Marketing Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Credit Scoring with Social Network Data

Yanhao Wei, Pinar Yildirim, Christophe Van den Bulte, Chrysanthos Dellarocas

MARKETING SCIENCE
Vol. 35, No. 2, March–April 2016, pp. 234–258
ISSN 0732-2399 (print) | ISSN 1526-548X (online)

http://dx.doi.org/10.1287/mksc.2015.0949

Credit Scoring with Social Network Data

Yanhao Wei
Department of Economics, University of Pennsylvania, Philadelphia, Pennsylvania 19104, yanhao@sas.upenn.edu

Pinar Yildirim, Christophe Van den Bulte
Marketing Department, The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104
{pyild@wharton.upenn.edu, vdbulte@wharton.upenn.edu}

Chrysanthos Dellarocas
Information Systems Department, Questrom School of Business, Boston University, Boston, Massachusetts 02215, dell@bu.edu

Motivated by the growing practice of using social network data in credit scoring, we analyze 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 consumers can strategically construct their social networks to attain higher scores. We find that those who are motivated to improve their scores may form fewer ties and focus more on similar partners. The impact of such endogenous tie formation on the accuracy of consumer scores is ambiguous. Scores can become more accurate as a result of modifications in social networks, but this accuracy improvement may come with greater network fragmentation. 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 managers and public policy.

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

History: Received: July 18, 2014; accepted: June 21, 2015; K. Sudhir served as the senior editor and Yuxin Chen served as associate editor for this article. Published online in *Articles in Advance* October 26, 2015.

1. Introduction

When a consumer applies for credit, attempts to refinance 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 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 loan obligations (Rusli 2013). Given that consumers 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 access to opportunities and hence on social inequality among citizens.

Until recently, assessing consumers' creditworthiness relied solely on their financial history. The financial credit score popularized by the Fair Isaac 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. In the past few years, however, the credit scoring industry has witnessed a dramatic change in data sources (Chui 2013, Jenkins 2014, Lohr 2015). An increasing number of firms rely on network-based data to assess consumer creditworthiness. One such company, Lenddo, reportedly assigns credit scores based on information in users' social networking profiles, such as education and employment history, how many followers they have, who they are friends with, and information about those friends (Rusli 2013).[1] 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 consumers 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,

[1] Network data can be collected from a variety of sources. Lenddo, for instance, obtains applicants' consent to scan a variety of their online social accounts (Facebook, Gmail, Twitter, LinkedIn, Yahoo, Microsoft Live) and sometimes also their phone activity.

from the perspective of lenders, is there an advantage to using network-based measures rather than measures based only on an individual's data? Second, as 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 operate in network-based credit scoring? Finally, and most important 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 the extensive and intensive margins, i.e., by increasing the number of those 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 (§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 consumers and reduce the uncertainty about their creditworthiness. We find that the accuracy of network-based scores depends primarily on information from the direct ties, i.e., the assessed consumers' ego-network. This implies that credit-scoring firms can efficiently assess an individual's creditworthiness using data from a subset of the overall network.

In §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 not to connect to people with lower scores. This can result in social fragmentation within a network: Those with better access to financing opportunities choose to segregate themselves from those with worse financing opportunities. As a result, consumers self-select into highly homogeneous yet smaller subnetworks. 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 is situated 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 consumers with poor financing opportunities in further hardship.

This result supports concerns about social credit scoring from consumer advocates and regulators such as the Consumer Financial Protection Bureau (CFPB) and the Federal Trade Commission (FTC) (Armour 2014).

In §§2 and 3, we study environments wherein all consumers, independent of their type, have similar needs for financing. We relax this assumption in §4 and introduce a formulation with discrete risk types that may vary in their needs for financing. 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 consumers. 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 different from between-group discrimination studied commonly in the literature (e.g., Arrow 1998, Becker 1971, Phelps 1972).

In §5, again within a discrete setting, we allow consumers to exert effort to improve their true creditworthiness or type. When social ties motivate effort, social credit scoring may benefit those with poor financial health in two ways, i.e., 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. We consider environments with explicit discrimination and with homophily. We find that when there are complementarities between the effort exerted by individuals, the between-group connections can motivate effort and thus lead to increased social mobility in both environments. The within-group connections also improve effort in a discriminatory environment. By contrast, when homophily is the only factor determining tie formation, a high number of low-type friends who exert low effort will reduce an individual's desire to exert effort. In §6, we analyze another way consumers can exert effort to improve their financial outcomes, i.e., by actively networking to endogenously alter the probability of meeting people with high creditworthiness. Our analysis demonstrates that low types exert effort to meet others more aggressively than high types only when they are in dire need of improving credit access. Otherwise, high types exert greater effort.

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 such as 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 the unattractive, and vice versa, underlies social customer scoring in predictive analytics (e.g., Benoit and Van den Poel 2012, Goel and Goldstein 2013, Haenlein 2011). It is also 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), in targeted online advertising (Bagherjeiran et al. 2010, Bakshy et al. 2012, Liu and Tang 2011), and in customer referral programs (e.g., Kornish and Li 2010, Schmitt et al. 2011). The basic insights also apply to employment settings, where firms have long used employee referral programs to attract better applicants (e.g., Castilla 2005) and many have started to use social network data to gain more information about applicants' character and work ethic (e.g., Roth et al. 2016).

The model construct that we label "social credit score" 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, p. 510). As such, our analysis involving endogenous tie formation contributes not only to research traditions in economics and sociology (e.g., Ball et al. 2001, Podolny 2008) but also to the recent marketing research on how status considerations affect consumers' networking behavior (Lu et al. 2013, Toubia and Stephen 2013), their acceptance of new products (Iyengar et al. 2015), 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 microfinance 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 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 form their network ties differently and to exert more effort in improving their financial health is different from, yet dovetails with, the evidence by Feigenberg et al. (2013) that group lending tends to trigger changes in network structure that in turn reduce loan defaults. The two different kinds of "social financing" practices acting at two different stages of the loan (customer selection and terms definition versus compliance) can lead to improved outcomes mediated through endogenous 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 consumers. There are two ways through which such discrimination can come about. The first is that consumers 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 consumers to protect their credit score. Formation of networks to attain a high credit score can be an indirect way of discrimination because some consumers are systematically excluded from others' networks. The second is that consumers may observe each other's effort to improve their score and may discriminate based on personal effort. Any low-type consumer who does not exert effort may face disengagement by fellow low-type contacts who exert effort and who want to disassociate their own credit score from hers.

Third, our work is 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 scoring in general) may foster or harm integration within versus 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 found lacking. Creditors in these markets often seek to enrich scores based on an individual's history with additional information (e.g., Guseva and Rona-Tas 2001, Sudhir et al. 2015, Rona-Tas and Guseva 2014).

The rest of the article develops as follows. In §2, we present a benchmark model with data collection from networks to assess creditworthiness, and then provide justification for the emergence of this industry. In §3, we investigate the possibility of networks forming endogenously to the social credit scoring practice. We extend our model to allow consumers to vary in their financing needs in §4. We consider the possibility of social mobility through effort in §5. We extend the model in several directions in §6 and conclude with implications for public policy and marketing practice in §7.

2. Model with Exogenous Network

Consider a society with a large population S. Each person i is denoted by a type x_i, and x_i follows $N(0, q^{-1})$ across individuals, with precision $q > 0$. We assume that each agent knows her own type and discovers that of fellow consumers upon meeting them.

The process of forming friendships is specified as follows. Each pair of consumers 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 on preference rather than opportunity (Zeng and Xie 2008): Individuals enjoy the company of others like them more than that of others unlike them. Person i accepts the formation of a friendship tie with j, iff, they have met and

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

On 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 $\mathbb{E}(n_i \mid x_i) = S\nu\sqrt{q/(q+1)}e^{-(q/(1+q))x_i^2/2}$.[2] To represent an environment with sufficient uncertainty about the creditworthiness of consumers, 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 characterize a society with sufficient uncertainty about individuals. They also allow us to assume that the product term $S\nu\sqrt{q/(q+1)}$ holds a constant, which we denote by N.[3]

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

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

where $\varepsilon_i \sim N(0, c^{-1})$ and is independent across individuals. The firm observes the signals of a finite set of consumers $\mathbf{y}$, which we refer to as the vector of signals as well. For these consumers, 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 that the lender knows are friends, and g^0 is the set of the dyads that the lender knows are not friends. Furthermore, for each person in $\mathbf{y}$, we allow g^0 to include all of the dyads that involve her and someone outside $\mathbf{y}$.[4]

First, we present some properties about the firm's posterior on the types of consumers in a network. Together with the nodes in $\mathbf{y}$, the ties in g^1 define a subnetwork involving only nodes on which a signal is observed. In this subnetwork, let d_i be the degree of i,[5] and $r(i, j)$ be the length of the shortest path (i.e., geodesic distance) between i and j.

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

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

and mean vector μ

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

Proposition 1 states that the lender's beliefs about the types of consumers in the network follow a multivariate normal distribution the parameters of which depend on the network structure. So two consumers with identical individual signals (such as personal financial history) may obtain different network-based scores because of social connections. These consumers 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 score gains and losses from the social network.

Equation (4) shows that the weight that contact j's signal receives depends on her location in 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 $r(i, j)$. If i and j are not connected in the subnetwork, the weight is zero.

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

[3] In a small society where everyone is likely to be friends with others, or in a society where each type is organized in perfectly homogeneous and mutually disconnected subgraphs (i.e., components), there is little to no uncertainty about an individual's type. This implies that network-based scores are less useful.

