this

chapter, we show that players belonging to different identity groups can choose to link if the reliability within their own group is lower than the reliability of the opposite group.

Bibliography

[1] Akerlof, G. A., and R. E. Kranton (2000): "Economics And Identity," *The Quarterly Journal of Economics*, **115**, 715-753.

[2] Bala, V., Goyal, S. (2000a): "A Non-Cooperative Model of Network Formation," *Econometrica*, **68**, 1181-1229.

[3] Bala, V., Goyal, S. (2000b): "A Strategic Analysis of Network Reliability," *Rev. Econ. Design*, **5**, 205-228.

[4] Ballester C., Calvo-Armengol, A. and Zenou, Y. (2004): " Who's Who in Crime Networks. Wanted: the Key Player," CEPR Discussion Paper No. 4421

[5] Ballester, C., Calvo-Armengol, A. and Zenou, Y. (2006): "Who's Who In Networks. Wanted: The Key Player," *Econometrica*, **74**, 1403-1417.

[6] Ballester, C., Calvo-Armengol, A. and Zenou, Y. (2010): "Delinquent Networks," *Journal of the European Economic Association*, **8**, 34-61.

[7] Battre, M., Deutscher, C. and Frick, B. (2008): "Salary Determination in the German "Bundesliga: A Panel Study," No 0811, *IASE Conference Papers*, International Association of Sports Economists.

[8] P. Billand, C. Bravard and Sarangi, S. (2008): "Existence of Nash Networks in One-way Flow Models," *Economic Theory*, **37**, 491-507.

[9] Billand, P., Bravard, C. and Sarangi, S. (2010): "Strict Nash Networks and Partner Heterogeneity," *forthcoming* in Int. J. Game Theory.

[10] Billand, P., Bravard, C. and Sarangi, S. (2011): "Heterogeneity and Link Imperfections in Nash Networks," *International Game Theory Review*, **13**, 181-194.

[11] Bisin, A., and T. Verdier (2000): "Beyond The Melting Pot: Cultural Transmission, Marriage, And The Evolution Of Ethnic And Religious Traits," *The Quarterly Journal of Economics*, **115**, 955-988.

[12] Bonacich, P. (1987): "Power and Centrality: A Family of Measures," *American Jounal of Sociology*, **92**, 1170-1182.

[13] Borgatti, S. P. (2006): "Identifying Sets of Key Players In A Social Network," *Computational and Mathematical Organization Theory*, **12**, 21-34.

[14] Calvo-Armengol, A., Patacchini, E. and Zenou, Y. (2009): "Peer Effects and Social Networks in Education," *Review of Economic Studies*, **76**, 1239-1267.

[15] Chandra, K. (2001): "Ethnic Bargains, Group Instability and Social Choice Theory," *Politics and Society*, **29**, 337-362.

[16] Chen Y., Paul, G., Cohen, R., Havlin, S. (2007): "Percolation Theory and Fragmentation Measures in Social Networks," *Physica*, **378**, 11-19.

[17] Currarini, S., M. O. Jackson, and P. Pin (2008): "An Economic Model of Friendship: Homophily, Minorities and Segregation," *Econometrica*, **77**, 1003-1045.

[18] Debreu, G., and Herstein, I.N. (1953): "Nonnegative Square Matrices," *Econometrica*, **21**, 597-607.

[19] De Marti, J. and Zenou, Y. (2009): "Ethnic Identity and Social Distance in Friendship Formation," CEPR Discussion Paper No 7566.

[20] Dev, P. (2009): "Identity and Fragmentation in Networks," Mimeo, ITAM.

[21] Dev, P. (2010): "Choosing Me and My Friends: Identity in a Non-Cooperative Network Formation Game with Cost Sharing," Mimeo, ITAM.

[22] Duch, J., Waitzman, J.S. and Amaral, L.A.N. (2010): "Quantifying the Performance of Individual Players in a Team Activity," PLoS ONE 5(6): e10937. doi:10.1371.

[23] Everett, M. G. and Borgatti, S. P. (1999): "The Centrality of Groups and Classes," *Journal of Mathematical Sociology*, **23**, 181-201.

[24] Everett, M. G. and S. P. Borgatti (2006): "A Graph-Theoretic Perspective on Centrality," *Social Networks*, **28**, 466-484.

[25] Freeman, L. C. (1977): "A Set of Measures of Centrality Based on Betweenness," *Sociometry*, **40**, 35-41.

[26] Freeman, L. C. (1979): "Centrality In Social Networks: Conceptual Clarification," *Social Networks*, **1**, 215-239.

[27] Frick, B. (2007): "The Football Players' Labor Market: Empirical Evidence from the Major European Leagues," *Scottish Journal of Political Economy*, **54**, 422-446.

[28] Galeotti, A. (2004): "One-way Flow Networks: The Role of Heterogeneity," Tinbergen Institute Discussion Paper No. 2004-031/1.

[29] Galeotti, A., (2006): "One-way Flow Networks: The Role of Heterogeneity," *Economic Theory*, **29**, 163-179.

[30] Galeotti A, Goyal S, Kamphorst J. (2006): "Network Formation with Heterogeneous Players," *Games Economic Behavior*, **54**, 353-372.

[31] Haller, H. and Sarangi, S. (2005): "Nash Networks with Heterogenous Links," *Mathematical Social Sciences*, **50**, 181-201.

[32] Hubbell, C. H. (1965): "An Input-Output Approach to Clique Information," *Sociometry*, **28**, 377-399.

[33] Hojman, D. A., and A. Szeidl (2008): "Core and Periphery in Networks," *Journal of Economic Theory*, **139**, 295-309.

[34] Jackson, M. and Wolinsky, A. (1996): "A Strategic Model of Social and Economic Networks," *Journal of Economic Theory*, **71**, 44-74.

[35] Katz, L. (1953): "A New Index Derived From Sociometric Data Analysis," *Psychometrika*, **32**, 241-254.

[36] Luhtanen, P., Belinskij, A., Häyrinen, M., and Vänttinen T. (2001): "A Comparative Tournament Analysis Between the EURO 1996 and 2000 In Soccer." *International Journal of Performance Analysis in Sport*, **1**, 74-82.

[37] Luxbacher, J. "Soccer: Steps to Success," *Sports and Recreation*, 2005, 23.

[38] Miller, A. and Wingert, N. "Winning Soccer," H. Regnery Co., 1975, 62.

[39] Sabidussi, G. (1966): "The Centrality Index of a Graph," *Psychometrica*, **31**, 581-603.

[40] Sade, D. S. (1972): "Sociometric of Macaca Mulatta I: Linkages and Cliques in Grooming Networks," *Folia Primatologica*, **18**, 196-223.

[41] Stephenson, K. and Zelen, M. (1989): "Rethinking Centrality: Methods and Examples," *Social Networks*, **11**, 1-37.

[42] Temurshoev, U. (2008): "Who's Who In Networks. Wanted: The Key Group," Net Institute Working Paper 08-08.

Vita

Emre Unlu was born in Eskisehir, Turkey. In August, 2007, he joined Louisiana State University to pursue his graduate studies in economics after completing his education in Turkey. In the process, he received a Master of Science degree in August, 2009. His research interests include social networks and applied microeconomics. Currently, he is a candidate for the degree of Doctor of Philosophy in economics at Louisiana State University.