Credit Scoring with Social Network Data

Yanhao Wei   Pinar Yildirim   Christophe Van den Bulte   Chrysanthos Dellarocas*

Abstract

Motivated by the growing practice of using social network data in credit scoring, this study analyzes the impact of using network based measures on customer score accuracy and on tie formation among customers. We develop a series of models to compare the accuracy of customer scores obtained with and without network data. We also investigate how the accuracy of social network based scores changes when individuals can strategically modify their social networks to attain higher credit scores. We find that, if individuals are motivated to improve their scores, they may form fewer ties with more similar partners. The impact of such endogenous tie formation on the accuracy of consumer credit scores is ambiguous. Scores can become more accurate as a result of modifications in social networks, but this accuracy improvement may come at the cost of more fragmented social networks. The threat of social exclusion in such endogenously formed networks provides incentives to low type members to exert effort that improves everyone's creditworthiness. We discuss implications for both managers and public policy.

Keywords: social networks, credit score, customer scoring, social status, endogenous tie formation

---

*Wei: Doctoral Candidate, Department of Economics, University of Pennsylvania. Yildirim: Assistant Professor of Marketing, The Wharton School, University of Pennsylvania. Van den Bulte: Professor of Marketing, The Wharton School, University of Pennsylvania. Dellarocas: Professor of Information Systems, School of Management, Boston University. We thank Winnie Cheng for making us aware of network based credit scoring practices. We benefited from comments by Lisa George, Upender Subramanian, and audience members at the 2014 Columbia-NYU-Wharton-Yale Four School Marketing Conference, the 2014 Marketing Science Conference, and the 12th ZEW Conference on Information and Communication Technologies at the University of Mannheim. We gratefully acknowledge financial support from the Rodney L. White Center for Financial Research and the Social Impact Initiative of the Wharton School of the University of Pennsylvania.

Noscitur ex sociis.

Latin proverb

You are known by the company you keep.

English proverb

1 Introduction

When a consumer applies for credit, attempts to re-finance a loan, or wants to rent a house, potential lenders often seek information about the applicant's financial background in the form of a credit score provided by a credit bureau or other analysts. A consumer's score can influence both the lender's decision to extend credit and the terms of the credit. In general, consumers with high scores are more likely to obtain credit, and to obtain it with better terms, including the annual percentage rate (APR), the grace period, and other contractual obligations of a loan (Rusli, 2013). Given that people use credit for a range of undertakings that affect social and financial mobility, such as purchasing a house, starting a business, or obtaining higher education, credit scores have a considerable impact on the access to opportunities and hence on social inequality among citizens.

Until recently, assessing a consumer's creditworthiness relied solely on his or her financial history. The financial credit score popularized by Fair, Isaac and Corporation (FICO), for example, relies on three key data to determine access to credit: consumers' debt level, length of credit history, and regular and on-time payments. Together, these elements account for about 80% of the FICO score. Within the past few years, however, the credit scoring industry has witnessed a dramatic change in data sources (Chui, 2013; Jenkins, 2014). In order to assess a consumer's creditworthiness, an increasing number of firms today rely on network based data. One such company, Lenddo, is reported to assign credit scores based on information in users' social networking profiles, such as education and career data, who they are friends with, information available about friends, and how many followers they have (Rusli, 2013). Similar to Lenddo, a growing number of start-ups specialize in using data from social networks. Such firms claim that their social network based credit scoring and financing practices broaden opportunities for a larger portion of the population and may benefit low-income individuals who would otherwise find it hard to obtain credit.

Our study is motivated by the growing use of such practices and investigates whether a move to network based credit scoring affects financing inequality. In particular, we address the following questions. First, from the perspective of lenders, is there an advantage to using network based measures rather than measures based on an individual's data? Second, as the use of social network data becomes common practice, how may consumers' endogenous network formation influence the accuracy of credit scores? Third, how does peer pressure work in network based credit scoring? Finally, and most importantly for public policy, how do these scores influence inequality in access to financing?

1.1 Main insights

Access to financing is correlated with one's credit score. Following Demirgüç-Kunt and Levine (2009), we assume that credit scores can influence access to financing at both the extensive and intensive margins, i.e., by increasing the number of individuals who are considered eligible for financing as well as by providing access to credit at better terms. Although network based scoring can affect access to financing at the extensive and intensive margin, the impact on each might be uneven for different segments of society.

We first develop a model with continuous risk types incorporating network based data (Section 2). Under the assumption of homophily, the notion that people are more likely to form social ties with others who are similar to them, we show that network data provide additional information about individuals and reduce the uncertainty about their creditworthiness. We find that the accuracy of network based scores is dependent mostly on information from the direct ties, i.e., the assessed consumers' ego-network. This implies that credit-scoring firms can assess an individual's creditworthiness efficiently using data from a subset of the overall network.

In Section 3, we extend our model to allow consumers in a network to form ties strategically to improve their credit scores. We find that they may then choose to drop friends with lower scores. This can result in social fragmentation within a network: individuals with better access to financing opportunities choose to segregate themselves from individuals with worse financing opportunities. As a result, individuals self-select into highly homogeneous yet smaller sub-networks. The impact of such social fragmentation on credit scoring accuracy is ambiguous. On the one hand, scores may more accurately reflect borrowers' risk as each agent will be located in a more homogeneous ego-network. On the other hand, scores may become less accurate because smaller ego-networks provide fewer data points and hence less information on each person. How important financial scores are relative to social relationships determines whether strategic tie formation improves or harms credit score accuracy. When accuracy declines, network based scoring could put deserving individuals with low financing opportunities in further hardship. This result supports concerns about social credit scoring from consumer advocates and regulators like the Consumer Financial Protection Bureau and the Federal Trade Commission (Armour, 2014).

In Sections 2 and 3, we study environments where *all* individuals in the society, independent of their type, have similar needs for financing. We relax this assumption in Section 4 where we introduce a formulation with discrete risk types that vary in their needs for financing. That model recognizes that financially strong consumers may have outside options for financing and so may need to rely less on their credit score than financially weaker consumers. When studying this environment, we pay particular attention to the strategic formation of social ties. An important result is the emergence of social exclusion or discrimination *among* low type individuals. They avoid associating with one another, because such associations signal even more strongly to lending institutions that their type is low. Such within group discrimination is a distinct phenomenon from between-group discrimination studied by others (e.g. Arrow, 1998; Becker, 1971; Phelps, 1972).

In Section 5, again within a discrete setting, we allow individuals to exert effort to improve their true creditworthiness or 'type'. We find that when there are complementarities between the effort exerted by individuals, both within and between-group connections motivate effort and thus may lead to increased social mobility, though between-group ties have a more direct and significant effect. Thus, social credit scoring may benefit individuals with poor financial health in two ways: Not only by letting them benefit from a positive signal from social ties with others having a stronger financial footing, but also by motivating them to invest more in their own financial health.

1.2 Related literature

Though motivated by and couched in terms of social credit scoring, the insights we develop go beyond that realm. Our models involve a relatively abstract notion of customer attractiveness or 'type' that has two properties: (1) social relationships are homophilic with respect to types and (2) a third party like a firm or society at large values higher types more and bestows some rewards (external to social relationships) that are monotonically increasing with one's type. The notion of homophily in customer value, i.e., the notion that attractive prospects or customers are more likely to be connected to one another than to unattractive ones, and vice versa, underlies social customer scoring in predictive analytics (e.g., Goel and Goldstein, 2013; Haenlein, 2011). It also is the basis for targeting friends and other network connections of valuable customers in new product launch (e.g., Haenlein and Libai, 2013; Hill et al., 2006), e.g., in targeted online advertising (Bagherjeiran et al., 2010; Bakshy et al., 2012; Liu and Tang, 2011), and customer referral programs (e.g., Kornish and Li, 2010; Schmitt et al., 2011). The basic insights also apply to employment settings, where firms are exploring the use of social network data to gain more information about a job candidate's character and work ethic (e.g., Roth et al., 2013).

The model construct that we label 'social credit score' actually captures a customer's attractiveness or type as perceived by a firm based on social network information, in which the firm bestows some benefits that are monotonically increasing with type. Hence, our insights about social credit scoring can also be interpreted as pertaining to consumers' social status more broadly, i.e., their "position in a social structure based on esteem that is bestowed by others" (Hu and Van den Bulte, 2014). As such, our analysis involving endogenous tie formation adds not only to research traditions in economics and sociology (e.g., Ball et al., 2001; Podolny, 2008) but also to the recent work in marketing on how status considerations affect consumers' networking behavior (Lu et al., 2013; Toubia and Stephen, 2013) and their appeal as customers (Hu and Van den Bulte, 2014).

Even when limited to the realm of financial credit scoring, our analysis relates to several streams of recent work. First is the large and growing amount of work on micro-finance and, more specifically, how group lending helps improve access to capital by reducing the negative consequences of information asymmetries between creditor and debtor (e.g., Ambrus et al., 2014; Bramoullé and Kranton, 2007a,b; Stiglitz, 1990; Townsend, 1994). Our analysis focuses on individual loans rather than group loans, and on a priori customer scoring rather than a posteriori compliance through group monitoring and social pressure. Hence, our result that social credit scoring can lead people to change their network ties and to exert more effort in improving their financial health is different from yet dovetails with the evidence by Feigenberg et al. (2010) that group lending tends to trigger changes in network structure that in turn reduce loan defaults. Hence, two different kinds of "social financing" practices acting at two different stages of the loan (customer selection and terms definition vs. compliance) can both lead to improved outcomes mediated through endogeneous changes in network structure.

Second, we provide new insights on the risk of discrimination and exclusion triggered by social financing (Ambrus et al., 2014; Armour, 2014). Our model allows for the possibility of discrimination against less creditworthy individuals. There are two ways through which such discrimination can come about. The first is that individuals may be subject to discrimination based on type. In an endogenous network, borrowers will be more selective in forming relationships, and may prefer to form relationships with higher-type individuals to protect their credit score. Formation of networks in order to attain a high credit score can be an indirect way of discrimination because some individuals are systematically excluded from others' networks. The second is that individuals may observe each other's effort to improve score and discriminate based on personal effort. Any low-type individual who does not exert effort may face disengagement by contacts who do exert effort and who want to disassociate their own credit score from his.

Third, our work is also relevant to ongoing debates on the impact of new social technologies on social integration versus balkanization. Rosenblat and Mobius (2004) find that a reduction in communication costs decreases the separation between individuals but increases the separation between groups. Along similar lines, van Alstyne and Brynjolfsson (2005) find that the internet can lead to segregation among different types of individuals. In this study, we identify conditions under which network based credit scoring (and customer valuation in general) may foster or harm integration within vs. between groups.

Finally, our work will be of topical interest to the growing number of scholars seeking to better understand consumers' financial behaviors, especially the role of homophily (Galak et al., 2011) and trust signaling (e.g., Herzenstein et al., 2011; Lin et al., 2013) in gaining access to credit. It will also be of interest to researchers focusing on the practices in emerging economies where consumer finance and access to credit are particularly important yet the traditional credit scoring apparatus is still found lacking. Creditors, in these markets, often seek to enrich scores based on individual's history with additional information (e.g., Guseva and Rona-Tas, 2001; Rona-Tas and Guseva, 2014; Sudhir et al., 2014).

The rest of the article develops as follows. In Section 2, we present a benchmark model of data collection from networks to assess one's creditworthiness, and then provide justification for the emergence of this industry. In Section 3, we investigate the possibility of networks forming endogenously to the social credit scoring practice. We extend our model to allow individuals to vary in their financing needs in Section 4. We consider the possibility of social mobility through effort in Section 5. In Section 6 we conclude with implications for public policy and managers.

2 Model with Exogenous Network

Consider a society with a large population S of individuals. Each individual i is represented with a type x_i, and x_i follows $N(0, q^{-1})$ across individuals, with precision $q > 0$. We assume that each agent knows his own type.

The process of forming friendships is specified as follows. Each pair of individuals meet with a very small independent probability of $\nu > 0$. Between i and j there is an independent match value $m_{ij} \sim \chi_2$. A friendship between i and j creates utility $m_{ij} - |x_i - x_j|$ for either individual. So, our model features homophily based preference rather than opportunity (Zeng and Xie, 2008): Individuals enjoy the company of others like them more than the company of others unlike them. Person i accepts the formation of a friendship tie with j iff they have met and:

$$m_{ij} > |x_i - x_j|. \tag{1}$$

Upon mutual consent of both parties, a friendship tie is created. The assumption of a χ_2 distribution implies that the probability i and j become friends upon meeting is:

$$Pr(m_{ij} > |x_i - x_j|) = e^{-|x_i - x_j|^2/2}. \tag{2}$$

Let G denote the set of friendships (ties) in society and n_i denote the number of friends of i, or, the *degree* of i under G. The expected number of friends for i is $\mathbb{E}(n_i|x_i) = S\nu\sqrt{\frac{q}{q+1}}e^{-\frac{q}{1+q}{x_i}^2/2}$.[1] In order to represent an environment with sufficient uncertainty about creditworthiness of individuals, we make three assumptions: (i) the society is large ($S \to +\infty$), (ii) the probability that any pair of individuals meet is very small ($\nu \to 0$), and (iii) types are diffuse ($q \to 0$.) These three properties

[1] $\mathbb{E}(n_i|x_i) = S\int_{-\infty}^{+\infty} \nu e^{-(t-x_i)^2/2}\sqrt{\frac{q}{2\pi}}e^{-qt^2/2}dt = S\nu\sqrt{\frac{q}{q+1}}e^{-\frac{q}{1+q}{x_i}^2/2}$.

characterize a society with sufficient uncertainty about individuals. They also allows us to assume that the product term $S\nu\sqrt{\frac{q}{q+1}}$ holds a constant, which we denote by N. [2]

Suppose that friendships in the society have been formed. The lender is interested in updating its information about the types of individuals using signals collected from the network. For any individual i, the lender may observe a noisy signal y_i about his type:

$$y_i = x_i + \varepsilon_i \tag{3}$$

where $\varepsilon_i \sim N(0, 1/c)$ and is independent across individuals. The firm observes the signals of a finite set of individuals denoted by $\mathbf{y}$, which we refer to as the vector of signals as well. For these individuals, the firm may observe the presence or absence of a tie. We use $g \equiv (g^1, g^0)$ to denote such information. Specifically, g^1 is the set of the dyads which the lender knows are friends, and g^0 is the set of the dyads which the lender knows are not friends. Furthermore, for each person in $\mathbf{y}$, we allow g^0 to include all the dyads that involve him and someone outside $\mathbf{y}$. [3]

We first present some properties about the firm's posterior on the types of individuals in a network. Together with the nodes in $\mathbf{y}$, the ties in g^1 define a sub-network involving only nodes on which a signal is observed. In this sub-network, let d_i be the degree of i, and $r(i,j)$ be the distance, i.e., the length of the shortest path between i and j.