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

[5] Note that d_i, the observed degree of i need not be the same as her true degree, n_i, as here we allow for observing any subnetwork of friends, $d_i \le n_i$.

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

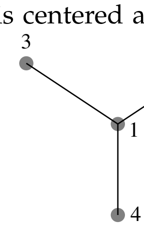
$$c\boldsymbol{\Sigma}_{ij} < \frac{c}{c+d_i}\frac{\delta^{r(i,j)}}{1-\delta},$$

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.

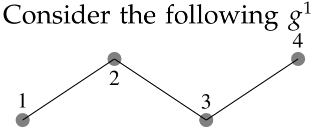


With $c=1$, $c\boldsymbol{\Sigma}$ equals

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

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 a firm's updating of beliefs about a consumer's type. 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 an equal distance to persons 1 and 3, their signals receive different weights: Individual 3's signal is diluted as she is linked to individual 4.

Propositions 1 and 2 together imply that agents who have lower distances to high-type consumers can receive a more favorable posterior in credit score assessment. Conversely, proximity to those with low signals may hurt an individual's assessment. Consumers cannot choose their distance as we have not yet considered active selection of friendship ties to attain such benefits (see §3). When the weight of a friend j's signal (on updating the beliefs about the type of i) is zero, this implies that either it is unknown whether there is a friendship between the ego and j, or that $j \in g^0$ and they are not friends. When two people are not friends, the interpretation is that they have not met due to the low meeting probability.

In the remainder of the paper, we assume that when evaluating a particular i, the firm observes the complete ego-network of i, i.e., all of the ties $ij \in G$, and receives a signal on each of i's friends. We collect the signals in the vector $\mathbf{y}_i$, which we will refer to as the set of i's friends. Note that this imposes an additional assumption on the previous analysis: We now require that g^1 equals 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 risk assessment of type i, $\Pr(x_i \mid \mathbf{y}_i)$ is normal with precision*

$$\rho_i = c + \frac{c}{c+1}n_i, \tag{5}$$

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 her type will be more precise. The assessment of an individual with a higher degree is likely to be closer to this true type, x_i.[6] More important, (5) implies that the precision of a 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 receive gradually lower weights in the beliefs about one's creditworthiness.

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

3. Endogenous Tie Formation

We next study consumers' incentives to form network ties 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, a consumer has an incentive not to form ties with low types 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

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

where the first part of the utility, $(m_{ij} - |x_i - x_j|)$, indicates a social utility taking into consideration homophily. 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 versus the utility from friendship ties with people she likes. All consumers gain utility from their posterior credit score at rate α.[7] If $\alpha = 0$, the individual cares only about forming friendships for social utility. If $\alpha \to +\infty$, then the agent cares little about social utility but cares greatly about improving her score.

Parameter α can also be interpreted as a measure of the desire for status. How much people care about how highly 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. Let each consumer i adopt a tie formation rule a priori (i.e., before meeting j) which states that she will accept friendship with j, iff,

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

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 simultaneously[8] and optimally by consumers. Although individual i would prefer to be friends with others similar to her, which was expressed in (1), she may have additional utility from adding high type or removing low type friends due to the improvement in her credit assessment. This suggests that consumers 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 $\lambda_i \geq 1$.[9] 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.[10] 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 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 with the tie formation probability in an exogenous setting (given by Equation (2)), consumers will be more selective in linking to others. Fewer ties will be formed in the endogenous case.

3.1. Credit Scoring with Endogenous Tie Formation

In this section we complete the analysis of endogenous relationship formation using an equilibrium concept. We use (λ, λ_i) to denote the common rule with the possible deviation of i. The expected utility of i becomes

$$\mathbb{E}(U_i \mid x_i, \lambda, \lambda_i) = \mathbb{E}\left(\sum_{j:\, ij \in G} m_{ij} - |x_i - x_j| \,\middle|\, x_i, \lambda, \lambda_i \right) + \alpha\,\mathbb{E}[\mu_i(\lambda) \mid x_i, \lambda, \lambda_i], \qquad (7)$$

[7] To allow for the possibility that some agents may have no interest in improving their scores when they meet others with similar types, §4 presents a discrete formulation of our matching model and we provide a special case wherein the high types have zero utility from credit scores.

[8] Note that in this model consumers form ties simultaneously. A model with sequential friendship formation would need to consider, in addition to tie formation rules, rules about the order in which consumers form ties, and would need to assume that individual beliefs about firms' financial assessment are consistent with equilibrium outcomes.

[9] The benefits of network-based scoring are 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 rejects ties with 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 important, (ii) there is no symmetric equilibrium where $\eta > 1$ or $\lambda < 1$.

[10] If we allowed consumers to form friendships without mutual consent, then everyone could link to anyone to improve her own score. The benefits of network-based scoring would be limited since a connection to a high type would not be informative of one's type.

where $\mu_i(\lambda) = \mathbb{E}(x_i \mid \mathbf{y}_i, \lambda)$ is the lender's posterior. Each person calculates her expected utility from being in a friendship network before the network is formed, implying that expected utility will depend on the friendship rule (λ, λ_i) adopted. The expectation $\mathbb{E}(\cdot)$ is taken before meeting others. We first display a version of Corollary 1 under a symmetric rule. In the following, when λ_i conforms with the common rule, we omit λ_i in the expectation conditionals.

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

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

and mean

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

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 consumers react to an environment with network-based scoring, will scores be less or more precise? In other words, will assessments based on network data yield a better assessment? Our answer to this question is a qualified yes. We explain the mechanism through which this improvement can be achieved via a lemma and a proposition.

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

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

A lower rate of mixing between types (a higher λ) results in a smaller number of ties per person. Ties are formed only between those 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 yet clear. We address it next.

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

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

$$1 < \lambda^* < \left(1 - \frac{\alpha}{N}\right)^{-2}. \tag{10}$$

In words, when networks are created endogenously, consumers are more selective in accepting friendships in equilibrium; the upper bound on selectivity is determined by how much importance consumers put on a high credit score and the expected degree in society.

Corollary 2. *If $c \geq N/(N-\alpha)$, then*

$$\mathbb{E}[\rho_i(\lambda^*) \mid \lambda^*] > \mathbb{E}[\rho_i(1) \mid \lambda = 1],$$

where $\rho_i \equiv \text{Precision}(x_i \mid \mathbf{y}_i, \lambda)$. 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 $\mathbb{E}[\rho_i(\lambda^*) \mid \lambda^*] < \mathbb{E}[\rho_i(1) \mid \lambda = 1]$. On average, the network-based scores are less accurate.

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

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

Corollary 2 identifies a further condition 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 consumers to form networks to improve their credit scores. The improvement in precision is conditional on consumers placing sufficiently low weight on financial outcomes relative to the utility derived from social connections. Paradoxically, when consumers 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 the density of relationships is generally higher. Lenddo, for instance, operates in Mexico, Colombia, and the Philippines, and reports that Mexico is its fastest growing market.[11]

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 consumers can borrow, even if they would qualify to receive credit using the individual score system. For simplicity and concreteness of discussion, we specify the perceived probability of repayment of credit by consumer 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 consumer i with type x_i. The firm determines the rate by solving

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

This formulation takes into account not only the expected creditworthiness of a consumer, 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 versus repayment

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

As the consumer's likelihood of a default increases, she faces a higher borrowing rate. Note that the financial utility of a consumer given in Equation (6) can be derived by assuming that the lending rate enters the utility through $-\alpha\log(r_i)$. If lending rates can be interpreted in the context of economic opportunities available to consumers, then a consumer 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.

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

4. Role of Signals from Social Contacts

In the preceding sections, we developed a model with continuous types and assumed that every individual had identical incentives to improve her credit score. In reality, there may be differences among consumers about how much utility they can gain from improving their credit score conditional on their type. In this section, we introduce a discrete version of the model to allow for this possibility. The discrete version allows us to analyze in greater detail how the firm uses signals of low versus high type friends when assessing a consumer'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 $\Pr(x_i = l) = \Pr(x_i = h) = \frac{1}{2}$. Whereas high types have a low risk of credit default, low types have a higher risk. With probability ν, any two consumers will meet. On 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-based 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 her social network.

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

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

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 3 suggests that low- and high-type signals observed for a consumer's social connections affect the lender's assessment of that consumer's creditworthiness in different directions. Note that $(\varepsilon + (1-\varepsilon)p)/(\varepsilon p + (1-\varepsilon)) < 1$ and $(\varepsilon p + (1-\varepsilon))/(\varepsilon + (1-\varepsilon)p) > 1$. Thus, high-type signals increase the likelihood that an agent will be categorized as high type, whereas low-type signals reduce this likelihood. Figures 1 and 2 illustrate how $\Pr(x_i = h \mid \mathbf{y}_i)$ 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 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.

Endogenous Network Formation. Equation (13) applies only when tie formation is based only on social utility and excludes the credit score ($\alpha = 0$). We now consider the case wherein consumer utility includes credit score. We construct the utility of a borrower similar to §3.2. P_i is the firm's assessment of borrower i's probability of repayment, which we may take as the posterior probability that i is a high type. The lending rate for borrower i is again given by $r_i = (1-P_i)/P_i$.

