Identity and Link Reliability in Social Networks

Identity and Link Reliability in Social Networks *

Pritha Dev, Sudipta Sarangi, Emre Unlu

May, 2014

Abstract

This paper 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. 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 imperfections to the model. 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.

Keywords: Social Networks, Identity, Link Reliability

JEL Classification: C72, D85

*We would like to thank Christophe Bravard, Pascal Billand, Francis Bloch, Surajit Bortokokey, Sudipta Sarangi, Matt Wiser for very helpful discussions and comments.
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 suggest that one of the key determinants in the network formation is the similarities in the characteristics of the players. In this paper, 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 reliability into the network formation model. Link 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 reliability probability, \( p \). Then, we relax this assumption heterogenize the link reliability and allow pairwise link reliability, \( p_{ij} \). 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 reliability, 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.

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 paper 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.
Rest of the paper 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 paper concludes with a discussion and possible extensions.

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 paper, 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 characteristics. 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 $I$. 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 can 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 characteristic. $n \times 1$ vector $\Theta$ represents the commitment profile of the population.

- Link offers: $l_i = \{l_{i1},...,l_{i(i-1)},l_{ii} = 0, l_{i(i+1)},...,l_{in}\}$ 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 $(I, \Theta$ 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 first 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 reliability: We allow an every existing link in the network to have a reliability probability, $p \in [0, 1]$. The reliability 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 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 probability is $p_2$. If player $i$ and $j$ have different identity characteristic ($I_i \neq I_j$) then we assume that their link probability is $p_3$. We assume that all link probabilities 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. Let $\mu_i(g)$ be the set of players that player $i$ is linked to directly or indirectly and let $\mu^d_i(g)$ be the set of players to whom player $i$ has formed a link and $\mu^{si}(g)$ be the set of players who share the same identity characteristics with player $i$.

**Definition 1:** The closure of $g$ is a non-directed network denoted by $h = cl(g)$ and defined as $cl(g) = \{ij \in N \times N : i \neq j \text{ and } g_{ij} = 1 \text{ or } g_{ji} = 1\}$. The benefits from network $g$ are derived from its closure $h = cl(g)$. 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 $h_{ij} = h_{ji} = 1$, 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 reliability 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 $h'_{i_k,i_{k+1}} = 1$ 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:

$$\lambda(h' | h) = \Pi_{[ij] \in h'} p_{ij} \Pi_{[ij] \not\in h'} (1 - p_{ij})$$

(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 $B_i(h)$:

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

(2)

where $h = cl(g)$. 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:

$$
\Pi_i(I, \Theta, G) = (B_i(cl(g)) - \sum_{j\neq i \in \mu^d_i} c(I_i, I_j, \theta_i, \theta_j)
$$

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:

$$
\Pi_i(I, \Theta, \ell) = \pi(B_i(cl(g)) - c(L_i, \Theta, I)
$$

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

$$
\Pi_i(I, \Theta, \ell) = \pi(B_i(cl(g)) - c(L_i, \Theta, I)
$$

Note that $L_i, \Theta, I$ 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 $\pi_i$ is a strictly increasing function in $N^si$ meaning that as the number of players who share the identity characteristic of player $i$ increases, player $i$ obtains higher payoff.

Given a network $g \in G$, 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 $g$ 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 success and 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) $\Pi_i$ 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 $\theta_i, \theta_j$ for each $j \in \mu_i$ if $I_i = I_j$. And $c(I_i, I_j, \theta_i, \theta_j)$ is strictly increasing in $\theta_i, \theta_j$ for each $j \in \mu_i$ if $I_i \neq I_j$. 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) $\Pi_i$ is a strictly decreasing function of $L_i$ which represents the total link offers made by player $i$.

**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 $g = (g_1, \ldots, g_n)$ 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 $\Pi(I, \Theta, \mathcal{L})$ which result in network $g$, such that for each player $i \in N$

$$\Pi(I, \Theta, \mathcal{L}) \geq \Pi(I'_i, \theta'_i, I'_i)$$

**Definition 5:** A set $C \subset N$ is called a component of $g$ if there exists a path in $cl(g)$ 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 $g$ is said to be minimally connected, if it is no longer connected after the deletion of any link. A network $g$ is called complete, if all links exist in $cl(g)$. A network with no links is called an empty network.
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 reliability 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 reliability 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 Reliability

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 \( p_0 < \arg\min\{c_{ij} \times (n-1)^{-1}\} \)
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, \( C_1(g) \) and player \( k \) belongs to another component, \( C_2(g) \). 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 \( g' \) 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 \( g' \) 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 \( g' \) 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 \( g' \) its profitable for player \( k \) to initiate the link \( k_j \). 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 $p_0 > max_{c_{ij}} \times (n-1)^{-1}$ holds then Nash network non-empty. However, if $p_0 < max_{c_{ij}} \times (n-1)^{-1}$ 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
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, $p > \arg\max\{c(I_i \neq I_j)\}$ 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 Reliability

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 $p_2 \geq \frac{1}{1+c_1/n_b} p_1$ and $p_3 \geq \frac{1}{1+c/(n_b+1)} p_2$ 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 $p_2 \geq \frac{1}{1+c_1/n_b} p_1$ and $p_1 \geq \frac{1}{1+c_2} p_2$ then each block is a connected component.

  b) If $p_2 < \frac{1}{1+c_1/n_b} p_1$ 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 $p_2 \geq \frac{1}{1+c_1/n_b} p_1$ 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 $p_2 < \frac{1}{1+c_1/n_b} p_1$ 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 reliability 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 $m \in C_1$. 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., $p_{kj} > p_{ij} \left( \frac{1}{1+ c/n} \right)$. Since $g$ is Nash and $g_{ij} = 1$, we know $p_{ij} > p_0$. By assumption, $p_{kj} > \frac{1}{1 + c/n} p_{ij}$. This shows the claim that $b_4 > b_3 > b_2 > b_1 > c$. If $p_1 \geq \frac{1}{1 + c/n} p_2$ and $p_2 \geq \frac{1}{1 + c/n} p_1$ and $p_3 \geq \frac{1}{1 + c/(n_k + 1)} p_2$ 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 $p_2 \geq \frac{1}{1 + c_1/n_k} p_1$ 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 $p_2 < \frac{1}{1 + c_1/n_k} p_1$ 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 $p_2 \geq \frac{1}{1 + c_1/n_k} p_1$ 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 $\theta_k$ 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.

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 Discussion and Extensions

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 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 paper, 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 paper 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 not 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.

5 References