Proposition 1. *Let vector* $\mathbf{x}$ *indicate the types of individuals contained in vector* $\mathbf{y}$. $\Pr(\mathbf{x}|g, \mathbf{y})$ *is a multivariate normal density with precision matrix* $\mathbf{\Sigma}^{-1}$*:*

$$\begin{aligned}
(\mathbf{\Sigma}^{-1})_{ii} &= c + d_i \\
(\mathbf{\Sigma}^{-1})_{ij} &= -\mathbf{1}_{\{ij \in g^1\}}
\end{aligned}$$

and mean vector μ*:*

$$\mu = c\mathbf{\Sigma}\mathbf{y}. \tag{4}$$

[2]Notice that in a small society where everyone is likely to be friends with others, or in a society where each type is organized in perfectly homogenous and mutually disconnected sub-graphs ("components"), there is little to no uncertainty about an individual's type, implying that network based scores are less useful.

[3]This type of information arises when the lender observes all of i's friends and their signals, which implies that he is not friends with the rest of the society. Corollary 1 demonstrates an example of such a situation.

Proposition 1 states that the lender's beliefs about the types of individuals in the network follow a multivariate normal distribution whose parameters depend on the network structure. So, two individuals with identical individual signals (such as personal financial history) may obtain different network based scores because of social connections. These individuals would obtain similar financing opportunities if credit scores relied solely on individual history. In the new regime, despite identical individual financial histories, it is possible that they will have unequal access to financing because of the score gains and losses from the social network.

Equation (4) shows that the weight that a contact j's signal receives relies on his location within the network. Proposition 2 states an upper bound on the weight of connection j's signal on i's posterior mean. When all else is equal, the upper bound on the weight of j decreases in the distance $r(i,j)$ which is measured as the length of the shortest path between i and j. If i and j are not connected in the sub-network, the weight is zero.

Proposition 2. *For all $i \neq j$ and $r(i,j) < +\infty$, the weight matrix of Proposition 1 satisfies*

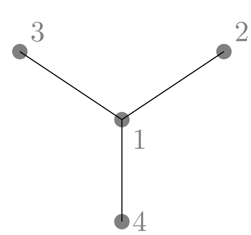
$$c\mathbf{\Sigma}_{ij} < \frac{c}{c+d_i}\frac{\delta^{r(i,j)}}{1-\delta},$$

where

$$\delta \equiv \frac{\max_{k\in\mathbf{y}}\{d_k\}}{c+\max_{k\in\mathbf{y}}\{d_k\}}.$$

To generate further insights about how the weight of a connection's signal changes with distance, we follow with two examples.

Example 1. *For a simple example, consider a star network g^1 that is centered at 1.*

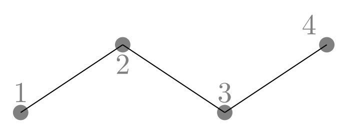


With $c = 1$, $c\mathbf{\Sigma}$ equals:

$$\begin{pmatrix} 0.4 & 0.2 & 0.2 & 0.2 \\ 0.2 & 0.6 & 0.1 & 0.1 \\ 0.2 & 0.1 & 0.6 & 0.1 \\ 0.2 & 0.1 & 0.1 & 0.6 \end{pmatrix}$$

By Proposition 1, this is a "weight" matrix, suggesting that to calculate the posterior mean of x_1, for example, the firm should weigh the signals (y_1, y_2, y_3, y_4) by $(0.4, 0.2, 0.2, 0.2)$. Note, further, that direct neighbors (friends) for nodes 2, 3, and 4 receive more weight than indirect neighbors (friends of friends).

Example 2. *Consider the following g^1.*



With $c = 1$, the weight matrix is:

$$\begin{pmatrix} 0.62 & 0.24 & 0.10 & 0.05 \\ 0.24 & 0.48 & 0.19 & 0.10 \\ 0.10 & 0.19 & 0.48 & 0.24 \\ 0.05 & 0.10 & 0.24 & 0.62 \end{pmatrix}$$

Note that direct neighbors are weighed more heavily than indirect neighbors, and that direct neighbors need not receive equal weight. For instance, the updating of x_2 weighs the signal from node 1 more heavily than that from node 3.

The above examples convey the intuition that distant signals on average receive lower weight in firm's updating of the beliefs about a consumer's type. For example, in examples 1 and 2, the weight of the signal of an individual who is two links away is always lower than the weight of the individual who is only one link away. In the second example, although individual 2 is at equal distance to persons 1 and 3, their signals receive different weights: Individual 3's signal is diluted as he is linked to individual 4.

Propositions 1 and 2 together imply that agents who have lower distances to high type individuals can receive a more favorable posterior in credit score assessment. Conversely, proximity to individuals with low signals may hurt an individual's assessment. It should be noted here that individuals cannot choose their distance as we have not *yet* considered active selection of friendship ties to attain such benefits (please see Section 3).

In the remainder of the paper up to Section 5, we assume that, when evaluating a particular i, the firm observes the complete ego-network of i, i.e., all the ties $ij \in G$, and receives a signal on each of i's friends ($ij \in G \Leftrightarrow y_i, y_j \in \mathbf{y}$). We collect the signals in the vector $\mathbf{y}_i$, which we will refer to as the set of i's friends as well. Note that this imposes an additional assumption on the previous analysis: We now require that g^1 consists of the complete set of i's direct ties. The posterior belief of the firm about an individual's type can then be stated as a special case of Proposition 1.

Corollary 1. *For the evaluation of i, $\Pr(x_i|\mathbf{y}_i)$ is normal with precision*

$$\rho_i = \left(1 + \frac{n_i}{c+1}\right) c, \tag{5}$$

and mean

$$\mu_i = \frac{1}{\rho_i}\left[c y_i + \frac{c}{c+1}\sum_{ij\in G} y_j\right].$$

Corollary 1 states that when an individual has a higher number of connections, the posterior about his type will have higher precision. The assessment of an individual with a higher degree is likely to be closer to this true type, x_i.[4] More importantly, (5) implies that the precision of lender's beliefs is higher than the precision of the individual signal of i, even with data only from the direct relationships of i. The corollary thus states useful information about the efficiency of risk assessment based on network data. If gathering data on the whole network is impossible or costly, efficiency gains can still be attained by using data from the focal consumer's immediate neighbors. Remember from Proposition 2 that first degree contacts of i receive a greater weight, and that data from longer paths in the network are expected to add at gradually lower weights to the beliefs about one's credit-worthiness.

[4] Notice that $\rho_i = 1/E((\mu_i - x_i)^2|\mathbf{y}_i)$, which is the inverse of the conditional mean squared error. Since in (5) ρ_i is increasing in n_i, we see that the conditional mean squared error is decreasing with n_i.

3 Endogenous Tie Formation

We next study individuals' incentives to form and modify network ties in order to improve their scores. This suggests that the probability that two agents will become friends depends on their type, x_i, and the expected utility from improving their credit score.

Facing network based scoring, an individual has an incentive to drop ties with low types in order to achieve a more favorable score. Such endogenous tie formation involves a trade-off between utility from friendship ties with people one likes and utility from a high score. To formally express this, we assume that the posterior mean μ_i enters the utility additively. The utility of individual i is:

$$U_i = \sum_{ij \in G} (m_{ij} - |x_i - x_j|) + \alpha \mu_i, \tag{6}$$

where the first part of the utility, $(m_{ij} - |x_i - x_j|)$, indicates a social utility, taking into consideration homophily; and the second part, $\alpha\mu_i$, indicates how much i enjoys having a high posterior mean. Here, α calibrates the relative importance an individual places on receiving a high credit score vs. the utility from friendship ties with people he likes. All individuals gain utility from their posterior credit score at rate α.[5] If $\alpha = 0$, the individual cares only about forming friendships for social utility. If $\alpha \to +\infty$, then the agent cares relatively little about social utility but highly about improving his score.

Parameter α can also be interpreted as a measure of the desire for status. How much people care about how others evaluate them (i.e., generate a posterior about their type based on characteristics of their network) captures the importance people place on their position in a social structure based on esteem that is bestowed by others, i.e., their status (Hu and Van den Bulte, 2014, and references therein). Let each individual i adopt a tie formation rule a priori (i.e., before meeting j) which states that he will accept friendship with j iff

$$\begin{cases} m_{ij} > \eta_i |x_i - x_j| & \text{for } x_j \geq x_i, \\ m_{ij} > \lambda_i |x_i - x_j| & \text{for } x_j < x_i. \end{cases}$$

[5] To allow for the possibility that some agents may have no interest in improving their scores when they meet others with similar types, Section 4 presents a discrete formulation of our matching model and allow the high types to incur zero utility from credit scores.

The parameters λ_i and η_i represent the degree to which i is willing to accept a lower and a higher type individual as a friend. These parameters are not exogenous but will be chosen optimally by i. Although individual i would prefer to be friends with others similar to him, which was expressed in (1), he may have additional utility from adding high type or removing low type friends due to the improvement in his credit assessment. This suggests that individuals will form relationships with others who have lower types only if the match value m_{ij} yields sufficiently high utility.

Comparing (6) with (1), a greater (lesser) desire to link to individuals with higher (lower) types would indicate that an agent should pick $\eta_i \leq 1$ and $\lambda_i \geq 1$. [6] Remember that forming a friendship tie requires mutual consent: for i and j to become friends, i should want to connect with j and j should want to connect with i. [7] Thus η_i becomes irrelevant and λ_i becomes the parameter that sets the level of mixing with 'others'. In the rest of the paper we omit any further references to η_i.

Consider the symmetric case where $\lambda_i = \lambda$ for all i. If everyone in the society applies the same rule with common λ, a friendship is established after meeting iff $m_{ij} > \lambda|x_i - x_j|$. With the common rule in place, the probability of becoming friends after meeting becomes:

$$Pr(|x_i - x_j|, \lambda) = e^{-\lambda|x_i - x_j|^2/2}.$$

Compared to the tie formation probability in an exogenous setting (given by Equation 2), individuals will be more selective in linking to others and fewer ties will be formed in the endogenous case.

[6] The benefits from network based scoring is measured by the difference between one's expected posterior mean and one's individual signal. This difference increases in λ_i (i.e., the rate at which the individual drops low-type friends) and decreases in η_i (i.e., the rate at which the individual adds high-type friends). Choosing $\eta_i > 1$ is worse than $\eta_i = 1$, because it decreases both the expected score benefit and the social utility of a tie. Similarly, choosing $\lambda_i < 1$ rather than $\lambda_i = 1$ would decrease the utility from a higher credit score and the social utility of a well-matching tie. Together, these two arguments imply that: (i) any symmetric equilibrium derived with restrictions is still an equilibrium even if we allow $\eta_i > 1$ or $\lambda_i < 1$; and more importantly, (ii) there is no symmetric equilibria where $\eta > 1$ or $\lambda < 1$.

[7] If we allowed individuals to form friendships without the consent of the other, then we would be in a trivial world where everyone can link to anyone to improve his own score. In such a world, the benefits of network data are limited since a connection to a high-type is not informative of one's type.

3.1 Credit scoring with endogenous tie formation

To complete the model for endogenous relationship formation, we use $\Lambda \equiv \{\lambda_j\}, \forall j \in S$ to denote the collection of everyone's rule. The expected utility of i becomes:

$$\mathbb{E}_\Lambda \left(U_i|x_i\right) = \mathbb{E}_\Lambda \left(\sum_{ij\in G} m_{ij} - |x_i - x_j| \; |x_i\right) + \alpha\mathbb{E}_\Lambda(\mu_i|x_i). \tag{7}$$

Each individual calculates his expected utility from being in a friendship network before the network is formed, implying that ex-ante utility will depend on the friendship rule Λ adopted. The expectation $\mathbb{E}_\Lambda$ is taken before meeting others. Similarly, we use $\mathbb{E}_\Lambda(n_i)$ to denote expected number of friends (i.e., degree) of i. We first display a version of Corollary 1 under a symmetric rule.

Lemma 1. *Under a common relationship formation rule λ, the posterior* $\Pr(x_i|\mathbf{y}_i)$ *is normal with precision*

$$\rho_i(\lambda) = \left(1 + \frac{n_i\lambda}{c+\lambda}\right)c, \tag{8}$$

and mean

$$\mu_i(\lambda) = \frac{1}{\rho_i(\lambda)}\left[cy_i + \frac{\lambda c}{c+\lambda}\sum_{ij\in G} y_j\right].$$

Compared to Corollary 1, in Lemma 1, ρ_i and μ_i are scaled by the selection rule λ. When borrowers are more selective in forming friendships with lower types (when λ is higher), a financial institution will put more weight on friends' signals to update beliefs about the type of an individual (i.e., to calculate the posterior). In broad terms, this selectivity addresses our second main research question: When individuals begin reacting to an environment with network based risk assessment, are credit scores going to be less precise or even more precise? In other words, can assessments based on network data yield better assessment of individual risk? Our answer to this question is a qualified yes. We explain the mechanism through which this improvement can be achieved via a lemma and then a proposition. Notice that when a symmetric rule is used, i.e., $\lambda_i = \lambda$ for all i, we use $\mathbb{E}_\lambda$ instead of $\mathbb{E}_\Lambda$ to denote the expectation under the rule.