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

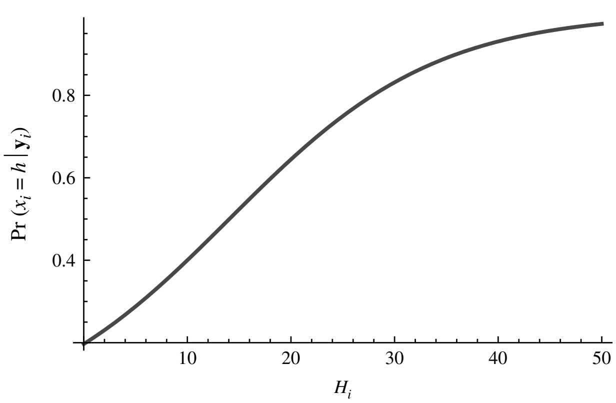
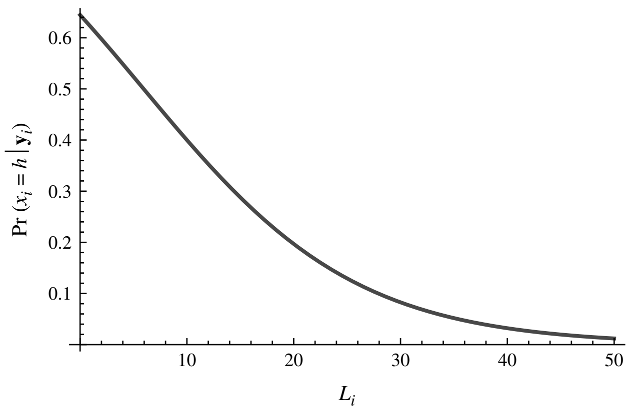


Figure 2 $\Pr(x_i = h \mid \mathbf{y}_i)$ **vs.** L_i $(\varepsilon = 0.4, p = 0.6, H_i = 10, y_i = -1)$



Because the lending rate enters the utility additively through $-\alpha_{x_i} \log(r_i)$, we have

$$U_i = \sum_{ij \in G} (m_{ij} - \mathbf{1}_{\{x_i \neq x_j\}}) + \alpha_{x_i} R_i, \qquad (14)$$

where $R_i \equiv \log(P_i/(1-P_i))$ is the log odds of repayment. A higher R_i implies a lower risk of extending credit to an individual. Furthermore, the parameter α_{x_i} calibrates the importance of improving access to financing. Note that this formulation allows low and high types to have two different levels of financial need. When $\alpha_h < \alpha_l$, high types' utility is less dependent on improving financing compared to the low types. When $\alpha_h = \alpha_l$, both types have identical financial needs. This exposition mirrors our continuous-type model, except that different types may weigh financial concerns (represented by R_i) differently when forming ties.

Let consumers choose tie formation rules before the meeting process. Intuitively, given the network-based score, consumers will be more selective towards low types and less selective towards high types. Because of the simplicity of the discrete-type model, the friendship rules we allow are general and flexible. More specifically, two high types will continue to form a tie with probability 1 after they meet. As to the friendship between low types, a low type i will set a threshold θ_i and accept another low type j, iff

$$m_{ij} - \theta_i > 0.$$

Because friendships are formed based on mutual consent, a friendship between a high and low type can only be formed when the high type accepts friendship. A high type i will accept a low type j, iff

$$m_{ij} - \beta_i > 0.$$

As in the continuous case, social credit scoring makes consumers wary of forming ties with low types. In the discrete case low and high types are allowed to differ in their need for financing, and low types face discrimination or social rejection from both low and high types. This result is interesting since discrimination is often thought to take place between groups or is believed to be exercised by one group on another. Interestingly, within-group discrimination arises endogenously with the use of the network-based scoring for the low types, in addition to the more common between-group discrimination. Within-group discrimination may make the surviving within-group ties more valuable, as we will see next in Lemma 4.

We define a symmetric profile characterized by two thresholds, i.e., (θ, β), where $\theta_i = \theta$ for all low type i and $\beta_i = \beta$ for all high type i. Let $(\theta, \beta, \theta_i)$ denote a symmetric profile except for possible deviation of a low type i. Let $\mathbb{E}(U_i \mid l, \theta, \beta, \theta_i)$ represent the expected utility before the meeting process for a low-type individual i

$$\mathbb{E}(U_i \mid l, \theta, \beta, \theta_i) = \mathbb{E}\left(\sum_{j:\, ij\in G} m_{ij} - \mathbf{1}_{\{x_i \neq x_j\}} \,\middle|\, l, \theta, \beta, \theta_i\right) + \alpha_{x_i = l}\, \mathbb{E}[R_i(\theta, \beta) \mid l, \theta, \beta, \theta_i],$$

where the lender's posterior assessment is $P_i(\theta, \beta) = \Pr(x_i = h \mid \mathbf{y}_i, \theta, \beta)$, consistent with the profile. Similarly, $\mathbb{E}(U_i \mid h, \theta, \beta, \beta_i)$ is the corresponding expected utility of a high type. Using this utility formulation, we first lay out the lender's prior about consumer types in Lemma 4.

Lemma 4. *Let (θ, β) be the symmetric criterion, $p_\theta \equiv \Pr(m_{ij} > \theta)$ the probability of two l-types forming a tie, and $p_\beta \equiv \Pr(m_{ij} > \beta)$ be the probability of a tie formation between h and l types. Then the posterior probability of i being high type is*

$$\Pr(x_i = h \mid \mathbf{y}_i, \theta, \beta) = \left[1 + \left(\frac{\varepsilon}{1-\varepsilon}\right)^{y_i} \left(\frac{\varepsilon p_\beta + (1-\varepsilon) p_\theta}{\varepsilon + (1-\varepsilon) p_\beta}\right)^{L_i} \cdot \left(\frac{\varepsilon p_\theta + (1-\varepsilon) p_\beta}{\varepsilon p_\beta + (1-\varepsilon)}\right)^{H_i} e^{(1/2)N(1-p_\theta)}\right]^{-1}, \quad (15)$$

where 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 with Lemma 3 in decomposing the contributions of high and low signals. When consumers form ties endogenously, the probability of a favorable risk assessment, $P_i(\theta, \beta)$ (or the corresponding $R_i(\theta, \beta)$), is increasing in the number of high signals (i.e., H_i) for any level of p_θ. By contrast, $R_i(\theta, \beta)$ increases in the number of friends with low signals (i.e., L_i) only if p_θ is sufficiently small,[12] and decreases in L_i otherwise. In other words, when l-types are very selective in forming ties among themselves (p_θ low), then in-group ties help to achieve a more favorable assessment from the firm, as low types have fewer ties than high types and a large friendship circle becomes a conspicuous signal, suggesting that one is more likely to be a h-type. That is the reason low-type signals can increase the high type perception, $P_i(\theta, \beta)$. Yet when low types are less selective towards other own types, the negative signal begins to dominate the positive impact from size of social circle and P_i decreases in L_i.

We now turn to the impact of how selective low types are in forming ties among themselves, characterized by selection rule θ. $R_i(\theta, \beta)$ is not always decreasing in L_i. In particular, we can define a value $\bar{\theta}(\beta)$ such that the *expected* effect of an additional low type *friend* on $R_i(\theta, \beta)$ is positive, iff $\theta > \bar{\theta}(\beta)$. Formally, $\bar{\theta}$ can be defined by

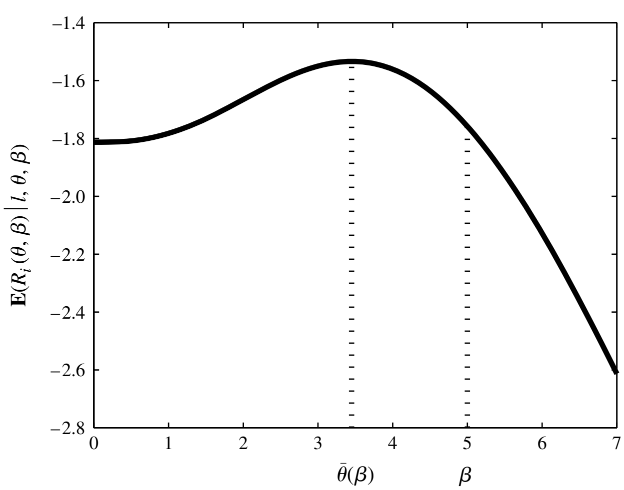
$$\left(\frac{\varepsilon p_\beta + (1-\varepsilon) p_{\bar{\theta}}}{\varepsilon + (1-\varepsilon) p_\beta}\right)^{1-\varepsilon} \left(\frac{\varepsilon p_{\bar{\theta}} + (1-\varepsilon) p_\beta}{\varepsilon p_\beta + (1-\varepsilon)}\right)^{\varepsilon} = 1.$$

It can be easily shown that $0 < \bar{\theta}(\beta) < \beta$. We show, in detail, how a consumer's odds of a favorable risk assessment vary with respect to the selectivity of l-types in Lemma 5.

Lemma 5. *The expected log odds for a low type under a common tie formation criterion (θ, β), $\mathbb{E}[R_i(\theta, \beta) \mid l, \theta, \beta]$, is strictly quasi-concave in θ and achieves its maximum at $\bar{\theta}(\beta)$. Furthermore, $0 < \bar{\theta}(\beta) < \beta$, i.e., the selectivity among low types which results in the most favorable risk assessment for a low type, is lower than the selectivity of high types towards the low types.*

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

Figure 3 **Expected Log Odds of Repayment vs. Selectivity** θ



Note. $\varepsilon = 0.2$, $f \sim \Gamma(3, 2)$, $\beta = 5$.

[12] Precisely, when $p_\theta < p_\beta + (\varepsilon/(1-\varepsilon))(1-p_\beta)$.

result in lower expected odds, whereas medium levels of selectivity among low types yield the most favorable risk assessment for them. 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 initially improves. Consumers benefit from disassociating themselves from l-types, thus improving the appearance of being an h-type. As selectivity increases further, however, a second and competing effect starts to dominate: Consumers' ego-networks begin to shrink extensively. Recall that the size of a borrower's network becomes a conspicuous signal of her type when consumers can form ties endogenously. Extreme selectivity leads to a smaller number of ties and so reveals the true low type of a borrower, thus reducing her chances of a favorable credit assessment.

Lemma 6. *The expected log odds for a low type is strictly decreasing in β for $\theta < \beta$. Higher levels of selectivity of high types towards low types reduce the chances of a favorable assessment for low types.*

The lemma states that, unlike the within-group exclusion that helps low types to some degree, between-type exclusion strictly reduces their chances of improving their financial outcomes. As high types exclude lower types from their networks, the latter's chances of a favorable assessment from the firm decreases, resulting in further hardship for this segment.

We will look for a symmetric equilibrium where no consumers have ex-ante incentive to deviate, and the company's posterior is consistent with their equilibrium behaviors. More precisely, (θ^*, β^*) is a symmetric equilibrium if, for all i, $\mathbb{E}(U_i \mid l, \theta^*, \beta^*, \theta_i)$ (or $\mathbb{E}(U_i \mid h, \theta^*, \beta^*, \beta_i)$, depending on i's type) is maximized by $\theta_i = \theta^*$ (or $\beta_i = \beta^*$). While ensuring that there will be no unilateral deviation, a Nash equilibrium in social networks does not necessarily rule out mutual improvement in the utility of consumers. For example, a very high acceptance criteria such as $\theta^* = \infty$ can always be part of an equilibrium because if no l-type accepts another l-type, an l-type would have no incentive for unilaterally deviating from this threshold. We remove "unintuitive" equilibria similar to the one described from consideration. Formally, we will not consider tie formation criteria (θ^*, β^*) an equilibrium if there is another profile (θ^{**}, β^*) such that (i) low types are better off, and (ii) given that high types choose β^* and every other low type chooses θ^{**}, a low type is willing to set her criterion θ^{**} as well. Similarly, we do not consider (θ^*, β^*) an equilibrium if there is a profile (θ^*, β^{**}) with unintuitive properties alike.

Note that from Lemma 5, for any equilibrium, $\theta^* < \beta^*$ should hold. In words, when both low and high types need financing, regardless of how dire the needs of the low types are (i.e., independent of the value of α_l), low types will face *within* and *between* group exclusion. More important, since high types are more successful in tie formation, they can afford to be selective in forming friendships. The low types, by contrast, cannot be picky choosers: If they set the friendship threshold too high, they find themselves on the downhill side of the expected log odds curve (Figure 3). They would achieve a higher score and higher social utility by being less selective. As a result, the within-group discrimination against low types is always lower than the between-group discrimination against them. This result is formally stated in Proposition 4.

Proposition 4. *Suppose $\alpha_h, \alpha_l > 0$. In any symmetric equilibrium (θ^*, β^*), we have $0 < \theta^* < \bar{\theta}(\beta^*)$ and $\beta^* > 1$, i.e., when both types gain utility from improving their credit scores, the within-group discrimination among low types is always lower than the between-group discrimination against them.*

In summary, two forces influence the network-based score in equilibrium to be more or less diagnostic for detecting a low type. Compared with the scenario before people react, higher exclusion among low types make social network-based scoring less powerful, by Lemma 5. Similarly, higher levels of exclusion on low types by high types increase the accuracy of the scores by Lemma 6.

4.2. Special Case: Lower Financing Needs for High Types

Until now, we have focused on an environment where the high types need financing. In reality, it is often the case that the need for financing (i.e., obtaining credit or a loan) is markedly more severe for low types. To address this possibility, we provide the outcomes from the special case when $\alpha_h = 0$. Note that, by continuity, this implies that similar results would hold if α_h is a very small positive number. Note also that when $\alpha_h = 0$, β is no longer material, and high types form a tie with low types only when $m_{ij} > 1$ (i.e., $\beta = 1$).

Proposition 5. *When $\alpha_h = 0$, there exists a unique equilibrium among low types such that $0 < \theta^* < \bar{\theta}(1)$. When only low types gain utility from improving their credit scores, there exists within-group discrimination among low types in equilibrium. This discrimination level is lower than the preference of high types to avoid forming relationships with low types due to mere homophily.*

Proposition 5 suggests that when high types put no or very little weight on access to financing, they may reject many social ties with low types due to homophily. In addition, due to financial concerns, l-type consumers are systematically excluded even from the networks of others similar to them. Put differently, existing financial inequality breeds within-group discrimination and social isolation among those of lower type and greater need.

4.3. Explicit Discrimination Against Low Types

We have shown how strategic discrimination against low types may emerge endogenously even in the presence of nonstrategic 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)

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

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 $p_1 \equiv \Pr(m_{ij} > 1) < 1$. Social utility is penalized whenever one becomes friends with an individual who is an l-type.

Parallel to Lemma 3, the following lemma gives the posterior before consumers strategically form their social ties to obtain better network-based scores. Note that mathematically the following lemma is a special case of Lemma 4 where $p_\theta = p_\beta = p_1$.

Lemma 7. *Let $p_1 \equiv \Pr(m_{ij} > 1)$ be the probability of formation of a tie with at least one low type. Then,*

$$\Pr(x_i = h \mid \mathbf{y}_i) = \left[1 + \left(\frac{\varepsilon}{1-\varepsilon}\right)^{y_i} \left(\frac{p_1}{\varepsilon + (1-\varepsilon)p_1}\right)^{L_i} \cdot \left(\frac{p_1}{\varepsilon p_1 + (1-\varepsilon)}\right)^{H_i} e^{(1/2)N(1-p_1)}\right]^{-1}, \quad (16)$$

where H_i is the number of friends with a high signal, and L_i is the number of friends with a low signal.

The lemma says that having a friend with a low signal actually improves one's score. When explicit discrimination is present, 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 $S\nu p$. Similar to the endogenous rise of discrimination, a larger social network is a conspicuous signal. A consumer with a larger network emits a stronger signal that she is a high type. Because in expectation low types have a smaller social circle, any tie becomes a signal of being high type.

What happens when both exogenous discrimination and endogenous tie formation are at work? Lemma 7 implies that consumers will be less selective towards low types in an attempt to obtain better scores. Similar to the thresholds we defined for the homophily case, we let low types choose a criterion $\theta_i \le 1$ towards their same type, and let high types choose $\beta_i \le 1$ towards low types. High types continue to form ties with probability 1 upon meeting. A tie between two different types forms only when the high type accepts the low type. It is not difficult to see that Lemmas 5 and 6 can be stated here without change. Furthermore, a result can be derived that corresponds to Proposition 4.

Proposition 6. *When low types are exogenously discriminated against and $\alpha_l, \alpha_h > 0$, in a symmetric equilibrium, $\bar\theta(\beta^*) < \theta^* < 1$ and $\beta^* < 1$.*

5. Effort to Become a High Type

Our results thus far have relied on the assumption that consumers are endowed with types that cannot be changed. In other words, we assumed that there is no social mobility. Although some type indicators (e.g., family, race, birthplace, country of origin) cannot be altered, other potential indicators, such as occupation or financial discipline, can be improved if low types exert effort (e.g., by investing in education). In this section, we extend our discussion to allow for this possibility. An array of factors may force l-type consumers 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 consumers. Let G_l denote the subnetwork among the low types. Furthermore, let H_i denote the number of h-type contacts of a low-type i, which are collectively represented with the vector $\mathbf{H}$ for all of the low types. Similarly, let L_i denote the number of l-type contacts of a low-type i. Each low-type consumer may exert effort $e_i \ge 0$ such that, with probability e_i, she will become a high type. Note that the effort therefore projects types of possible future contacts. We assume that given the network and parameters of our model, $e_i \le 1$ for all low-type i. High-type consumers exert zero effort and remain high types.

The utility that a low-type individual i derives from exerting effort e_i is composed of two parts

$$U_i(\mathbf{e}, G) = \sum_{j:\, ij \in G} \{m_{ij} - \mathbf{1}_{\{x_j=l\}}(1-e_j)\} + u_i, \quad (17)$$

where

$$u_i = ae_i - b\left[\frac{e_i}{2} - \phi\left(H_i + \sum_{j:\, ij\in G,\, x_j=l} e_j\right)\right]e_i. \quad (18)$$

The term in curly brackets in Equation (17) captures consumer i's expected social utility under the assumption of explicit discrimination (§4.3) and exertion of own and friends' effort. Given the effort of a friend e_j, there is $1-e_j$ probability that j will remain a low type, in which case i's utility from forming ties with j will be discounted by a unit normalized to 1.

The term u_i expresses the nonsocial benefits and costs of exerting effort. First, term ae_i captures the expected intrinsic benefits of becoming a high type. Second, the cost of effort is captured with the marginal cost $be_i/2$ that is increasing in effort. Third, under social network-based scoring, a (potential) high-type friend j has a positive effect on i's credit score and thus reduces i's financing burden. We formally express this network effect by allowing the marginal cost of effort for i to decrease in the number of high-type friends she has and in the efforts of her low-type friends to become high types, at rate $\phi > 0$. Alternatively, $b\phi(H_i + \sum_{ij\in G, x_j=l} e_j)e_i$ can be thought of as an interaction term, representing how the return to one's own effort (e_i) is expected to be amplified by the number of friends one expects will be considered high type. Some investors, for instance, may prefer friends who are also invited to participate in exclusive investment opportunities (Bursztyn et al. 2014). In a very different setting, one is likely to gain admission to an exclusive bar or dance club if oneself and the rest of one's party are attractively dressed.