Lemma 2. *The expected degree under a symmetric rule* λ *satisfies*

$$\mathbb{E}_\lambda(n_i) = \frac{N}{\sqrt{\lambda}} \ . \tag{9}$$

A lower rate of mixing between types (a higher λ) results in a smaller number of ties per person. Ties are formed only between individuals who are highly similar to each other in type. Such self-selection reduces the expected number of connections among consumers but increases the information value of any single link and the signal it conveys. The net effect on the formation of ties is not clear yet. We address it next.

Proposition 3 shows that, under the limits of S, ν and q, there is a symmetric equilibrium $\lambda_i = \lambda^*$ which maximizes (7) for any individual i, where $\lambda_j = \lambda^*$, and $\mu_i = \mu_i(\lambda)$, $\forall j$. In other words, there exists a common tie formation rule no individual wants to deviate from, with which the posterior is consistent.

Proposition 3. *For* $0 < \alpha < N$*, there exists at least one symmetric equilibrium, and any symmetric equilibrium* λ^* *must satisfy*

$$1 < \lambda^* < \left(1 - \frac{\alpha}{N}\right)^{-1} . \tag{10}$$

Corollary 2. *If* $c \geq \sqrt{\frac{N}{N-\alpha}}$*, then* $\mathbb{E}_{\lambda^*}\left[\rho_i(\lambda^*)\right] > \mathbb{E}_1\left[\rho_i(1)\right]$*. On average, the network based score becomes more accurate when consumers are averse to connecting with lower type peers. Otherwise, if* $c \leq 1$*, then* $\mathbb{E}_{\lambda^*}\left[\rho_i(\lambda^*)\right] < \mathbb{E}_1\left[\rho_i(1)\right]$*. On average, the network based scores are less accurate.*

Social credit scoring changes the incentives of individuals to form relationships. There are two directions of change. Compared to the exogenous setting ($\lambda = 1$), in the endogenous setting with $\lambda = \lambda^* > 1$, individual relationships are formed more selectively. This has several consequences. First, relationships are more strongly homophilous, that is, individuals form relationships with others who are closer to their own type. This first effect has a positive impact on network scores for lenders: The accuracy of their assessment will improve as a result of obtaining signals from closer types. Network based scores will prove to be even more precise due to data from others who are expected to be more similar in type.

Second, individuals will reject friendship ties with others who have lower types, implying that ego-networks will shrink in size (Lemma 2). This second effect has a negative impact on the accuracy of network scoring. The two forces, the homogenization and the shrinkage of ego-networks work against each other. The net effect is ambiguous.

Corollary 2 identifies a further condition, which we interpret using the parameter α, to characterize situations in which the net effect is positive and network score accuracy improves with endogenous tie formation. For some sufficiently small α, lenders may benefit from using network based credit scoring as it becomes even more precise with self-selection of individuals to form networks to improve their credit scores. The improvement in precision is conditional on consumers placing low weight on financial outcomes relative to the utility derived from social connections. Paradoxically, when individuals care greatly about their score or status, they may reduce the size of their social networks so much that network based scoring becomes less reliable in equilibrium.

Can societal tissue make network based scoring more effective in some societies than others? Corollary 2 states that the parameter range under which network based scores are more precise is larger when the average number of friends is higher. If everything else remains the same, the benefits of network based scoring may be greater in societies where people maintain a large number of connections, which are likely to be societies with collectivist cultures (Hofstede, 2001). Interestingly, several start-ups turning to social scoring have been growing in countries known to have collectivist cultures where density of relationships is generally higher. Lenddo, for instance, operates in Mexico, Colombia, and Philippines and reports that Mexico is its fastest growing market.[8]

3.2 Lending Rates with Endogenous Network Formation

We now relate our scoring formulation to lending rates, i.e., access to finance at the intensive margin. The discussion in this section implies that network based scoring affects the rates at which individuals can borrow, even if these individuals would qualify to receive credit using the individual score system. For simplicity and concreteness of discussion, we specify the perceived probability of

[8] http://techonomy.com/2014/02/lenddos-borrowers-mexico-philippines-get-credit-via-facebook/

repayment of credit by individual i, P_i as

$$P_i = \frac{1}{1 + e^{-\mu_i}}$$

which increases from 0 to 1 as the lender's assessment of the borrower's posterior mean, μ_i, increases from $-\infty$ to $+\infty$. Consider a risk-neutral lender who earns a rate of r^o from a non- risky investment. Let r_i be the lending rate to be charged to individual i with type x_i. The firm determines the rate by solving:

$$P_i \cdot (1 + r_i) + (1 - P_i) \cdot 0 = 1 + r_o.$$

Notice that this formulation takes into account not only the expected creditworthiness of an individual, μ_i, but also the outside options of the lender, r^o. For $r^o = 0$, the borrowing rate for i equals the log odds of default vs. repayment:

$$r_i = \frac{1 - P_i}{P_i} = e^{-\mu_i}. \tag{11}$$

As the consumer's likelihood of a default increases, he faces a higher borrowing rate. Notice that the financial utility of consumers given in Equation (6) can be derived by assuming that the lending rate enters the utility through $\alpha\mu_i = -\alpha \log(r_i)$. If lending rates can be interpreted within the context of economic opportunities available to consumers, then an individual with a better network score will be likely to receive a loan on better terms. This links network based credit scores to financing access at the intensive margin.

4 Asymmetric Needs for Financing & The Role of Signals

In the preceding sections, we developed a model with continuous types and assumed that every individual has some incentive to improve his credit score. In reality, there may be differences between the low and high type individuals about the utility gained from improving their credit score. In general, the need for improving scores and the need for financing are greater for the low types. We introduce a discrete version of the model to address this possibility. The new model allows us to analyze in greater detail how the firm utilizes signals of low vs. high type friends when assessing an individual's creditworthiness. This enables us to disentangle and contrast the role of high and low type contact signals in the network.

4.1 Credit scoring and tie formation with high and low types

Consider a society with two types of borrowers: high types (h) and low types (ℓ) where the prior is uniform, with $Pr(x_i = \ell) = Pr(x_i = h) = \frac{1}{2}$. Whereas high types have a low risk of credit default, low types have a higher risk. With probability ν, any two individuals will meet. Upon meeting, they learn each other's type and their match value $m_{ij} > 0$, which is i.i.d. across pairs, with positive distribution density f. For i, the utility of becoming friends with j is

$$m_{ij} - \mathbf{1}_{\{x_j \neq x_i\}}, \tag{12}$$

where the disutility of becoming friends with a different type is normalized to 1. The utility of not becoming friends is 0. Given the specification, the probability that two same-type consumers will become friends conditional on meeting is 1, while the probability of two different types becoming friends is $p \equiv \Pr(m_{ij} > 1) < 1$; hence the network features preference homophily. We retain the assumptions $S \to +\infty$ and $\nu \to 0$ and set $S\nu = N$ for some positive number N. With the discrete formulation, the expected number of friends for any type is $\frac{1}{2}S\nu(1+p)$: increasing the degree of homophily (a lower p) reduces the expected number of friends.

Network based score. We assume that the lender may observe a signal y_i which is -1 or 1, indicating a low or high type. The signal is credible but incorrect with probability $\varepsilon < \frac{1}{2}$. This implies, for example, that if the lender receives a signal from an ℓ-type consumer, with probability $1 - \varepsilon$ it observes $y_i = -1$ and with the remaining probability it observes $y_i = 1$. Let $\mathbf{y}_i$ be the collection of signals from i and the friends of i. We first explore how the firm perceives the probability of an agent being of h-type conditional on the structure of his social network.

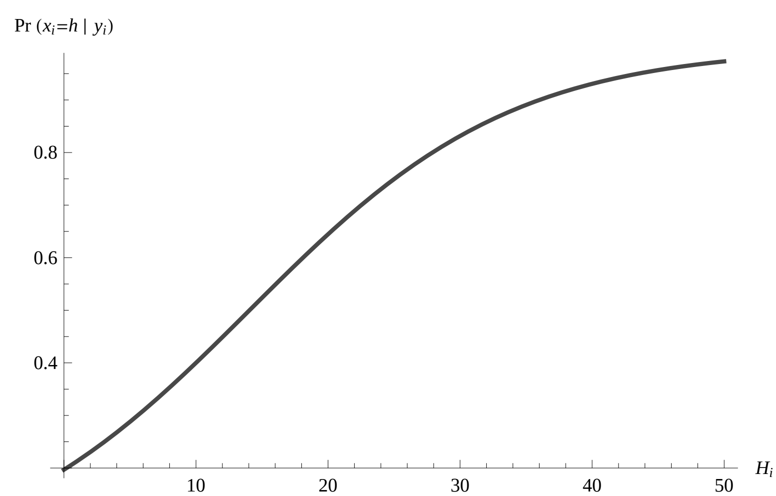
Lemma 3. *In evaluating i, the posterior for him to be high type is*

$$\Pr(x_i = h|\mathbf{y}_i) = \left[1 + \left(\frac{\varepsilon}{1-\varepsilon}\right)^{y_i} \left(\frac{\varepsilon p + (1-\varepsilon)}{\varepsilon + (1-\varepsilon)p}\right)^{L_i} \left(\frac{\varepsilon + (1-\varepsilon)p}{\varepsilon p + (1-\varepsilon)}\right)^{H_i}\right]^{-1} \tag{13}$$

where y_i is the signal observed for agent i, H_i is the number of friends with high signal, L_i is the number of friends with low signal.

Lemma 3 suggests that low and high type signals observed for an individual's social connections affect the lender's assessment of that individual's creditworthiness in different directions. Notice that $\left(\frac{\varepsilon+(1-\varepsilon)p}{\varepsilon p+(1-\varepsilon)}\right) < 1$ and $\left(\frac{\varepsilon p+(1-\varepsilon)}{\varepsilon+(1-\varepsilon)p}\right) > 1$. Thus, high type signals increase the likelihood that an agent will be categorized as being of high type, whereas low type signals reduce this likelihood. Figures 1 and 2 illustrate how $\Pr(x_i = h|\mathbf{y}_i)$ changes with H_i and L_i. The firm would prefer to extend credit to ℓ-types with a higher number of h-type connections, if everything else remained the same. This suggests that in a given network where ℓ-types are fairly segregated from the h-types due to homophily, ℓ-types who are bridges between ℓ-types and h-types may be favored by the lender (compared to ℓ-types surrounded by the same-types). Put differently, in-group centrality of ℓ-types will hurt their financing opportunities whereas between-group centrality will improve them.

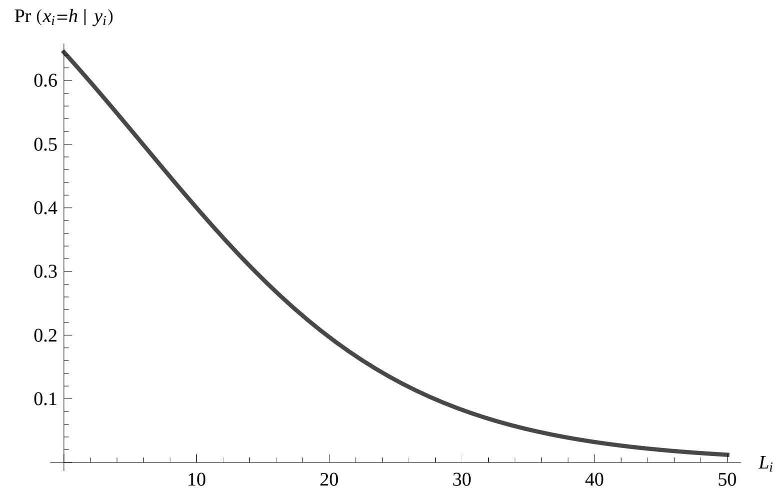
Figure 1: $\Pr(x_i = h|\mathbf{y}_i)$ vs. H_i ($\varepsilon = 0.4, p = 0.6, L_i = 10, y_i = -1$)



Endogenous Network Formation. Let's denote the lending rate for borrower i with r_i and the no-risk investment rate with r^o. Assume that high types always repay while low types always default.

We assume that individuals with low types gain utility from the firm's posterior assessment and that high types' utility from the posterior is zero. The firm's assessment of an individual's

Figure 2: $\Pr(x_i = h|\mathbf{y}_i)$ vs. L_i ($\varepsilon = 0.4, p = 0.6, H_i = 10, y_i = -1$)



probability of repayment, P_i, is then equivalent to $\Pr(x_i = h|\mathbf{y}_i)$. Note that for $r^o = 0$, we have $r_i = \frac{1-P_i}{P_i}$. We construct the utility of a borrower similar to Section 3.2, where the lending rate enters the utility of the ℓ-type additively through $-\alpha \log(r_i)$:

$$U_i = \begin{cases} \sum_{ij \in G} \left(m_{ij} - \mathbf{1}_{\{x_j \neq x_i\}} \right) + \alpha R_i & \text{if } x_i = \ell \\ \sum_{ij \in G} \left(m_{ij} - \mathbf{1}_{\{x_j \neq x_i\}} \right) & \text{if } x_i = h, \end{cases} \tag{14}$$

where $R_i \equiv \log\left(\frac{P_i}{1-P_i}\right)$. A higher R_i implies a lower risk of extending credit to an ℓ-type. Notice that this formulation mirrors our continuous-type model, except for the difference that financial concerns (represented by R_i) does not enter the utility of high types.