It is important to make two notes here. First, the derivation of the functional form of u_i is a reduced-form approach to motivate complementarity between one's effort and the effort of her friends. It is possible to derive this form of complementarity based on the results provided in the earlier sections. (In the online appendix (available as supplemental material at http://dx.doi.org/10.1287/mksc.2015.0949), we offer a more detailed description of how Equation (18) can be derived through this route.) As demonstrated in §4, under network-based credit scoring with non-zero financing needs for both types, low types will face within-group and between-group discrimination. Under such pressure, l-type consumers would exert effort to increase their social and credit scoring utility from friendships. The benefits of exerting effort depend on the expected number of low- and high-type friends.

Second, it is possible to consider alternate specifications of social utility. For instance, we could also investigate an environment with pure homophily instead of discrimination, in which case Equation (17) would be replaced with

$$U_i(\mathbf{e}, G) = \sum_{j:ij\in G} \left\{ m_{ij} - e_i \mathbf{1}_{\{x_j=l\}}(1-e_j) - (1-e_i)\left[\mathbf{1}_{\{x_j=h\}} + \mathbf{1}_{\{x_j=l\}} e_j\right]\right\} + u_i. \quad (19)$$

In an environment with homophily, consumer i will become a high type with probability e_i, in which case there will be a disutility for a tie with consumer j who, after exerting effort e_j, remains a low type (which happens with probability $1-e_j$). With probability $1-e_i$, consumer i will remain a low type, in which case she will face a disutility from ties with high types (including low types who become high types after exerting effort e_j).

Next, given the utility form in (18), we will first derive the optimal effort level in a given network.

5.1. Effort in an Exogenous Network

We are interested in the Nash equilibrium under which consumers simultaneously choose their efforts when the network is exogenously given. Proposition 7 summarizes the optimal level of effort for a consumer conditional on her social network, following Ballester et al. (2006).

Proposition 7. *Let $\mathbf{A}_l$ be sociomatrix (i.e., the adjacency matrix) of G_l.*

(i) *Under a discriminating social utility, if the largest-magnitude eigenvalue of $\mathbf{A}_l$ is smaller than $|\phi|^{-1}$, then the equilibrium effort is*

$$\begin{aligned}\mathbf{e}^* &= (\mathbf{I} - \phi\mathbf{A}_l)^{-1}(ab^{-1} + \phi\mathbf{H})\\ &= (\mathbf{I} + \phi\mathbf{A}_l + \phi^2\mathbf{A}_l^2 + \cdots)(ab^{-1} + \phi\mathbf{H}). \quad (20)\end{aligned}$$

(ii) *Under a homophilic social utility, if the largest-magnitude eigenvalue of $\mathbf{A}_l$ is smaller than $|2b^{-1} + \phi|^{-1}$, the equilibrium effort is*

$$\begin{aligned}\mathbf{e}^* &= [\mathbf{I} - (2b^{-1} + \phi)\mathbf{A}_l]^{-1}[(a + \mathbf{H} - \mathbf{L})b^{-1} + \phi\mathbf{H}]\\ &= [\mathbf{I} + (2b^{-1} + \phi)\mathbf{A}_l + (2b^{-1} + \phi)^2\mathbf{A}_l^2 + \cdots]\\ &\quad \cdot [(a + \mathbf{H} - \mathbf{L})b^{-1} + \phi\mathbf{H}]. \quad (21)\end{aligned}$$

Proposition 7 states that the effort exerted by consumers to improve their score relies on several factors. A discriminatory environment and an environment with homophily differ in the role of the low types in inducing effort. In both environments, a consumer with a higher number of high-type friends is likely to exert more effort, as her overall cost of borrowing is lower. In an environment with discrimination, if two l-type consumers are connected to the same number of h-type friends, the one with a higher number of l-type friends is incentivized to exert more effort. This is perhaps surprising, as sufficiently high within-group connectivity can be a stronger motivator of effort. By contrast, in homophily, increasing proportions of low-type friends can reduce effort due to enhanced social utility when a consumer with low-type friends remains low type with low effort.

The expression for the equilibrium level of effort given in Equations (20) and (21) is a form of Bonacich centrality. The effort exerted by an agent to improve her credit score is proportional to her Bonacich centrality measure, which is the "summed connections to others, weighted by their centralities" (Bonacich 1987, p. 1172). With a discriminating social utility, a consumer who is at the center of a social network is likely to be exposed to higher positive network effects, and therefore may exert greater effort. As a result, consumers who are more central in the network are more prone to social mobility when there are complementarities. In an environment with pure homophily, there will be two conflicting forces determining centrality and social mobility relationships. First, being central in a network of high types and low types who exert effort can increase a consumer's chances of social mobility. Second, if a low-type consumer is central among other low types who exert little effort, she will reduce her effort to "fit" and be similar to her network to enhance her social utility. Therefore, in tie formation based on homophily, it is possible for central low types to exert low effort leading to permanent low class membership and financial hardship.

5.2. Effort with Endogenous Network Formation Among Low Types Under Discriminating Utility

As we have specified in (17)–(19), 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 her friends' network formation. Moreover, in the last section, we saw that i's effort depends on her position in the network. This mutual dependence between the network position and effort suggests the possibility of multiple stable situations. With discriminating social utility, 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 under the discrimination environment. In the first stage, consumers choose friends, and friendships are formed bilaterally. In the second stage, consumers exert effort. 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

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

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 $ij \notin G$, $U_i(G) \geq U_i(G + ij)$ 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 consumers and explicit discrimination, and assume that $a = 1$, $b = 5$, and $m_{ij} = \frac{1}{2}$ for all i, j. Let $\phi b = \frac{1}{5}$. It can be easily verified that the empty network and the complete network are pairwise stable. For the empty network, each person exerts effort $\frac{1}{5}$ and obtains utility of $\frac{1}{10}$. For the complete network, each person exerts effort $\frac{1}{2}$ and has utility $\frac{5}{8}$.

The example demonstrates that the empty network is pairwise stable because everyone exerts very low effort. A single link between a pair will not generate a sufficiently large change. The disutility of friendship with a low type (which is normalized to 1) prevents any pair from becoming friends. Moreover, a complete network is pairwise stable because everyone exerts reasonable effort. The effort reduces the disutility of friendship between low types; the friendship utility between any pair is exactly zero. Breaking any one link increases the costs of effort for the pair; thus, they will decrease their efforts. This leads to higher costs for their friends, and eventually everyone's effort will decrease. As a result, everyone receives less utility from 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, social pressure is expected to be less effective and social mobility may remain limited. By 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 that the performance is conditional on the network structure of society.

6. Extensions

6.1. Uncertainty About Friends' Types

In our main model, the underlying assumption was that upon meeting, consumers learn about each others' types with certainty. In reality, types may be observed with some noise. Consider the case wherein consumers meet others but observe their types imperfectly. Let consumer i observe a signal of x_j upon meeting with j, which is correct with probability $1 - \tau$ with $0 < \tau < \frac{1}{2}$. This implies that the added utility from homophily relies on how the uncertainty about the other's type is resolved: Expected social utility is $m_{ij} - \tau$ if the signal is the same as one's own type, and $m_{ij} - 1 + \tau$ otherwise. Respectively, probabilities $p_\tau \equiv P(m_{ij} > \tau)$ and $p_{1-\tau} \equiv P(m_{ij} > 1 - \tau)$ define how likely two consumers are to become friends upon meeting.

Compared with the benchmark model, the added uncertainty implies that ties will be less informative for the firm to predict a consumer's type. To see this, first note that under this formulation, the probability that two consumers of the same type will form a tie upon meeting is

$$q_s \equiv (1 - \tau)^2 p_\tau + (1 - (1 - \tau)^2) p_{1-\tau}, \tag{22}$$

and that two consumers of opposite types will form a tie is

$$q_d \equiv \tau^2 p_\tau + (1 - \tau^2) p_{1-\tau}. \tag{23}$$

Using these probabilities, we can formulate how the firm will assess a borrower's type to be high as given in Lemma 8.[13]

[13] The derivation of Lemma 8 follows the derivation of Lemma 4.

Lemma 8. *When consumers learn about each others' types with uncertainty,*

$$\Pr(x_i = h \mid \mathbf{y}_i) = \left[1 + \left(\frac{\varepsilon}{1-\varepsilon}\right)^{y_i} \left(\frac{\varepsilon q_d + (1-\varepsilon) q_s}{\varepsilon q_s + (1-\varepsilon) q_d}\right)^{L_i} \cdot \left(\frac{\varepsilon q_s + (1-\varepsilon) q_d}{\varepsilon q_d + (1-\varepsilon) q_s}\right)^{H_i}\right]^{-1},$$

where H_i is the number of friends with high signal, and L_i is the number of friends with low signal.

We are interested in how the presence of noise in detecting each other's true types in social relationships may influence the firm's ability to rely on social credit scores. We compare Lemma 8 with Lemma 3. Since $p_1 < q_d < q_s < 1$, $1 < (\varepsilon q_d + (1-\varepsilon) q_s)/(\varepsilon q_s + (1-\varepsilon) q_d) < (\varepsilon p_1 + (1-\varepsilon))/(\varepsilon + (1-\varepsilon) p_1)$ and $(\varepsilon + (1-\varepsilon) p_1)/(\varepsilon p_1 + (1-\varepsilon)) < (\varepsilon q_s + (1-\varepsilon) q_d)/(\varepsilon q_d + (1-\varepsilon) q_s) < 1$. In words, signals from contacts carry less weight in forming beliefs about a consumer's type when types cannot be perfectly observed in friendship.

There are two observations related to this finding. First, the level of information sharing between consumers can change the appropriateness of a social network for credit scoring. For example, if an online network allows consumers to frequently communicate and exchange in-depth information, this may positively influence the efficiency of credit assessment by reducing the uncertainty about friends' types. Second, the ability of peers to observe each other's types may correlate with the characteristics of the network, including tie strength. For example, parameter τ could reflect the strength of ties correlating with the ability to convey complex or subtle information (Van den Bulte and Wuyts 2007, pp. 71–72) and hence with one's ability to observe a friend's type. Next, in §6.2, we discuss this in detail.

6.2. Friendship Formation and Strength of Ties

In §6.1, we maintained the assumption that all relationships carry equal information and pointed out that the informativeness of a link may relate to tie strength. We will adjust the earlier model slightly to extend the earlier discussion.

Specifically, we assume that consumers can form weak and strong ties, and that they learn about others' type with certainty only if they have strong ties with them. After meeting, a match value $m_{ij} > 0$ and the tie type are randomly determined. If the tie is strong, consumers obtain the utility $m_{ij} - \mathbf{1}_{\{x_i \neq x_j\}}$ by forming a friendship. If the tie is weak, types remain unknown, and the social utility of forming a tie is $m_{ij} - \kappa$. Parameter κ captures the disutility from forming a weak tie.

Because weak ties do not carry information about the type or the type difference between the ego and the friend, a firm cannot use them to update its posterior belief about a consumer's type. Only the strong ties will reveal information about a contact's type and become eligible for the firm to use to determine the social score.

The general implication is straightforward. Because strong ties are more homophilous than weak ties and since they provide a greater ability to learn about one's contacts, the accuracy of social scoring increases with the relative prevalence of strong versus weak ties.

6.3. Effort to Enhance Probability of Meeting High Types

In §5, the model was built such that the low-type consumers exerted effort to climb social ladders by improving their type. Under some circumstances, consumers cannot change their type but can exert effort to increase the probability of meeting high types. Networking is an example of such directed effort. In this section we explore this possibility, which also allows us to endogenize the probability of meeting between two consumers.

We use the settings of the discrete-type model in §4 and allow individual i to choose an effort level e_i. Conditional on the effort exerted, the individual is likely to meet another person randomly with probability $(M/S)e_i$, where M is a constant that calibrates the chance of meeting another person proportional to the effort exerted in a society of size S. A meeting between i and j happens when either of the two "runs into" the other. Suppose a common effort e is exerted by everyone but i. Then the expected number of meetings for i becomes

$$S\left[1 - \left(1 - e_i \frac{M}{S}\right)\left(1 - e\frac{M}{S}\right)\right] = \left(e_i + e - \frac{e_i e M}{S}\right) M.$$

When $S \to +\infty$, the expected number of meetings increases to $(e_i + e)M$.

First consider the scenario of exogenous tie formation where an individual's utility depends only on the social utility from friendships. Recall that, upon meeting with j, i always forms a tie if j is of the same type, and forms a tie, iff, $m_{ij} > 1$ if j is of the different type. Let f be the density of the matching value distribution. The expected social utility for i, given that a symmetric effort e is used except for possible deviation of i to e_i, can be derived from

$$\mathbb{E}\left(\sum_{j:\, ij \in G} m_{ij} - \mathbf{1}_{\{x_i \neq x_j\}} \,\middle|\, x_i, e, e_i\right) = (e_i + e)\frac{M}{2}\left(\int_0^\infty t f(t)\, dt + \int_1^\infty (t-1) f(t)\, dt\right).$$

Let Λ denote the term in the last parentheses. Let $\frac{1}{2}e_i^2$ be the cost of effort. The equilibrium effort is then given by

$$e^* = \frac{M\Lambda}{2}. \tag{24}$$

Under this common effort level, the firm's posterior on type is again given by (13) in Lemma 3.

Next, we set this equilibrium effort level as the baseline, and compare it to that when social relationships affect financial benefits. Because the credit score introduces asymmetric desirability of low-type and high-type friends, the effort levels exerted by low types and high types will, in general, be different. In principle, the firm's posterior needs to incorporate the difference in efforts. Here for simplicity, we focus on how effort level will differentiate between types but omit how it would affect a firm's posterior assessment.

Formally, we let the credit score enter utility additively through $\alpha_{x_i} R_i$ with P_i given simply by (13). We characterize a symmetric equilibrium, by which we mean the effort pair (e_l^*, e_h^*) where every low type chooses e_l^* and every high type chooses e_h^* such that no consumer has an incentive to deviate.

The following proposition summarizes how the consumer motivation to meet others changes compared with the effort they would exert simply to maximize their utility from friendships.

Proposition 8.
(i) *For* $\alpha_l = \alpha_h > 0$, $e_h^* > e^* > e_l^*$. *When both types have identical needs for financing, high types exert more effort than low types.*
(ii) *For* α_l *sufficiently larger than* α_h, $e_l^* > e_h^* \geq e^*$. *When low types have higher needs for financing, they exert more effort than high types.*

The proposition suggests that when l and h type consumers have identical needs for financing, high types exert higher levels of effort to increase their probability of meeting others compared to low types and compared with the effort exerted when consumers only want to maximize social utility. This is because high types have a higher marginal return on effort than the low types (i.e., are more likely to form new ties as a result of effort). As a result, independent of their financial needs, high types always exert more effort than they would when they earn utility from improving their access to credit in addition to the gains in social utility. Low types, by contrast, have lower returns, but when high types make an effort to meet others, they also benefit from it. With some probability, a meeting will take place between a low and a high type and a friendship will be formed if m_{ij} is sufficiently high.

If, on the other hand, the low types' utility from improving their credit scores is very high (α_l very high), this pattern result could reverse. Low types would feel an immense pressure to increase the probability of becoming friends with high types, resulting in a higher level of effort exerted by low types compared to that of high types.

Finally, an environment where the low types exert sufficiently high levels of effort could help to create a bridge between the two types, possibly reducing the social separation. Therefore, low types who have more to gain from improving their financing ($\alpha_l > \alpha_h$) could exert sufficiently great networking effort to connect the two types.

7. Conclusion

7.1. Main Insights

Increasing access to financing is important in many countries where institutions and contract enforcement are weak (e.g., Feigenberg et al. 2013, 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 those 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 applicants' 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 benefits to collecting information from a consumer's network rather than only individualized data. Simply put, when consumers have an above average chance of interacting with others of similar creditworthiness, then network ties provide additional reliable signals about their true creditworthiness. Hence, social scoring can reduce lenders' misgivings about engaging applicants 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 consumers, particularly those in more dire need of improving their credit score, may 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 subnetworks where consumers only connect 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 accuracy is not straightforward. Although there will be fewer ties conveying information about one's contacts useful in updating lenders' prior beliefs, 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, these are: (i) consumers place sufficiently low importance on the posterior mean of the firm's beliefs about their type (low α), (ii) high precision on individual credit scores (high c), and (iii) relatively dense network (high N).

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. Not surprisingly, connections to those with high-type signals have an overall positive impact. More interesting is that the impact of connections to low-type signal consumers can be positive or negative, depending on the tie formation rules used in society. As demonstrated in Figure 3, consumers with poor financial health would prefer others like them not to be too selective but also not to be too liberal in their willingness to associate with people with poor financial health. Intuitively, as the selectivity of same-type consumers increases, the impact of negative signals received from some of the low-type friends weakens. As the selectivity increases even further, low types' social circles will shrink such that it will be harder for them to emit a high-type signal, since size of social circles is a conspicuous signal of type. As a result, disadvantaged consumers would prefer some intermediate level of ostracism and social isolation.

In our extensions, we discuss two scenarios that may reduce the reliability of social scores. First, if consumers cannot observe their social contacts' types perfectly upon meeting, the added noise will imply that homophily will play a lesser role in the formation of social networks. As a result, firms' ability to detect a borrower's type by looking at her friends will be limited. Similarly, if the network consists mainly of weak rather than strong ties, this will also reduce social scores' diagnosticity, since strength correlates with how well consumers know each other. In both of these scenarios, contacts' signals carry lower value to the firm in assessing the risk of a borrower.

We also consider the possibility of exerting effort in two different ways. First, we move away from the static type model and allow consumers to improve their type. We find that when there is discrimination against low types, low- and high-type contacts play a role in motivating effort, but high types, in general, have a stronger effect. In an environment with only homophily, these results hold as well, unless a consumer is highly embedded in a network with many low-type friends who exert low effort. Such consumers are not motivated to exert effort towards improving themselves and are more likely to remain a low type. Second, when types are sticky and cannot be altered, we allow consumers to exert effort to improve their chances of meeting other people. This second model shows that consumers' networking effort will depend on their need for financing. When high and low types have comparable needs for financing, high types have higher returns on their effort of creating new ties and thus exert more effort to meet others. Because the types are revealed only after meeting, low types' likelihood of meeting a higher type increases when high types exert effort, too. Therefore they choose to free ride on others' efforts. This outcome reverses when low types are in dire need of financing, and they become the primary driver of meetings in society.