Let individuals choose tie formation rules before the meeting process. Due to the simplicity of the discrete-type model, friendship rules we allow are general and flexible. Formally, the rules we allow are those that can be written as a function $f(m_{ij}, x_i, x_j)$ such that an individual will accept friendship with j iff $f(m_{ij}, x_i, x_j) > 0$. Since high types do not have financial motives in forming friendship ties, their behaviors remain $f(m_{ij}, x_i, x_j) = \left(m_{ij} - \mathbf{1}_{\{x_j \neq x_i\}} \right)$, similar to Section 3. Low types, without loss of generality, will be willing to form friendship with any high type yielding $\left(m_{ij} > \mathbf{1}_{\{x_j \neq x_i\}} \right)$. Doing so improves not only their social utility, but also their chances of being perceived as a high type. They will be selective, however, in forming ties with other ℓ-types, since ℓ-types, as demonstrated in Lemma 3, pose a threat to the credit evaluation even though they may add social utility. Low type i will use a criterion $\theta_i \geq 0$ to accept a tie with another ℓ-type j iff: [9]

$$m_{ij} > \theta_i \cdot \mathbf{1}_{\{x_j = \ell\}}.$$

We seek to understand how individuals form ties when R_i enters the utility function of ℓ-types as in (14). Let Θ denote the collection of the criteria of all individuals and let a symmetric criterion θ indicate $\theta_i = \theta$ for all $i = \ell$ types. The expected utility for an ℓ-type i is the sum of the expected utility from friendship and the expected utility from the posterior credit score:

$$\mathbb{E}_\Theta(U_i | x_i = \ell) = \mathbb{E}_\Theta \left(\sum_{ij \in G} m_{ij} - \mathbf{1}_{\{x_j = h\}} | x_i = \ell \right) + \alpha \mathbb{E}_\Theta(R_i | x_i = \ell). \tag{15}$$

As in the continuous case, social credit scoring makes individuals wary of forming ties with low types. However, in the discrete case where low types have a greater need for financing than the high types whose needs are normalized at zero, low types face discrimination or social rejection only from the other low types.[10] This pure within-group discrimination arises endogenously with the use of the network based scoring. We focus on this discrimination result. Discrimination is often thought to take place between groups, or is believed to be exercised by one group on another. Our results suggest that ℓ-type individuals are systematically excluded from the networks of others similar to them. This is a unique societal outcome directly resulting from the asymmetric financing needs of the two groups and the differential between them. Put differently, existing financial inequality breeds within-group discrimination and social isolation among those of lower type and greater need. In turn, this discrimination may make the surviving within-group ties more valuable, as we see next in Lemma 4.

Lemma 4. *Let θ be the symmetric criterion, and let $\widehat{p} \equiv \Pr(m_{ij} > \theta)$ indicate the probability of two ℓ-types forming ties. Then*

$$\Pr(x_i = h | \mathbf{y}_i) = \left[1 + \left(\frac{\varepsilon}{1-\varepsilon} \right)^{y_i} \left(\frac{\varepsilon p + (1-\varepsilon)\widehat{p}}{\varepsilon + (1-\varepsilon)p} \right)^{L_i} \left(\frac{\varepsilon \widehat{p} + (1-\varepsilon)p}{\varepsilon p + (1-\varepsilon)} \right)^{H_i} e^{\frac{1}{2}N(1-\widehat{p})} \right]^{-1} \tag{16}$$

[9] Note that under the individual data based scoring $\theta_i = 0$ is optimal.

[10] When $\alpha_\ell > \alpha_h > 0$, the extremity of discrimination is determined by the acceptance of low types of other low types, but high types will also discriminate against them.

where y_i is the signal observed for agent i, H_i is the number of friends with high signal, and L_i is the number of friends with low signal.

Lemma 4 presents a slightly different result compared to Lemma 3 in decomposing the contributions of high and low signals. When individuals form ties endogenously, the probability of a favorable risk assessment, $\Pr(x_i = h|\mathbf{y}_i)$, is increasing in the number of high signals (i.e., H_i) for any level of $\widehat{p}$. However, it now increases in the number of friends with low signals (i.e., L_i) for $\widehat{p} < p + \frac{\varepsilon(1-p)}{1-\varepsilon}$, and decreases in L_i otherwise. In other words, when ℓ-types are very selective in forming ties amongst themselves ($\widehat{p}$ low), then in-group ties help rather than hurt and this strengthens their value. As a result, low types have fewer ties than the high types and a large friendship circle becomes a conspicuous signal, suggesting that one is likely an h-type. That is the reason low-type friends increase the high type perception, $\Pr(x_i = h|\mathbf{y}_i)$ and so strengthen the signal up to a certain level only. As the number of low type friends increases further, the negative associations eventually exceed the positive impact and $\Pr(x_i = h|\mathbf{y}_i)$ decreases in L_i again.

We now turn to the impact of how selective low types are in forming ties amongst themselves, characterized by the selection rule θ. Let us define $\overline{\theta}$ by $\Pr(m_{ij} > \overline{\theta}) = p + \frac{\varepsilon(1-p)}{1-\varepsilon}$. Notice that $\overline{\theta} \in (0,1)$. To investigate the equilibrium selection rules, we use $P_i(\theta)$ to denote the posterior in (16) and $R_i(\theta) = \log\left(\frac{P_i(\theta)}{1-P_i(\theta)}\right)$. The common selection rule θ influences the firm's evaluation of individual's risk. In particular, Proposition 4 states the relationship between θ and the expected risk assessment received from the firm.

Proposition 4. *As a function of θ, the expected log odds of repayment for a low type i,*

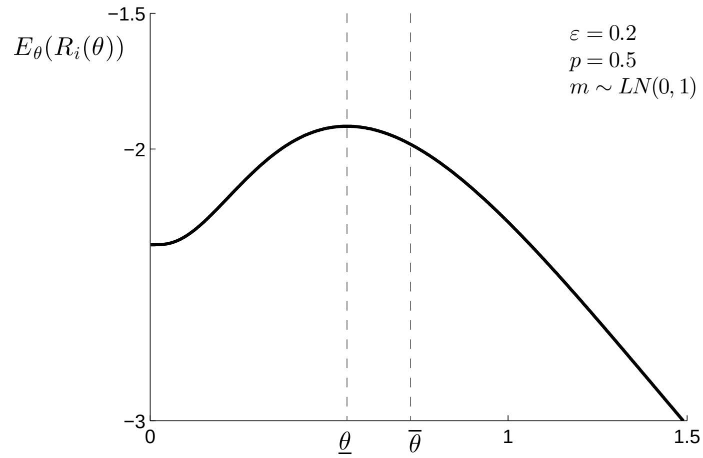
$$\mathbb{E}_\theta \left[R_i(\theta) | \, x_i = \ell \right]$$

is strictly quasi-concave on $[0, +\infty)$. Let $\underline{\theta}$ denote the point where $\mathbb{E}_\theta \left[R_i(\theta) \right]$ is maximized. Then $\underline{\theta} \in (0, \overline{\theta})$. Further, for any θ, the expected log odds is smaller under social scoring than under individual-based scoring.

Figure 3 plots a numerical example for the expected log odds of repayment as a function of θ. Notice that very high or very low levels of within-group selectivity results in lower expected odds;

whereas medium levels of selectivity among low types yield the most favorable risk assessment for them.

Figure 3: Expected Log Odds of Repayment as a Function of Selectivity θ.



The inverse U-curve relationship stems from two competing forces that shape low-type borrowers' chances of receiving a loan. As the level of selectivity begins to increase from zero, the expected assessment improves at first. Consumers benefit from disassociating themselves from ℓ-types, improving the appearance of being an h-type. As selectivity increases further, however, a second and competing effect starts to dominate: Individuals' ego-networks start to shrink extensively. Recall that the size of a borrower's network becomes a conspicuous signal of his type when individuals can form ties endogenously. Extreme selectivity leads to a smaller number of ties and so reveals the true low type of a borrower, reducing his chances of a favorable credit assessment.

To complete the analysis, we identify an equilibrium θ^* in which given that $\theta_j = \theta^*$ for all $j \neq i$, any low type i maximizes his utility in (15) by choosing $\theta_i = \theta^*$, where R_i is $R_i(\theta^*)$. First, notice that any $\theta^* \geq \overline{\theta}$ is an equilibrium. Remember that at high levels of selectivity among low types (when $\theta^* \geq \overline{\theta}$) , the log odds $R_i(\theta^*)$ is increasing in both L_i and H_i. Choosing a more stringent criterion $\theta_i > \theta^*$ is not a profitable deviation because it not only decreases the expected utility from friendship but also lowers the expected posterior and $R_i(\theta^*)$. Further, accepting a criterion slightly less restrictive ($\theta_i < \theta^*$) unilaterally does not change the outcome for the individual because only mutual acceptance leads to tie formation.

Proposition 5. *Any equilibrium must satisfy $\theta^* > 0$, and any $\theta^* < \underline{\theta}$ sufficiently close to $\underline{\theta}$ is an equilibrium.*

Since there are multiple equilibria, we consider refinement by using a selection rule. Given that expected log odds is a measure of how likely a low type will be (wrongly) seen as a high type by the firm, to obtain access to financing, a low type would prefer a higher expected log odds, which, by Proposition 4, is differentiably maximized at $\underline{\theta}$. Moreover, the expected utility from friendship is strictly decreasing in θ. So any $\theta^* \geq \underline{\theta}$ is less preferable than some $\theta^* < \underline{\theta}$. Therefore we can eliminate $\theta^* \geq \underline{\theta}$ as an equilibrium.

In summary, asymmetric valuation of financing and endogenous tie friendship formation may lead to within-group discrimination. Low levels of discrimination cannot sufficiently eliminate the negative association from being connected to low-type individuals and high levels of discrimination hurt by reducing the number of social ties. Consumers prefer a medium level of discrimination.

4.2 Non-strategic Discrimination against Low Types

We have shown how strategic discrimination against those of low type and high need may emerge endogenously even in the presence of non-strategic homophily among low types. To extend the discussion on discrimination, we analyze an environment with exogenous discrimination against ℓ-types. To formally express such discrimination, we construct the utility for i of becoming friends with j in a manner similar to but different from the specification in Equation (12):

$$m_{ij} - \mathbf{1}_{\{x_j=\ell\}}.$$

Keeping the discrete matching formulation with this slight modification, the probability that two h-type individuals will become friends conditional on meeting is 1 and the probability that any other type of pairs will become friends is $p \equiv \Pr(m_{ij} > 1) < 1$. The social utility is penalized whenever one becomes friends with an individual who is an ℓ-type. With such explicit or non-strategic discrimination, the expected number of friends varies for each type: For a high type, the expected degree is $\frac{1}{2}S\nu(1+p)$, whereas for a low type it is $S\nu p$. We will see that this difference influences the incentives of individuals to enter relationships, since, similar to the endogenous rise of discrimination, a larger social network can become a conspicuous signal. An individual with a larger network emits a stronger signal that he is of high type.

Network Based Score. Let the credit scoring institution observe a credible signal y_i which takes a value -1 (1) to indicate a ℓ (h) type with an error rate of ε.[11] The posterior belief of the credit institution about an individual's credit-worthiness, conditional on borrower's network, is given in Lemma 5.

Lemma 5. *With explicit discrimination, the posterior of x_i is*

$$\Pr(x_i = h|\mathbf{y}_i) = \left[1 + \left(\frac{\varepsilon}{1-\varepsilon}\right)^{y_i} \left(\frac{p}{\varepsilon + (1-\varepsilon)p}\right)^{L_i} \left(\frac{p}{\varepsilon p + (1-\varepsilon)}\right)^{H_i} e^{\frac{1}{2}N(1-p)}\right]^{-1} \tag{17}$$

where y_i is the signal observed for agent i, H_i is the number of friends with high signal, L_i is the number of friends with low signal.

Notice that Lemma 5 is a special case of Lemma 4, with $\widehat{p} = p$. The lemma states that under the special case, *all* social connections boost the posterior probability of being classified as low risk, as $\Pr(x_i = h|\mathbf{y}_i)$ is increasing in both L_i and H_i. The probability, however, increases faster in H_i: high-type signals have a stronger impact than low-signals. Low-type signals contribute positively to $\Pr(x_i = h|\mathbf{y}_i)$ for two reasons. First a large ego-network is a conspicuous signal of being a high type. Second, a low-type signal may still be wrong due to the imperfect assessment of the credit scoring institution. Therefore each additional low signal strengthens the belief that the individual is likely to be of h-type.

Endogenous Network Formation. We consider how individuals form ties under network based scoring when explicit discrimination against low types is present. Similar to the analysis in Section 4.1, R_i denotes a low type i 's log odds of repayment and it enters the utility additively.

$$U_i = \begin{cases} \sum_{ij \in G} \left(m_{ij} - \mathbf{1}_{\{x_j = \ell\}}\right) + \alpha R_i & \text{for } x_i = \ell, \\ \sum_{ij \in G} \left(m_{ij} - \mathbf{1}_{\{x_j = \ell\}}\right) & \text{for } x_i = h. \end{cases}$$

[11] We retain the environment $S \to +\infty$ and $\nu \to 0$ to set $S\nu = N$ for some positive number N.

Let individuals choose tie formation rules $f(m_{ij}, x_i, x_j)$ before meeting such that i will accept friendship with j iff $f(m_{ij}, x_i, x_j) > 0$. Since high types do not obtain utility from credit score improvement, their friendship rules remain the same as in Section 4.1. They become friends with any other high type and they choose to be friends with ℓ-types if $m_{ij} > \mathbf{1}_{\{x_j=\ell\}}$ is satisfied. Low types accept friendship with any high type, since doing so increases both their social utility and expected financial utility, but they may act selectively against fellow low types. They choose a criterion $\theta_i \geq 0$ to accept friendship with other low types iff $m_{ij} > \theta_i \cdot \mathbf{1}_{\{x_j=\ell\}}$.

We use $\mathbb{E}_\Theta$ to denote the expectation taken before the meeting process when criteria Θ is used, and use the notation $\mathbb{E}_\theta$ for the special case of a symmetric criterion θ. The total expected utility for i is the sum of the expected utility from friendship and the expected utility from the posterior:

$$\mathbb{E}_\Theta(U_i|x_i) = \begin{cases} \mathbb{E}_\Theta\left(\sum_{ij\in G} m_{ij} - \mathbf{1}_{\{x_j=\ell\}}|x_i = \ell\right) + \alpha\mathbb{E}_\Theta(R_i|x_i = \ell) & \text{for } x_i = \ell, \\ \mathbb{E}_\Theta\left(\sum_{ij\in G} m_{ij} - \mathbf{1}_{\{x_j=\ell\}}|x_i = h\right) & \text{for } x_i = h. \end{cases} \tag{18}$$

We look for an equilibrium θ^*, by which we mean that for any low type i, given that $\theta_j = \theta^*$ for all $j \neq i$, $\theta_i = \theta^*$ maximizes his utility (18), where R_i is given by $R_i(\theta^*)$. Notice that, for a symmetric criterion θ, the friendship tie formation probabilities among different types are expressed as in Section 4.1, hence both Lemma 4 and Proposition 4 apply here without any change.

For the equilibrium selection rules, any $\theta^* \geq 1$ is a equilibrium. To see this, note that when $\theta^* \geq 1$ the odds ratio $R_i(\theta^*)$ is increasing in both L_i and H_i. Choosing a larger criterion $\theta_i > \theta^*$ is not benefitable because it not only decreases the expected utility from friendship but also lowers the expected posterior and log odds of repayment. In contrast, choosing $\theta_i < \theta^*$ makes no difference because friendship must be mutual.

Proposition 6. *Any equilibrium must satisfy $\theta^* > \underline{\theta}$ and any $\theta^* < 1$ sufficiently close to 1 is an equilibrium.*

Consider which equilibria are more plausible. A low type would prefer a higher expected log odds, and by Proposition 4, the expected log odds is strictly decreasing in θ around and after $\theta = 1$. The expected utility from friendship is differentiably maximized at $\theta = 1$. This tells us that any $\theta^* \geq 1$ is less preferable for a low type than some $\theta^* < 1$. Hence, we may drop those equilibria

$\theta^* \geq 1$. As in the case without non-strategic or explicit discrimination, in equilibrium, low types prefer some intermediate level of endogenous discrimination against themselves.

5 Effort to Become a High Type

Our results thus far relied on the assumption that individuals in society are endowed with 'types' that cannot be changed. In other words, we assumed that there is no social mobility. Although some indicators of type (e.g., family, race, birth place, country of origin) cannot be altered, other potential indicators such as educational, occupational, or financial achievements can be improved if low types exert effort. In this section, we extend our discussion to allow for this possibility. An array of factors may force ℓ-type individuals to exert effort, but we will focus on factors endogenous to tie formation such as the reduction of borrowing costs and the threat of social exclusion.

We model the mechanism in the following fashion. Consider a friends network G among ℓ and h type individuals. Further, let G_ℓ denote the sub-network among the low types, and let H_i denote the number of h-type contacts of i, which collectively are represented with the vector $\mathbf{H}$ for all i. Each low-type individual may exert effort $e_i \geq 0$ such that with probability e_i he will become a high type. We assume that given the network and the parameters of our model, $e_i \leq 1$ for all low type i, and that high type consumers exert zero effort. We represent the utility of an h type individual in the following fashion:

$$U_i(\mathbf{e}, G) = \sum_{ij \in G} \left[m_{ij} - \mathbf{1}_{\{x_j = \ell\}} \cdot (1 - e_j) \right]. \tag{19}$$

The utility expresses the social utility from friendship, maintaining explicit discrimination against ℓ-types similar to Section 4.2.

The utility an ℓ-type individual i derives from exerting effort e_i is composed of two parts:

$$U_i(\mathbf{e}, G) = \sum_{ij \in G} \left[m_{ij} - \mathbf{1}_{\{x_j = \ell\}} \cdot (1 - e_j) \right] + u_i \tag{20}$$

where

$$u_i = a e_i - \frac{b e_i^2}{2} + \phi b \left(H_i + \sum_{ij \in G_\ell} e_j \right) e_i. \tag{21}$$

The term u_i expresses the benefits from exerting effort net cost of effort. The term ae_i expresses the direct utility from exerting effort which increases one's probability of becoming a high type and the cost of effort is given with $\frac{be_i^2}{2}$. Parameters a and b calibrate productivity and cost of effort, and for tractability we assume them to be respectively linear and quadratic.

The effort term e_j enters the utility equations (19) and (20) because effort increases the probability that i will give j a goodwill credit for effort and will not discriminate against him. As in Section 4.2, there is a disutility for friendship with a low type. However, the disutility can be reduced by the effort of the friend, $(1-e_j)$. If for example, j exerts maximum level of effort, $e_j = 1$, then any i will always want to connect to him upon meeting. Any level of effort $e_j > 0$ reduces the discrimination against j.

Effort by an individual j's contribution can also reduce his network's financing burden. Social credit scores (and thus financing rates) are determined not only based on one's individual signal but also based on his network's signals. We formally express this 'network effect' by including a term for reduction in cost of borrowing of the individual: $\phi b \left(H_i + \sum_{ij \in G_\ell} e_j \right) e_i$. When $\phi > 0$, improvement in cost of borrowing reflects complementarities between one's effort and that exerted by one's network contacts. Others' efforts increases the returns to one's own effort. The complementarity between one's effort and that of one's network has intuitive appeal, however, the formulation is flexible enough to accommodate other structures. For example, if $\phi = 0$, effort of others have a null effect on one's cost of borrowing and if $\phi < 0$, effort of others is substitutable to one's own effort.

Next, we will derive the optimal effort level, first assuming an exogenous and then an endogenous network.

Effort in an Exogenous Network. We are interested in the Nash equilibrium when people simultaneously choose their efforts and i's network is exogenously given. Proposition 7 summarizes the optimal level of effort for an individual conditional on his social network, following Ballester et al. (2006).

Proposition 7. *Let $\mathbf{A}_\ell$ be sociomatrix (i.e., the adjacency matrix) of G_ℓ. If the largest-magnitude* *eigenvalue of $\mathbf{A}_\ell$ is smaller than $|\phi|^{-1}$, then the equilibrium effort is:*

$$\mathbf{e}^* = (I + \phi \mathbf{A}_\ell + \phi^2 \mathbf{A}_\ell^2 + ...)(a/b + \phi \mathbf{H}) \tag{22}$$

$$= (I - \phi \mathbf{A}_\ell)^{-1}(a/b + \phi \mathbf{H}). \tag{23}$$

Proposition 7 states that the effort exerted by individuals to improve their score relies on several factors. First, the sign of network effect (ϕ) matters. Second, types matter. If an individual exerts high effort, it is more likely that he will become a high type, and thus it is easier for him to make friends. Moreover, an individual with a higher number of friends is likely to exert more effort, as his overall cost of borrowing is lower. This reinforcing effect makes it possible for substantially different networks to be pairwise stable. Third, strong within-group connectivity among the same type individuals (as measured in degree) influence effort. If two borrowers have identical between-group connectivity, the one with stronger within-group connectivity will exert more effort, when $\phi > 0$. In simpler terms, if two ℓ-type individuals are connected to the same number of h-type friends, the one with higher number of ℓ-type friends is incentivized to exert more effort. Perhaps more surprisingly, sufficiently high within-group connectivity may be a stronger motivator of effort. Overall, our model supports the claim that network based scoring is likely to improve agents' credit scores at different rates, conditional on their network structure.

Observation 1. *The expression for the equilibrium level of effort given in Equation* (23) *is Bonacich centrality. The effort exerted by an agent to improve his credit score is proportional to his centrality measure.*

Proposition 7 shows that the Nash effort is proportional to the Bonacich centrality measure, which is the 'summed connections to others, weighted by their centralities of connections to others' (Bonacich, 1987, p. 1172). When $\phi > 0$, an individual who is located at the center of a social network is likely to be exposed to higher positive network effects, therefore may exert greater effort. As a result, individuals who are more central in the network are more prone to social mobility when there are complementarities. On the contrary, when $\phi < 0$, then individuals who are more central exert lower levels of effort. In this environment, individuals who are less central may be subject to higher levels of social mobility.

Effort with Endogenous Network Formation Among Low Types. As we have specified in (19) and (20), the friendship utility of a friend of i depends on the effort that i will exert. Hence the effort of i plays an important role in his friends' network formation. Moreover, in the last section we saw that i's effort depends on his position in the network. This mutual dependence between the network position and effort suggests the possibility of multiple stable situations. For example, in one society people may exert low effort, and as a result, may become sparsely connected. This in turn gives little incentive for them to exert effort. Conversely, in another society, people may exert high effort and thus may become more densely connected, reinforcing their high-effort behavior.

To further explore how effort mitigates the likelihood of exclusion, we consider a two-stage game. In the first stage, individuals choose friends and friendships are formed bilaterally. In the second stage, individuals exert efforts. Let $\mathbf{e}^*(G)$ be the Nash effort for a given network G, which is characterized in Proposition 7. The first-stage reduced form utility for i depends on G only:

$$U_i\left(\mathbf{e}^*(G), G\right).$$

We look for pairwise-stable networks G under U. G is pairwise stable if (i) for any $ij \in G$, we have both $U_i(G) > U_i(G - ij)$ and $U_j(G) > U_j(G - ij)$; (ii) for any $ij \notin G$, either $U_i(G) \geq U_i(G + ij)$ or $U_j(G) \geq U_j(G + ij)$.

Example 3 provides an application of different stability outcomes in equilibrium.

Example 3. *Consider a society with four low-type individuals, and assume $a = 1$, $b = 5$, and $m_{ij} = \frac{1}{2}$ for all i, j. Let $\phi b = \frac{1}{5}$. It can be verified that both the empty network and the complete network are pair-wise stable. For the empty network, each individual exerts effort $\frac{1}{5}$ and obtains utility of $\frac{1}{10}$. For the complete network, each individual exerts effort $\frac{1}{2}$ and has utility $\frac{5}{8}$.*

The example demonstrates that the empty network is pair-wise stable because everyone exerts very low effort, and a single link between a pair won't generate a change sufficiently large. The disutility of friendship with a low type (which is normalized to 1) prevents any pair from becoming friends. Moreover, a complete network is pair-wise stable because everyone exerts reasonable effort. The effort reduces the disutility of friendship between low types, and the friendship utility between any pair is exactly zero. Breaking any one link increases the costs of effort for the pair, and they will decrease their efforts. This leads to higher costs for their friends and eventually the effort of everyone will decrease. As a result, everyone receives less utility from both the friendship and effort.

Overall, the example suggests that the network structure in different societies may facilitate social pressure to exert effort at different rates. In particular, in societies where network structure is sparse, it is expected to be less effective and social mobility may remain limited. In contrast, in denser societies, social pressure can be more effective, motivating higher levels of social mobility. The difference suggests that network based scoring practices are expected to reach different levels of success in different societies; and the performance is conditional on the network structure of society.

6 Conclusion

6.1 Main Insights

Increasing access to financing is important in many countries where institutions and contract enforcement are weak (e.g., Feigenberg et al., 2010; Rona-Tas and Guseva, 2014). In low-income countries, in particular, part of the credit access problem stems from the fact that reliable data on financial history do not exist, are limited, costly to collect, or hard to verify. In these countries, lenders tend to be very conservative in accepting borrowers' credit applications. This, of course, makes it even harder for individuals who are in financial hardship to obtain credit and generate a financial track record. Group lending has proven to be a popular way to address this problem. An alternative and possible complement is to use additional available data to assess individuals' creditworthiness. Using social data is one such option.

Motivated by the importance of consumer access to credit and by the increasing use of network based credit scoring, we analyzed the potential implications of such practices for consumers. Our study shows that there are indeed benefits to collecting information from a consumer's network rather than only individualized data. Simply put, when people have an above-average chance of interacting with others of similar creditworthiness, then network ties provide additional reliable signals about a consumer's true creditworthiness. Hence, social scoring can reduce lenders' misgivings about engaging people with limited personal financial history, which include many who are economically disadvantaged and "underbanked".

As these new scoring methods gain popularity, consumers may adapt their personal networks, which in turn may affect the usefulness of these scores. If one's network can influence one's financing chances, some individuals, particularly those in more dire need of improving their credit score, may be inclined to form social ties more selectively. If all consumers behave in this manner and forming social ties requires mutual agreement, the end result of such behavior will be social fragmentation into sub-networks where people connect only to others who are very similar to them. Though we expect that such fragmentation and balkanization will be deemed socially undesirable by many, its implications for network scoring's accuracy is not straightforward. People will have fewer ties conveying information about one's contacts useful in updating lenders' prior beliefs, but each of the ties will be more informative. We find, however, that there are situations in which social scoring is beneficial even when consumers adjust their networks. Specifically, when consumers place sufficiently low importance on the posterior mean of the firm, higher accuracy in risk assessment with network based scoring is possible even when individuals form their ties endogenously.

To focus on the role of connections to consumers with different levels of financial strength in the emergence of balkanized societal structures, we introduce discrete types and discrete type matching. Unsurprisingly, connections to individuals with high type signals have an overall positive impact. More interesting is that the impact of connections to low type signal individuals can be positive or negative, depending on the tie formation rules used in society. We find that consumers with poor financial health and in great need for credit would prefer others not to be too selective but also not to be too liberal in their willingness to associate with people having poor financial health and a great need for credit. As a result, disadvantaged consumers would prefer some intermediate level of ostracism and social isolation.

We also considered the possibility of exerting effort to improve one's type. We find that both low and high type contacts play a role in motivating effort, but high types, in general, have a stronger effect. Also, exerting effort to improve one's true type is more effective in dense than in sparse networks.