One possible outcome of social scoring, which is not addressed in this research, is that consumers strategically manipulate the perception of their type by trading friendships for financial access. In particular, realizing their higher financial status, high types may want to offer their friendships in exchange for monetary rewards. To model an environment wherein friendships are traded, we may need to consider several additional layers of complexity. First, rationally, traded friendships would need to be formed such that the credit scoring firm should be unable to distinguish a fake relationship from a true friendship. Otherwise, low types would have no incentive to pay for a high type's friendship. Second, high types must be financially motivated and the benefit from forming a friendship with a low type must exceed the losses from less favorable risk assessment. Third, trading friendships must be rare enough that a credit-scoring firm still benefits and desires to use data from the social networks. Altogether, modeling an environment of this sort would require a fairly complicated model, which goes beyond the purposes of the current study. Despite the complication, our expectation for the findings would be fairly simple: In line with the extensions discussed in §§6.1 and 6.2, if social ties have lower informative value and homophily is diluted, social credit scores will be less diagnostic in detecting one's true creditworthiness.

7.2. Implications for Public Policy

The link between credit scores and income is hard to ignore.[14] It is reported that most U.S. consumers with an income under $60K have a poor credit score.[15] Moreover, a significant portion of the individualized credit score calculation relies on a consumer's existing debt level. Those with higher amounts of debt, all else

[14] This is so even though 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.

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

equal, are expected to have lower credit scores. With network-based assessment, it is possible for immigrants, underbanked consumers, recent college graduates, and others who do not have a credit history but who are creditworthy to signal this to lenders with higher accuracy. The benefits introduced through network-based systems may help 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 Equal Credit Opportunity Act (ECOA) prohibits lenders from discrimination based on sex, race, color, religion, national origin or age. To the extent that some of these characteristics correlate with creditworthiness and that homophily along those dimensions correlate with homophily along levels of creditworthiness, a side-effect of social credit scoring could be discrimination in access to credit along characteristics prohibited by the ECOA (National Consumer Law Center 2014, pp. 27–29). Aside from strict legality, there is also a concern that social scoring opens an additional back door to discrimination along dimensions that many may find objectionable (Dixon and Gelman 2014, Pasquale 2015).

Matters are even more complex as our results also show that social scoring may lead consumers with low creditworthiness to prefer being discriminated against (in tie formation at least) to some moderate extent. Thus moderate levels of discrimination and social ostracism by fellow consumers may actually help rather than harm disadvantaged consumers. Also, one hitherto ignored societal benefit of social scoring is that it can motivate rather than demotivate financially disadvantaged citizens to exert greater effort to improve their creditworthiness. 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.

Finally, our findings here are of interest to policy makers keen on understanding the mutual interaction between social status and network structure. As noted at the outset, our mathematical analysis of credit scoring applies broadly to social status. Some people command less respect than others. Differences in status are rarely based solely on differences in true but hard-to-observe ability or character. People often use the company that others keep as a signal when assessing the respect they deserve. Our analysis of the benefits and challenges of social credit scoring, including improved diagnosticity paired with the risk of unwitting discrimination and the seeming paradox of optimal ostracism, extends to situations wherein citizens, employees or customers are valued and accorded status based on the company they keep.

7.3. Implications for Management

To managers in the financial industry, our analysis suggests that lenders can expect to reduce their risk in the short run by incorporating network-based measures. This dovetails with new governmental policies on risk. For example, as part of the regulations by the Basel Committee on Banking and Supervision, European banks have been encouraged to reduce the level of risk they undertake (Sousa et al. 2013). Regulations in the banking industry encourage U.S. financial institutions to better manage risk as well. These regulations have come at a time when big data analytics are enabling financial institutions to access larger and richer data sets. 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 toward using social data may prove to be useful in the post-crisis environment.

Our study also offers some insight to managers outside the financial industry who use social scoring for targeting customers when launching new products, targeting ads or designing referral programs. (i) The effectiveness of social scoring need not decrease when customers purposely adapt their networks to improve their score and their access to the benefits it entails. (ii) Marketers do not need information on the complete network. Data on the focal consumer's immediate contacts already provide an improvement in scoring accuracy. (iii) Social scoring is likely to be most diagnostic in societies and communities (online or not) where consumers maintain many strong, rather than weak, ties. (iv) Smart marketers will go beyond generic ties and seek to leverage specific ties that correlate highly with the traits they seek in their target customers. A car manufacturer such as Audi, for instance, will benefit from focusing on Twitter connections pertaining to cars (personal communication). (v) The benefits of social scoring to the marketer are greater when the benefits of having a high score matter little to customers or at least has little impact on those with whom they chose to form ties. More generally, the benefits of social scoring are greater when they involve networks of ties that not only exhibit great homophily but also are built and maintained for intrinsic rather than extrinsic reasons. Examples of the former used in social scoring include telephone call data and kinship data (Benoit and Van den Poel 2012, Hill et al. 2006). Examples of the latter are many ties in general-purpose online social networking platforms, where linking is very easy and often occurs between casual contacts. (vi) Customers with a high number of connections (degree centrality) in an undirected network such as Facebook or LinkedIn are not necessarily the most attractive. This is not only because centrality in such networks cannot distinguish between opinion seekers and opinion leaders (in-degree versus out-degree centrality) but also because (as our analysis shows) the most active networkers may be high-type or low-type customers, depending on whether low types value the benefits of a high consumer score more than high types. (vii) Marketers should be concerned that social customer scoring may create the impression of unfair discrimination. This is not only a legal and an ethical issue but also a commercial one. For instance, in January 2015, users of WeChat, the Chinese chat app, protested against discrimination after they were not targeted to see an ad for BMW, the luxury car maker. Some believed that the targeting algorithm involved social scoring based on those to whom the potential targets were connected (Clover 2015). Because social scoring uses inputs beyond one's traits and history, marketers must balance improved diagnostics against actual and perceived fairness.

The insights in this paper also provide some guidance on data collection and system design. Several firms already create credit scores using social network data. One important question they face is whether the number of friends is a useful signal for measuring creditworthiness. Our study demonstrates that even when it is not a signal directly linked to one's type, the practice of network scoring would endogenously make the number of friends a useful signal. Thus social credit scoring may shape credit assessment in its own image, i.e., help to construct the reality it is meant to describe, just as modern option theory did for valuing financial derivatives (MacKenzie 2006, MacKenzie and Millo 2003). This makes the implications of our analysis all the more important.

Supplemental Material

Supplemental material to this paper is available at http://dx.doi.org/10.1287/mksc.2015.0949.

Acknowledgments

The authors thank the senior editor, associate editor, and two anonymous referees for their comments and suggestions. The authors also benefited from comments by Lisa George, Yogesh Joshi, Upender Subramanian, Yi Zhu, and participants at the 2014 Columbia-NYU-Wharton-Yale Four School Marketing Conference, the 2014 Marketing Science Conference, the 12th ZEW Conference on Information and Communication Technologies at the University of Mannheim, the 2015 INSEAD Marketing Camp, and the 2015 CRES Strategy Conference at Washington University in St. Louis. The authors 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.

Appendix. Proofs

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

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

Thus

$$\begin{aligned}&\Pr(\mathbf{x}|g,\mathbf{y})\\ &\quad\propto \prod_{i\in\mathbf{y}} e^{-q{x_i}^2/2}\times\prod_{i\in\mathbf{y}} e^{-c(y_i-x_i)^2/2}\times\prod_{ij\in g^1}\nu e^{-(x_i-x_j)^2/2}\\ &\quad\times\prod_{ij\in g^0:\, i,\, j\in\mathbf{y}}[1-\nu e^{-(x_i-x_j)^2/2}]\times\prod_{ij\in g^0:\, j\notin\mathbf{y}}\left(1-\frac{\mathbb{E}(n_i|x_i)}{S}\right).\end{aligned} \tag{25}$$

In the expression above, $(1-\mathbb{E}(n_i\mid x_i)/S)$ 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 the term $\prod_{ij\in g^0:\ j\notin\mathbf{y}}(1-\mathbb{E}(n_i\mid x_i)/S)$. If $\{ij\in g^0\colon j\notin\mathbf{y}\}$ is not empty, then by our assumption on the information structure, it multiplies across everyone in the rest of the society. So its value under the limits of S, ν, and q is

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

which is not a function of $\mathbf{x}$ thus does not contribute to the conditional density. Note that the rest of the terms in the right-hand side of (25) multiply across finite items. It is easy to see that as $\nu\to 0$ and $q\to 0$,

$$\Pr(x\mid g, y)\propto\prod_{i\in\mathbf{y}} e^{-c(y_i-x_i)^2/2}\times\prod_{ij\in g^1} e^{-(x_i-x_j)^2/2}. \tag{26}$$

This implies that $\Pr(\mathbf{x}\mid g,\mathbf{y})$ 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 match 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 $(\boldsymbol{\Sigma}^{-1})_{i1}\mu_1+(\boldsymbol{\Sigma}^{-1})_{i2}\mu_2+\cdots$. The corresponding coefficients in the right-hand side of (26) 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 a special case of Proposition 1, where i is fixed and $\mathbf{y}=\{j\mid ij\in G\}$, $g^1=\{ij\mid ij\in G\}$ and $g^0=\{ij\mid ij\notin G,\ j\neq i\}$. □

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 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 up 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 write

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

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

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

Now consider a node j whose distance from i in the subnetwork 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 $(\mathbf{I} + \mathbf{B} + \mathbf{B}^2 + \cdots)_{ij}$ is

$$\sum_{k=r(i,j)}^{+\infty} \delta^k = \delta^{r(i,j)}/(1-\delta). \quad \square$$

Proof of Lemma 1. Derivation of the lemma follows similarly to the proof of Corollary 1.