6.2 Implications for public policy

The link between credit scores and income is hard to ignore.[12] It is reported that consumers with lower than 60K income levels are at unhealthy levels of credit scores.[13] Moreover, a significant portion of the individualized credit score calculation relies on an individual's existing debt level. Someone owing higher amounts, all else equal, is expected to have a lower credit score. With network based assessment, it is possible for individuals who are creditworthy to signal this to lenders with higher accuracy. The benefits introduced through network based systems may help to overcome a portion of the financing problems, particularly if networks are created based on attributes correlated to financial health.

However, our analysis also raises an important concern about discrimination against already financially disadvantaged and "underbanked" groups. For instance, the U.S. Equal Credit Opportunity Act (ECOA) prohibits lenders to discriminate based on sex, race, color, religion, national origin, or age. To the extent that some of these characteristics correlates with creditworthiness and that homophily along those dimensions correlate with homophily along levels of creditworthiness, there is a concern that a side-effect of social credit scoring may be discrimination in access to credit along characteristics prohibited by the ECOA. Aside from strict legality, there is a concern that social scoring opens a backdoor to discrimination along dimensions that some may find objectionable. Our results also show that social scoring may lead people with low creditworthiness to prefer being discriminated against- in tie formation at least- to some moderate extent. The financial discrimination and social exclusion implications of social credit scoring, and how they balance against its benefits, warrant attention from policy makers and researchers alike.

6.3 Implications for management

To managers in the financial industry, our analysis suggests that lenders can expect to reduce their risk by incorporating network based measures in the short run. This dovetails with new

[12] It is important to note here that FICO and other leading institutions state that income is not a part of one's individual credit score, as it is a self-reported item of assessment.

[13] `http://www.creditsesame.com/about/press/consumers-who-earn-60000-or-less-have-dangerously-high-credit-usage-levels-according-to-credit-sesame/`

governmental policies on risk. For example, as part of the regulations posed by the Basel Committee on Banking and Supervision, banks in Europe have been encouraged to reduce the level of risk they undertake (Sousa et al., 2013). Regulations in the banking industry encourage financial institutions to better manage risk in the U.S. as well. These regulations have come at a time when big data analytics are enabling financial institutions to access larger and richer datasets. Indeed, it has been reported that social media and social network data are being used not only by start-ups, but also by established and more institutionalized credit scoring firms, such as Experian (Armour, 2014). The trend to use social data may prove to be useful as the U.S. recuperates in the post crisis environment.

Our study also offers a key insight to managers outside the financial industry who use social scoring for targeting customers when launching new products, targeting ads, or designing referral programs. The effectiveness of social scoring need not decrease when customers purposely adapt their networks in order to improve their score and their access to the benefits it entails.

Appendix

Proof of Proposition 1: Since the signals $\mathbf{y}$, once conditional on the types $\mathbf{x}$, are independent of the network, we have $\Pr(\mathbf{y}|\mathbf{x}) = \Pr(\mathbf{y}|g, \mathbf{x})$. Using Bayes' rule we have

$$\begin{aligned} \Pr(\mathbf{x}|g,\mathbf{y}) &\propto \Pr(\mathbf{x})\Pr(g,\mathbf{y}|\mathbf{x}) \\ &= \Pr(\mathbf{x})\Pr(\mathbf{y}|\mathbf{x})\Pr(g|\mathbf{x}) \end{aligned}$$

Thus

$$\begin{aligned} \Pr(\mathbf{x}|g,\mathbf{y}) \propto & \prod_{i\in\mathbf{y}} e^{-q{x_i}^2/2} \times \prod_{i\in\mathbf{y}} e^{-c(y_i-x_i)^2/2} \times \prod_{ij\in g^1} \nu e^{-(x_i-x_j)^2/2} \times \\ & \prod_{ij\in g^0: i,j\in\mathbf{y}} [1-\nu e^{-(x_i-x_j)^2/2}] \times \prod_{ij\in g^0: j\notin\mathbf{y}} \left(1-\frac{\mathbb{E}(n_i|x_i)}{S}\right). \end{aligned} \tag{A.1}$$

In the expression above, $(1 - \mathbb{E}(n_i|x_i)/S)$ is the probability that "i is not friends with j" for some i whose type is x_i and some j whose type is unknown. Fix some $i \in \mathbf{y}$ and consider

$\prod_{ij\in g^0:j\notin \mathbf{y}}(1-\mathbb{E}(n_i|x_i)/S)$. If $\{ij \in g^0 : j \notin \mathbf{y}\}$ is not empty, then by our assumption on the information structure, the product includes everyone in the rest of the society so its value under the limits of S, ν and q is

$$\lim_{S\to\infty,\mathbb{E}(n_i|x_i)\to N}\left(1-\frac{\mathbb{E}(n_i|x_i)}{S}\right)^{S-|y|}=e^{-N}$$

which isn't a function of $\mathbf{x}$, and thus does not contribute to the conditional density. Noticing that the rest of the terms in the right hand side of (A.1) are finite products, it is easy to see that as $\nu\to 0$ and $q\to 0$, in the limit,

$$\Pr(x|g,y)\propto\prod_{i\in\mathbf{y}}e^{-c(y_i-x_i)^2/2}\times\prod_{ij\in g^1}e^{-(x_i-x_j)^2/2}. \tag{A.2}$$

This implies that $\Pr(\mathbf{x}|g,\mathbf{y})$ is a multivariate normal density $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$. To find the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, all we need to do is matching the coefficients. The coefficients of x_i^2, x_ix_j and x_i in the quadratic form $-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})$ are $-\frac{1}{2}(\boldsymbol{\Sigma}^{-1})_{ii}$, $-(\boldsymbol{\Sigma}^{-1})_{ij}$ and $(\boldsymbol{\Sigma}^{-1})_{i1}\mu_1+(\boldsymbol{\Sigma}^{-1})_{i2}\mu_2+...$, while the corresponding coefficients of (A.2) are $-\frac{1}{2}(c+d_i)$, $1_{\{ij\in g^1\}}$ and cy_i. Matching them gives us the results in the Proposition. □

Proof of Corollary 1: This is just a special case of Proposition 1, where i is fixed and $\mathbf{y}=\{j|\, ij\in G\}$, $g^1=\{ij|\, ij\in G\}$ and $g^0=\{ij|\, ij\notin G, j\neq i\}$. □

Proof of Proposition 2: Let $\mathbf{D}$ be the diagonal matrix where $D_{ii}=c+d_i$, and $\mathbf{B}=\mathbf{D}^{-1}\mathbf{A}$ where $\mathbf{A}$ is the adjacency matrix of g^1. We can express the precision matrix by

$$\boldsymbol{\Sigma}=(\mathbf{I}-\mathbf{B})^{-1}\mathbf{D}^{-1}$$

Let $\mathbf{B}_0$ denote the matrix $\mathbf{B}$ when $c=0$. Since $\mathbf{B}_0$ is a stochastic matrix (i.e., each row summing to 1), its largest-magnitude eigenvalue is 1. When $c>0$, $\mathbf{B}$ is non-negative and it is easy to see that

$$\mathbf{B}<\delta\mathbf{B}_0$$

By the Perron-Frobenius Theorem, we know that the largest-magnitude eigenvalue of $\mathbf{B}$ is smaller than that of $\delta\mathbf{B}_0$, which is δ. Given that $\delta < 1$, we may write

$$\mathbf{\Sigma} = (I + \mathbf{B} + \mathbf{B}^2 + ...)\mathbf{D}^{-1}$$

Because for any $k \geq 1$, $\mathbf{B}^k$ is non-negative and $\left\|\mathbf{B}^k\right\| < \delta^k$, we have,

$$(\mathbf{B}^k)_{ij} < \delta^k$$

Now consider a node j whose distance from i in the sub-network defined by g^1 is $r(i,j) \geq 1$. Because $\mathbf{A}$ is the adjacency matrix of g^1, and there is no path between i and j whose length is less than $r(i,j)$, we know $(\mathbf{B}^k)_{ij} = 0$ for all $k < r(i,j)$. Hence an upper bound of $(I + \mathbf{B} + \mathbf{B}^2 + ...)_{ij}$ is $\delta^{r(i,j)}/(1-\delta)$. $\square$

Proof of Lemma 1: Derivation of the lemma follows similarly to the proof of Corollary 1. $\square$

Proof of Lemma 2: Under a symmetric rule λ, i and j become friends iff they have met and $m_{ij} > \lambda(x_i - x_j)^2$. Thus

$$\mathbb{E}_\lambda(n_i|x_i) = S\int_{-\infty}^{+\infty} \nu e^{-\lambda(t-x_i)^2/2}\sqrt{\frac{q}{2\pi}}e^{-qt^2/2}dt = S\nu\sqrt{\frac{q}{q+\lambda}}e^{-\frac{\lambda q}{\lambda+q}{x_i}^2/2}$$

Recall that $S\nu\sqrt{\frac{q}{q+1}} = N$. Taking $q \to 0$ gives the result. $\square$

Proof of Proposition 3: Let Λ denote the profile where i chooses λ_i and everyone else in the society chooses λ. For simplicity of notation, the expectation sign $\mathbb{E}(\cdot)$ throughout this proof refers to the expectation conditional on x_i under Λ, that is, $\mathbb{E}_\Lambda(\cdot|x_i)$. Similarly, the notation $\Pr(\cdot)$ also refers to the probability conditional on x_i under Λ.

Let's calculate the expected social utility, $\mathbb{E}\sum_{ij\in G}(m_{ij} - |x_j - x_i|)$, which we will denote more compactly as $\mathbb{E}u_i$. First we look at the type difference with friends. For any j we have:

$$\begin{aligned} \Pr(x_j, ij \in G) &= \Pr(x_j)\Pr(ij \in G|x_j) \\ &= \sqrt{\frac{q}{2\pi}}e^{-qx_j^2/2} \times \begin{cases} \nu e^{-\lambda_i(x_i-x_j)^2/2} & \text{if } x_j \leq x_i \\ \nu e^{-\lambda(x_i-x_j)^2/2} & \text{if } x_j > x_i. \end{cases} \end{aligned} \tag{A.3}$$

(A.3) enables us to calculate the probability of being friends with j:

$$\begin{aligned} \Pr(ij \in G) &= \int_{-\infty}^{+\infty} \Pr(x_j, ij \in G)dx_j \\ &= \frac{1}{2}\left(\frac{1}{\sqrt{\lambda_i}} + \frac{1}{\sqrt{\lambda}}\right)\nu\sqrt{\frac{q}{q+1}}e^{-\frac{q}{q+1}x_i^2/2} \end{aligned}$$

and in particular, its limiting value:

$$S\Pr(ij \in G) \to \frac{1}{2}N\left(\frac{1}{\sqrt{\lambda_i}} + \frac{1}{\sqrt{\lambda}}\right) \tag{A.4}$$

In a similar way, (A.3) also enables us to calculate the conditional type difference and its limiting value:

$$\begin{aligned} \mathbb{E}(-|x_j - x_i||ij \in G) &= \int_{-\infty}^{+\infty} -|x_j - x_i|\Pr(x_j|ij \in G)dx_j \\ &\to \sqrt{\frac{2}{\pi}}\left(\frac{1}{\lambda_i} + \frac{1}{\lambda}\right) \Big/ \left(\frac{1}{\sqrt{\lambda_i}} + \frac{1}{\sqrt{\lambda}}\right) \end{aligned} \tag{A.5}$$

Next we look at the matching value. We have

$$\begin{aligned} \Pr(m_{ij}, ij \in G) &= \Pr(m_{ij})\Pr(ij \in g|m_{ij}) \\ &= m_{ij}e^{-m_{ij}^2/2}\left(\nu\sqrt{\frac{q}{2\pi}}\int_{x_i-m_{ij}/\sqrt{\lambda_i}}^{x_i+m_{ij}/\sqrt{\lambda}} e^{-qx_j^2/2}dx_j\right) \\ S\Pr(m_{ij}, ij \in G) &\to \frac{N}{\sqrt{2\pi}}m_{ij}^2e^{-m_{ij}^2/2}\left(\frac{1}{\sqrt{\lambda}} + \frac{1}{\sqrt{\lambda_i}}\right) \end{aligned}$$

which, with (A.4), implies that

$$\Pr(m_{ij}|ij \in G) \to \sqrt{\frac{2}{\pi}} m_{ij}^2 e^{-m_{ij}^2/2}$$

This is the density of a χ_3 distribution. So we have

$$\mathbb{E}(m_{ij}|ij \in G) \to 2\sqrt{\frac{2}{\pi}} \tag{A.6}$$

Now using (A.4), (A.5) and (A.6), we have for the expected social utility

$$\begin{aligned}
\mathbb{E}u_i &= \sum_{j \neq i} \Pr(ij \in G) \left[\mathbb{E}(-|x_j - x_i||ij \in G) + \mathbb{E}(m_{ij}|ij \in G)\right] \\
&= S \Pr(ij \in G) \left[\mathbb{E}(-|x_j - x_i||ij \in G) + \mathbb{E}(m_{ij}|ij \in G)\right] \\
&\to \frac{N}{\sqrt{2\pi}} \left[2\left(\frac{1}{\sqrt{\lambda_i}} + \frac{1}{\sqrt{\lambda}}\right) - \left(\frac{1}{\lambda_i} + \frac{1}{\lambda}\right)\right] \qquad \text{(A.7)}
\end{aligned}$$

The second equality comes from the observation that both $\Pr(ij \in G)$, $\mathbb{E}(-|x_j - x_i||ij \in G)$ and $\mathbb{E}(m_{ij}|ij \in G)$ are not j-dependent (even without taking limits). A result from this is that the social utility is maximized at $\lambda_i = 1$, which should be the case.

Next we look at the expected utility from the network based score. Notice in this case, the bias from using network based scoring is:

$$\begin{aligned}
\mathbb{E}\mu_i(\lambda) - x_i &= \mathbb{E}\left[\frac{\lambda \sum_{ij \in G}(x_j - x_i)}{c + \lambda + \lambda n_i}\right] \\
&= \mathbb{E}\left[\frac{\lambda}{c + \lambda + \lambda n_i} \mathbb{E}\left(\sum_{ij \in G}(x_j - x_i)\middle| n_i\right)\right] \\
&= \mathbb{E}\left[\frac{\lambda n_i}{c + \lambda + \lambda n_i} \mathbb{E}\left(x_j - x_i \middle| ij \in G\right)\right] \\
&= \mathbb{E}\left(\frac{\lambda n_i}{c + \lambda + \lambda n_i}\right) \mathbb{E}\left(x_j - x_i \middle| ij \in G\right)
\end{aligned}$$

The first equality comes from the fact that y are unbiased signals. The second equality makes use of the iterated law of expectation. The last equality makes use of the fact that $\mathbb{E}(x_j - x_i \big| ij \in G)$ is not a function of n_i.

Using (A.3), we may calculate, in a way similar to (A.5),

$$\mathbb{E}\left(x_j - x_i \middle| ij \in G\right) \to \sqrt{\frac{2}{\pi}}\left(\frac{1}{\lambda} - \frac{1}{\lambda_i}\right) \Big/ \left(\frac{1}{\sqrt{\lambda}} + \frac{1}{\sqrt{\lambda_i}}\right)$$

So under the limits, we have the equality:

$$\mathbb{E}\mu_i(\lambda) - x_i = \underbrace{\mathbb{E}\left(\frac{\lambda n_i}{c + \lambda + \lambda n_i}\right)}_{\varphi} \times \underbrace{\sqrt{\frac{2}{\pi}}\left(\frac{1}{\lambda} - \frac{1}{\lambda_i}\right) \Big/ \left(\frac{1}{\sqrt{\lambda}} + \frac{1}{\sqrt{\lambda_i}}\right)}_{\xi} \tag{A.8}$$

It is more difficult to find an explicit expression for the limit of φ, so we will deal with it implicitly.

From this point on, notations $\mathbb{E}U_i$, $\mathbb{E}u_i$ and $\mathbb{E}\mu_i$ all refer to their limiting values. We want to find the "best response" correspondence for i, that is, the value of λ_i that maximizes $\mathbb{E}U_i$ for any λ. We will use the derivative of it:

$$F(\lambda_i, \lambda) := \frac{\partial \mathbb{E}u_i}{\partial \lambda_i} + \alpha \frac{\partial(\mathbb{E}\mu_i(\lambda) - x_i)}{\partial \lambda_i}$$

Note that by (A.8),

$$\frac{\partial(\mathbb{E}\mu_i(\lambda) - x_i)}{\partial \lambda_i} = \frac{\partial \varphi}{\partial \lambda_i}\xi + \frac{\partial \xi}{\partial \lambda_i}\varphi$$

Notice that (i) ξ has the same sign as $\lambda_i - \lambda$, (ii) $\partial\xi/\partial\lambda_i > 0$, and (iii) $0 < \varphi < 1$, (iv) $\partial\varphi/\partial\lambda_i < 0$. The first three points are easy to see. The last point can be seen by noticing that n_i is binomially distributed, and under the limits, Poisson distributed with expectation $S\Pr(ij \in G)$ as given in (A.4).

Using (i)-(iv), we have

$$\frac{\partial(\mathbb{E}\mu_i(\lambda) - x_i)}{\partial \lambda_i} < \frac{\partial \xi}{\partial \lambda_i}, \quad \text{for } \lambda_i \geq \lambda \tag{A.9}$$

and

$$\frac{\partial(\mathbb{E}\mu_i(\lambda) - x_i)}{\partial \lambda_i} > 0, \text{ at } \lambda_i = 1 \tag{A.10}$$

By (A.7), $\frac{\partial \mathbb{E}u_i}{\partial \lambda_i} = 0$ at $\lambda_i = 1$. Together with (A.10), we have

$$F(1, \lambda) > 0 \tag{A.11}$$

Next it is useful to look at a simpler case where φ is ignored in the derivative of the expected utility:

$$\widetilde{F}(\lambda_i, \lambda) := \frac{\partial \mathbb{E} u_i}{\partial \lambda_i} + \alpha \frac{\partial \xi}{\partial \lambda_i}$$

One can verify that as long as $\alpha < N$, $\widetilde{F}(\lambda_i, \lambda) \leq 0$ for $\lambda_i \geq \lambda^o$ where λ^o is the invariant solution to $\widetilde{F}(\cdot, \lambda) = 0$:

$$\lambda^o \equiv \left(1 - \frac{\alpha}{N}\right)^{-1}$$

Now using (A.5), we have

$$F(\lambda_i, \lambda) < \widetilde{F}(\lambda_i, \lambda) \leq 0, \text{ for any } \lambda_i \geq \max(\lambda, \lambda^o) \tag{A.12}$$

Define $\Xi_i(\lambda) := \mathrm{argmax}_{\lambda_i \geq 1} \mathbb{E} U_i$ to denote "best response" correspondence. Using Berge's Theorem one can show that it is upper-semicontinuous. Furthermore, (A.11) and (A.12) imply

$$1 < \Xi_i(\lambda) < \max(\lambda, \lambda^o)$$

This tells us that any fixed point of $\Xi_i(\cdot)$ must be between 1 and λ^o. Using Kakutani Fixed-Point Theorem one can show that a fixed point exists. □

Proof of Corollary 2: For the precision,

$$\begin{aligned} \mathbb{E}_\lambda[\rho_i(\lambda)] &= \left(1 + \frac{\lambda \mathbb{E}_\lambda n_i}{c + \lambda}\right) c \\ &= \left(1 + \frac{\sqrt{\lambda} N}{c + \lambda}\right) c \end{aligned}$$

The first equality uses (8) for the expression of $\rho_i(\lambda)$. The second equality comes from (9). This says that the expected precision, as a function of λ, is strictly quasi-concave and maximized at c. If $c \leq 1$, then the function is decreasing after 1. If $c \geq \sqrt{\frac{N}{N-\alpha}}$, then the function is no smaller at $(1 - \alpha/N)^{-1}$ than at 1, which, given the strict quasi-concavity and (10), implies that it is larger at λ^* than at 1. □

Proof of Lemma 3: Using Bayes' rule, we have

$$
\begin{aligned}
\Pr(x_i = h|\mathbf{y}_i) = & \Pr(\mathbf{y}_i|x_i = h)\Pr(x_i = h) \\
= & \sum_{\mathbf{x}_i}\left[\prod_{ij\in G}\nu\left(1_{\{x_j=h\}} + p1_{\{x_j=\ell\}}\right)\prod_{ij\in G}\Pr(y_j|x_j)\right]\times \\
& \prod_{ij\notin G, j\neq i}\left(1-\frac{1}{2}\nu(1+p)\right)\times\frac{1}{2}\Pr(y_i|x_i = h)
\end{aligned}
$$

where $\sum_{\mathbf{x}_i}$ is the summation across all possible vectors of friends' types, which has 2^{n_i} items. Another way of expressing the posterior is

$$
\begin{aligned}
\Pr(x_i = h|\mathbf{y}_i) \quad = \quad & \prod_{ij\in G}\nu\left[\Pr(y_j|x_j = h) + p\Pr(y_j|x_j = \ell)\right]\times \\
& \prod_{ij\notin G, j\neq i}\left(1-\frac{1}{2}\nu(1+p)\right)\times\frac{1}{2}\Pr(y_i|x_i = h)
\end{aligned} \tag{A.13}
$$

Similarly we can find the corresponding expression for $\Pr(x_i = \ell|\mathbf{y}_i)$.

$$
\begin{aligned}
\Pr(x_i = \ell|\mathbf{y}_i) \quad = \quad & \prod_{ij\in G}\nu\left[p\Pr(y_j|x_j = h) + \Pr(y_j|x_j = \ell)\right]\times \\
& \prod_{ij\notin G, j\neq i}\left(1-\frac{1}{2}\nu(1+p)\right)\times\frac{1}{2}\Pr(y_i|x_i = \ell)
\end{aligned}
$$

The ratio is then

$$
\frac{\Pr(x_i = \ell|\mathbf{y}_i)}{\Pr(x_i = h|\mathbf{y}_i)} = \left(\frac{\varepsilon}{1-\varepsilon}\right)^{y_i}\left(\frac{\varepsilon p + 1 - \varepsilon}{\varepsilon + p - \varepsilon p}\right)^{L_i}\left(\frac{p - \varepsilon p + \varepsilon}{1 - \varepsilon + \varepsilon p}\right)^{H_i}
$$

This ratio, together with $\Pr(x_i = \ell|\mathbf{y}_i) + \Pr(x_i = h|\mathbf{y}_i) = 1$, proves the proposition. □

Proof of Lemma 4: Similar to the proof of Lemma 3, we can find the expression $\Pr(x_i = h|\mathbf{y}_i)$, which turns out to be the same as (A.13). However the expression for $\Pr(x_i = \ell|\mathbf{y}_i)$ changes:

$$\begin{aligned}\Pr(x_i = \ell|\mathbf{y}_i) &= \prod_{ij\in G} \nu\left[p\Pr(y_j|x_j = h) + \widehat{p}\Pr(y_j|x_j = \ell)\right] \times \\ &\quad \prod_{ij\notin G, j\neq i} \left(1 - \frac{1}{2}\nu(\widehat{p} + p)\right) \times \frac{1}{2}\Pr(y_i|x_i = \ell)\end{aligned}$$

Thus the ratio becomes:

$$\begin{aligned}\frac{\Pr(x_i = \ell|\mathbf{y}_i)}{\Pr(x_i = h|\mathbf{y}_i)} &= \left(\frac{1 - \nu(\widehat{p} + p)/2}{1 - \nu(1 + p)/2}\right)^{S-n_i} \times \\ &\quad \left(\frac{\varepsilon}{1-\varepsilon}\right)^{y_i} \left(\frac{\varepsilon p + \widehat{p} - \varepsilon\widehat{p}}{\varepsilon + p - \varepsilon p}\right)^{L_i} \left(\frac{p - \varepsilon p + \varepsilon\widehat{p}}{1 - \varepsilon + \varepsilon p}\right)^{H_i}\end{aligned}$$

Taking limits of S and ν completes the proof. □

Proof of Proposition 4: We want to study the expected log odds as a function of θ. By Lemma 4, we have,

$$\begin{aligned}\mathbb{E}_\theta\left[R_i(\theta)|x_i = \ell\right] &= (1-2\varepsilon)\log\left(\frac{\varepsilon}{1-\varepsilon}\right) - \frac{1}{2}N\left[\varepsilon p + (1-\varepsilon)\widehat{p}\right]\log\left(\frac{\varepsilon p + (1-\varepsilon)\widehat{p}}{\varepsilon + (1-\varepsilon)p}\right) - \\ &\quad \frac{1}{2}N\left[\varepsilon\widehat{p} + (1-\varepsilon)p\right]\log\left(\frac{\varepsilon\widehat{p} + (1-\varepsilon)p}{\varepsilon p + (1-\varepsilon)}\right) - \frac{1}{2}N(1-\widehat{p}) \qquad \text{(A.14)}\end{aligned}$$

Since $\widehat{p} \equiv \int_\theta^{+\infty} f(t)dt$ where f is the density of the matching value, the derivative of the above expected log odds w.r.t. θ is

$$\begin{aligned}\frac{\partial\mathbb{E}_\theta\left[R_i(\theta)|x_i = \ell\right]}{\partial\theta} &= \frac{\partial\mathbb{E}_\theta\left[R_i(\theta)|x_i = \ell\right]}{\partial\widehat{p}} \cdot \frac{\partial\widehat{p}}{\partial\theta} \\ &= \frac{1}{2}N\left[(1-\varepsilon)\log\left(\frac{\varepsilon p + (1-\varepsilon)\widehat{p}}{\varepsilon + (1-\varepsilon)p}\right) + \varepsilon\log\left(\frac{\varepsilon\widehat{p} + (1-\varepsilon)p}{\varepsilon p + (1-\varepsilon)}\right)\right] f(\theta) \qquad \text{(A.15)}\end{aligned}$$

Note that the term in the brackets is strictly increasing in $\widehat{p}$, thus strictly decreasing in θ, and by assumption f is positive. Thus the expected log odds is either strictly monotonic or is strictly increasing up to some point then strictly decreasing. This implies strict quasi-concavity.

Furthermore, when θ is close to 0, $\widehat{p}$ is close to 1, so the derivative is positive; when θ is large, $\widehat{p}$ is close to 0, so the derivative is negative. This implies that expected log odds has a maximum somewhere.

Finally, when $\theta = \overline{\theta}$ (thus $\widehat{p} = p + \frac{\varepsilon(1-p)}{1-\varepsilon}$), it can be shown that the derivative is negative. This implies that the maximizer is on the left side of $\overline{\theta}$. $\square$

Proof of Proposition 5: Let Θ denote the profile where everyone else chooses θ and i chooses θ_i. For the simplicity of notation, the expectation sign $\mathbb{E}(\cdot)$ throughout this proof refers to the expectation conditional on $x_i = \ell$ under Θ, that is, $\mathbb{E}_\Theta(\cdot|x_i = \ell)$.

With $P_i(\theta)$ given by (16), we have for any $\theta_i \geq \theta$,

$$\begin{aligned}\mathbb{E}R_i(\theta) &= (1-2\varepsilon)\log\left(\frac{\varepsilon}{1-\varepsilon}\right) - \frac{1}{2}N\left[\varepsilon p + (1-\varepsilon)\int_{\theta_i}^{+\infty} f(t)dt\right]\log\left(\frac{\varepsilon p + (1-\varepsilon)\widehat{p}}{\varepsilon + (1-\varepsilon)p}\right) - \\ &\quad \frac{1}{2}N\left[\varepsilon\int_{\theta_i}^{+\infty} f(t)dt + (1-\varepsilon)p\right]\log\left(\frac{\varepsilon\widehat{p} + (1-\varepsilon)p}{\varepsilon p + (1-\varepsilon)}\right) - \frac{1}{2}N(1-\widehat{p})\end{aligned}$$

where $\widehat{p} \equiv \int_\theta^{+\infty} f(t)dt$ is the probability that the matching value is larger than θ. Thus for any $\theta_i > \theta$,

$$\frac{\partial \mathbb{E}R_i(\theta)}{\partial \theta_i} = \frac{1}{2}N\left[(1-\varepsilon)\log\left(\frac{\varepsilon p + (1-\varepsilon)\widehat{p}}{\varepsilon + (1-\varepsilon)p}\right) + \varepsilon\log\left(\frac{\varepsilon\widehat{p} + (1-\varepsilon)p}{\varepsilon p + (1-\varepsilon)}\right)\right] f(\theta_i) \tag{A.16}$$

Also, using $\mathbb{E}u_i$ short for the social utility $\mathbb{E}\sum_{ij\in G}\left(m_{ij} - 1_{\{x_j=h\}}\right)$, we have for any $\theta_i \geq \theta$,

$$\mathbb{E}u_i = \frac{1}{2}N\left(\int_1^{+\infty}(t-1)f(t)dt + \int_{\theta_i}^{+\infty} tf(t)dt\right)$$

Thus for any $\theta_i > \theta$,

$$\frac{\partial \mathbb{E}u_i}{\partial \theta_i} = -\frac{1}{2}N\theta_i f(\theta_i) \tag{A.17}$$

With (A.16) and (A.17), first let's look the potential equilibrium where everyone chooses 0. When $\theta = 0$ (thus $\widehat{p} = 1$), it is seen that for $\theta_i > 0$ that is sufficiently close to 0,

$$\alpha\frac{\partial \mathbb{E}R_i(\theta)}{\partial \theta_i} > -\frac{\partial \mathbb{E}_\Theta u_i}{\partial \theta_i}$$

which means $\partial \mathbb{E}U_i/\partial\theta_i > 0$. This tells us that when everyone uses $\theta = 0$, any individual has the incentive to increase its own criterion above 0. So this cannot be an equilibrium.

Next we look at the potential equilibiria in the left neighborhood of $\underline{\theta}$. Note that (A.15) and (A.16) (coincidentally) have the same expression. This tells us that (A.16) is zero at $\theta = \underline{\theta}$. Using this fact, we see that by fixing a $\theta < \underline{\theta}$ that is sufficiently close to $\underline{\theta}$, we have for all $\theta_i > \theta$,

$$\alpha \frac{\partial \mathbb{E} R_i(\theta)}{\partial \theta_i} < \frac{1}{2} N \theta f(\theta_i) \leq -\frac{\partial \mathbb{E}_\Theta u_i}{\partial \theta_i}$$

which means $\partial \mathbb{E} U_i / \partial \theta_i < 0$. This tells us that when everyone chooses this θ, any $\theta_i > \theta$ is not a profitable deviation for i. Since friendship must be mutual, $\theta_i < \theta$ does not make a difference and is not a profitable deviation either. So $\theta^* = \theta$ is an equilibrium. □

Proof of Lemma 5: This is a special case of Lemma 4, with $\widehat{p} = p$. □

Proof of Proposition 6: Let Θ denote the profile where everyone else chooses θ and i chooses θ_i. For the simplicity of notation, the expectation sign $\mathbb{E}(\cdot)$ throughout this proof refers to the expectation conditional on $x_i = \ell$ under Θ, that is, $\mathbb{E}_\Theta(\cdot | x_i = \ell)$.

With $P_i(\theta)$ given still by (16), $\mathbb{E} R_i(\theta)$ is expressed similarly as in the proof of Proposition 5, and its derivative w.r.t. θ_i for $\theta_i > \theta$ is given by (A.16).

Also, using $\mathbb{E} u_i$ short for the social utility $\mathbb{E} \sum_{ij \in G} \left(m_{ij} - 1_{\{x_j = \ell\}} \right)$, we have for any $\theta_i \geq \theta$,

$$\mathbb{E} u_i = \frac{1}{2} N \left(\int_1^{+\infty} t f(t) dt + \int_{\theta_i}^{+\infty} (t-1) f(t) dt \right)$$

Thus

$$\frac{\partial \mathbb{E} u_i}{\partial \theta_i} = -\frac{1}{2} N (\theta_i - 1) f(\theta_i)$$

It is not difficult to see that when $\theta \leq \underline{\theta}$, we have for any $\theta_i > \theta$ that is sufficiently close to θ,

$$\alpha \frac{\partial \mathbb{E} R_i(\theta)}{\partial \theta_i} \geq 0, \quad \frac{\partial \mathbb{E} u_i}{\partial \theta_i} > 0$$

which imply $\partial \mathbb{E} U_i / \partial \theta_i > 0$. This says that when everyone uses $\theta \leq \underline{\theta}$, any individual has the incentive to increase its own criterion above θ. So these cannot be equilibria.

Fixing a $\theta < 1$ that is sufficiently close to 1 (thus $\widehat{p}$ is close to p), we have for all $\theta_i > \theta$,

$$\alpha \frac{\partial \mathbb{E} R_i(\theta)}{\partial \theta_i} < \frac{1}{2} N(\theta - 1) f(\theta_i) \leq -\frac{\partial \mathbb{E} u_i}{\partial \theta_i}$$

which implies $\partial \mathbb{E} U_i / \partial \theta_i < 0$. This says that when everyone chooses θ, any $\theta_i > \theta$ is not a profitable deviation for i. Since friendship must be mutual, $\theta_i < \theta$ does not make a difference and is not a profitable deviation either. So $\theta^* = \theta$ is an equilibrium. □

Proof of Proposition 7: Taking the derivative of Equation (20) w.r.t. e_i, we have the first-order condition each i:

$$e_i^* = \frac{a}{b} + \phi \left(H_i + \sum_{ij \in G_\ell} e_j^* \right)$$

which we may write in the matrix form:

$$\mathbf{e}^* = \frac{a}{b} + \phi \mathbf{H} + \phi \mathbf{A}_\ell \mathbf{e}^*$$

This implies

$$\left(\mathbf{I} - \phi \mathbf{A}_\ell \right) \mathbf{e}^* = \frac{a}{b} + \phi \mathbf{H}$$

By the Perron-Frobenius Theorem, the largest-magnitude eigenvalue of $\mathbf{A}_\ell$ is real and positive. Furthermore, if this eigenvalue is smaller than $|\phi|^{-1}$, then $\|\phi \mathbf{A}_\ell\| < 1$. This implies that the series $\sum_{k=0}^{\infty} \phi^k \mathbf{A}_\ell^k$ exists. One can readily check that the series is the inverse of $(\mathbf{I} - \phi \mathbf{A}_\ell)$. □

References

Ambrus, A., M. Mobius, and A. Szeidl (2014). Consumption Risk-Sharing in Social Networks. *American Economic Review 104*(1), 149–182.

Armour, S. (January 8, 2014). Borrowers Hit Social-media Hurdles. *Wall Street Journal Online*. Accessed: 2014-06-30.

Arrow, K. J. (1998). What Has Economics to Say About Racial Discrimination? *Journal of Economic Perspectives 12*(2), 91–100.

Bagherjeiran, A., R. P. Bhatt, R. Parekh, and V. Chaoji (2010). Online Advertising in Social Networks. In B. Furht (Ed.), *Handbook of Social Network Technologies and Applications*, pp. 651–689. New York, NY: Springer.

Bakshy, E., D. Eckles, R. Yan, and I. Rosenn (2012). Social Influence in Social Advertising: Evidence from Field Experiments. In *Proceedings of the 13th ACM Conference on Electronic Commerce (EC'12)*, pp. 146–161.

Ball, S., C. Eckel, P. Grossman, and W. Zame (2001). Status in Markets. *Quarterly Journal of Economics 116*(1), 161–188.

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

Becker, G. S. (1971). *The Economics of Discrimination (2nd Edition).* Chicago, IL: University of Chicago Press.

Bonacich, P. (1987). Power and Centrality: A Family of Measures. *American Journal of Sociology 92*(5), 1170–1182.

Bramoullé, Y. and R. Kranton (2007a). Risk Sharing Across Communities. *American Economic Review 97*(2), 70–74.

Bramoullé, Y. and R. Kranton (2007b). Risk-sharing Networks. *Journal of Economic Behavior & Organization 64*(3-4), 275–294.

Chui, M. (2013). Social Media to Boost Financial Services. *InFinance: The Magazine for Finsia Members 127*(2), 34–36.

Demirgüç-Kunt, A. and R. Levine (2009). Finance and Inequality: Theory and Evidence. *Annual Review of Financial Economics 1*(1), 287–318.

Feigenberg, B., E. M. Field, and R. Pande (2010). Building Social Capital Through Microfinance. Working Paper 16018, National Bureau of Economic Research.

Galak, J., D. Small, and A. T. Stephen (2011). Micro-finance Decision Making: A Field Study of Prosocial Lending. *Journal of Marketing Research 48*(SPL), S130–S137.

Goel, S. and D. G. Goldstein (2013). Predicting Individual Behavior with Social Networks. *Marketing Science 33*(1), 82–93.

Guseva, A. and A. Rona-Tas (2001). Uncertainty, Risk, and Trust: Russian and American Credit Card Markets Compared. *American Sociological Review 66*(5), 623–646.

Haenlein, M. (2011). A Social Network Analysis of Customer-level Revenue Distribution. *Marketing letters 22*(1), 15–29.

Haenlein, M. and B. Libai (2013). Targeting Revenue Leaders for a New Product. *Journal of Marketing 77*(3), 65–80.

Herzenstein, M., S. Sonenshein, and U. M. Dholakia (2011). Tell me a Good Story and I May Lend You Money: The Role of Narratives in Peer-to-peer Lending Decisions. *Journal of Marketing Research 48*(SPL), S138–S149.

Hill, S., F. Provost, and C. Volinsky (2006). Network-based Marketing: Identifying Likely Adopters via Consumer Networks. *Statistical Science 21*(2), 256–276.

Hofstede, G. H. (2001). *Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations Across Nations*. Thousand Oaks, CA: Sage Publications.

Hu, Y. and C. Van den Bulte (2014). Nonmonotonic Status Effects in New Product Adoption. *Marketing Science, in press*.

Jenkins, P. (2014). Big Data Lends New Zest to Banks' Credit Judgments. *Financial Times (U.S. ed.)*. Accessed: June 24, 2014.

Kornish, L. J. and Q. Li (2010). Optimal Referral Bonuses with Asymmetric Information: Firm-offered and Interpersonal Incentives. *Marketing Science 29*(1), 108–121.

Lin, M., N. R. Prabhala, and S. Viswanathan (2013). Judging Borrowers by the Company They Keep: Friendship Networks and Information Asymmetry in Online Peer-to-peer Lending. *Management Science 59*(1), 17–35.

Liu, K. and L. Tang (2011). Large-scale Behavioral Targeting with a Social Twist. In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*, pp. 1815–1824.

Lu, Y., K. Jerath, and P. V. Singh (2013). The Emergence of Opinion Leaders in a Networked Online Community: A Dyadic Model with Time Dynamics and a Heuristic for Fast Estimation. *Management Science 59*(8), 1783–1799.

Phelps, E. S. (1972). The Statistical Theory of Racism and Sexism. *American Economic Review 62*(4), 659–661.

Podolny, J. M. (2008). *Status Signals: A Sociological Study of Market Competition.* Princeton, NJ: Princeton University Press.

Rona-Tas, A. and A. Guseva (2014). *Plastic Money: Constructing Markets for Credit Cards in Eight Postcommunist Countries.* Stanford, CA: Stanford University Press.

Rosenblat, T. S. and M. Mobius (2004). Getting Closer or Drifting Apart? *Quarterly Journal of Economics 119*(3), 971–1009.

Roth, P. L., P. Bobko, C. H. Van Iddekinge, and J. B. Thatcher (2013). Social Media in Employee-Selection-Related Decisions: A Research Agenda for Uncharted Territory. *Journal of Management, in press*.

Rusli, E. M. (April 1, 2013). Bad Credit? Start Tweeting. *Wall Street Journal*. Accessed: 2014-04-10.

Schmitt, P., B. Skiera, and C. Van den Bulte (2011). Referral Programs and Customer Value. *Journal of Marketing 75*(1), 46–59.

Sousa, M. R., J. Gama, and E. Brandao (2013). Introducing Time-changing Economics into Credit Scoring. Technical Report 513, University of Porto.

Stiglitz, J. E. (1990). Peer Monitoring and Credit Markets. *World Bank Economic Review 4*(3), 351–366.

Sudhir, K., J. Priester, M. Shum, D. Atkin, A. Foster, G. Iyer, G. Jin, D. Keniston, S. Kitayama, M. Mobarak, Y. Qian, I. Tewari, and W. Wood (2014). Emerging Markets. *Yale School of Management, working paper*.

Toubia, O. and A. T. Stephen (2013). Intrinsic vs. Image-Related Utility in Social Media: Why Do People Contribute Content to Twitter? *Marketing Science 32*(3), 368–392.

Townsend, R. M. (1994). Risk and Insurance in Village India. *Econometrica 62*(3), 539–91.

van Alstyne, M. and E. Brynjolfsson (2005). Global Village or Cyber-Balkans? Modeling and Measuring the Integration of Electronic Communities. *Management Science 51*(6), 851–868.

Zeng, Z. and Y. Xie (2008). A Preference-opportunity-choice Framework with Applications to Intergroup Friendship. *American Journal of Sociology 114*(3), 615.