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

$$\begin{aligned}\mathbb{E}(n_i \mid x_i, \lambda) &= S\int_{-\infty}^{+\infty} \nu e^{-\lambda(t-x_i)^2/2} \sqrt{\frac{q}{2\pi}} e^{-qt^2/2}\, dt \\ &= S\nu \sqrt{\frac{q}{q+\lambda}} e^{-((\lambda q)/(\lambda+q)) {x_i}^2/2}.\end{aligned}$$

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

Proof of Proposition 3. For notational simplicity, the expectation sign $\mathbb{E}(\cdot)$ throughout this proof refers to the conditional expectation $\mathbb{E}(\cdot \mid x_i, \lambda, \lambda_i)$, which is computed conditional on the type x_i and a symmetric rule λ except for possible deviation of i to λ_i. Similarly, the notation $\Pr(\cdot)$ also refers to the probability with the same conditionals.

First we calculate the expected social utility, $\mathbb{E}\sum_{ij\in G}(m_{ij} - |x_j - x_i|)$, which we denote more compactly as $\mathbb{E}\, u_i$. For any j we have

$$\begin{aligned}\Pr(x_j, ij \in G) &= \Pr(x_j)\Pr(ij \in G \mid x_j) \\ &= \sqrt{\frac{q}{2\pi}} e^{-qx_j^2/2} \cdot \begin{cases} \nu e^{-\lambda_i(x_i-x_j)^2/2} & \text{if } x_j \leq x_i, \\ \nu e^{-\lambda(x_i-x_j)^2/2} & \text{if } x_j > x_i. \end{cases}\end{aligned} \tag{27}$$

Equation (27) enables us to calculate the probability of being friends with j

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

and in particular, its limiting value

$$S\Pr(ij \in G) \to \frac{1}{2}N\left(\frac{1}{\sqrt{\lambda_i}} + \frac{1}{\sqrt{\lambda}}\right). \tag{28}$$

Similarly, (27) also enables us to calculate the conditional type difference and its limiting value

$$\begin{aligned}\mathbb{E}(-|x_j - x_i| \mid ij \in G) &= \int_{-\infty}^{+\infty} -|x_j - x_i| \Pr(x_j \mid ij \in G)\, dx_j \\ &\to \sqrt{2/\pi}\left(\frac{1}{\lambda_i} + \frac{1}{\lambda}\right)\Big/\left(\frac{1}{\sqrt{\lambda_i}} + \frac{1}{\sqrt{\lambda}}\right).\end{aligned} \tag{29}$$

Next we turn to the matching value. We have

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

So,

$$S\Pr(m_{ij}, ij \in G) \to \frac{N}{\sqrt{2\pi}} m_{ij}^2 e^{-m_{ij}^2/2}\left(\frac{1}{\sqrt{\lambda}} + \frac{1}{\sqrt{\lambda_i}}\right),$$

which, with (28), implies that

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

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

$$\mathbb{E}(m_{ij} \mid ij \in G) \to 2\sqrt{\frac{2}{\pi}}. \tag{30}$$

Expected social utility can be computed by summing over i's expected social utility from each j in the society

$$\begin{aligned}\mathbb{E}\, u_i &= \sum_{j \neq i} \Pr(ij \in G)\big[\mathbb{E}(-|x_j - x_i| \mid ij \in G) + \mathbb{E}(m_{ij} \mid ij \in G)\big] \\ &= S\Pr(ij \in G)\big[\mathbb{E}(-|x_j - x_i| \mid ij \in G) + \mathbb{E}(m_{ij} \mid ij \in G)\big].\end{aligned}$$

Equipped with (28)–(30), we find its limiting value,

$$\mathbb{E}\, u_i \to \frac{N}{\sqrt{2\pi}}\left[2\left(\frac{1}{\sqrt{\lambda_i}} + \frac{1}{\sqrt{\lambda}}\right) - \left(\frac{1}{\lambda_i} + \frac{1}{\lambda}\right)\right]. \tag{31}$$

A nice intuitive result from this is that the social utility is maximized at $\lambda_i = 1$. Any deviation from that distorts the friendship formation and is suboptimal in terms of social utility.

Next we look at the expected utility from the network-based score. The bias from using network-based scoring is

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

The first equality comes from the fact that y_i (and y_j) are unbiased signals of x_i (and x_j). The second equality uses the iterated law of expectation. The last equality uses the fact that $\mathbb{E}(x_j - x_i \mid ij \in G)$ is not a function of n_i.

Using (27), we calculate, in a way similar to (29),

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

So under the limits, we have the equality

$$\mathbb{E}\mu_i(\lambda) - x_i = \underbrace{\mathbb{E}\left(\frac{\lambda n_i}{c + \lambda + \lambda n_i}\right)}_{\varphi} \cdot \underbrace{\sqrt{\frac{2}{\pi}}\left(\frac{1}{\lambda} - \frac{1}{\lambda_i}\right)\Big/\left(\frac{1}{\sqrt{\lambda}} + \frac{1}{\sqrt{\lambda_i}}\right)}_{\xi}. \tag{32}$$

Because it is somewhat difficult to find an explicit expression for φ, even under limits, we address 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.

First, to find the equilibrium, we look at the best response correspondence for i, that is, the value of λ_i that maximizes $\mathbb{E}\,U_i$ for any λ. We use the derivative of $\mathbb{E}\,U_i$

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

By (32),

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

Note that (i) ξ has the same sign as $\lambda_i-\lambda$, (ii) $\partial\xi/\partial\lambda_i>0$, (iii) $0<\varphi<1$, and (iv) $\partial\varphi/\partial\lambda_i<0$. The first three points are clear. The last point can be seen by noting that n_i is binomially distributed, and under the limits, Poisson distributed with the mean given in (28).

Using (i)–(iv), we see two useful properties for the second component of F

$$\frac{\partial(\mathbb{E}\,\mu_i(\lambda)-x_i)}{\partial\lambda_i} < \frac{\partial\xi}{\partial\lambda_i}, \quad \text{for } \lambda_i\ge\lambda, \tag{33}$$

and

$$\frac{\partial(\mathbb{E}\,\mu_i(\lambda)-x_i)}{\partial\lambda_i} > 0, \quad \text{at } \lambda_i=1. \tag{34}$$

Next we translate these two properties into two properties of F. First, (34) shows that

$$F(1,\lambda)>0 \tag{35}$$

because $\partial\,\mathbb{E}\,u_i/\partial\lambda_i=0$ at $\lambda_i=1$, by (31).

To obtain the second property, we look at a simpler case of F where φ is ignored in the derivative

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

It can be easily verified that as long as $\alpha<N$, there is an invariant solution to $\tilde F(\cdot,\lambda)=0$

$$\lambda^o \equiv \left(1-\frac{\alpha}{N}\right)^{-2}$$

and $\tilde F(\lambda_i,\lambda)\le 0$ for $\lambda_i\ge\lambda^o$. Together with (33), this shows that

$$F(\lambda_i,\lambda) < \tilde F(\lambda_i,\lambda)\le 0, \quad \text{for any } \lambda_i\ge\max(\lambda,\lambda^o). \tag{36}$$

These two properties about F are sufficient to derive the proposition. Define $\Xi_i(\lambda) := \arg\max_{\lambda_i\ge1}\mathbb{E}\,U_i$ as the best response correspondence. Using Berge's Theorem we show that it is upper semi-continuous. Furthermore, (35) and (36) imply

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

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

Proof of Corollary 2. For the precision,

$$\mathbb{E}[\rho_i(\lambda)\mid\lambda] = c+\frac{c\lambda}{c+\lambda}\mathbb{E}_\lambda n_i = c+\frac{c\sqrt{\lambda}}{c+\lambda}N.$$

The first equality uses (8) for the expression of $\rho_i(\lambda)$. The second equality comes from (9).

If $c\le 1$, then the precision is decreasing in λ after 1. So it is smaller at λ^* than at 1. If $c\ge\sqrt{N/(N-\alpha)}>1$, then the precision is no larger at 1 than at $(1-\alpha/N)^{-1}$, which is the upper bound of λ^*. Noting that the precision is also quasiconcave in λ, we see that it is smaller at 1 than at λ^*. □

Proof of Lemma 3. Using the definition of conditional probability, we have

$$\Pr(x_i=h,\mathbf{y}_i)=\Pr(\mathbf{y}_i\mid x_i=h)\Pr(x_i=h).$$

The prior $\Pr(x_i=h)=1/2$ by our assumption. The likelihood $\Pr(\mathbf{y}_i\mid x_i=h)$ has three parts: (i) the probability that i is friends with those whose signals are collected in $\mathbf{y}_i$ and that these friends have the signals as collected in $\mathbf{y}_i$, (ii) the probability that i is not friends with anyone outside $\mathbf{y}_i$, and (iii) the probability that i's own signal is as that collected in $\mathbf{y}_i$. Formally,

$$\Pr(\mathbf{y}_i|x_i=h) = \sum_{\mathbf{x}_i}\Bigl[\prod_{ij\in G}\nu(\mathbf{1}_{\{x_j=h\}}+p_1\mathbf{1}_{\{x_j=l\}})\prod_{ij\in G}\Pr(y_j|x_j)\Bigr] \times\prod_{ij\notin G,\,j\ne i}\bigl(1-\tfrac12\nu(1+p_1)\bigr)\cdot\Pr(y_i|x_i=h)\cdot\tfrac12,$$

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

Another way to express the probability $\Pr(x_i=h,\mathbf{y}_i)$ is

$$\Pr(x_i=h,\mathbf{y}_i) = \prod_{ij\in G}\nu\bigl[\Pr(y_j|x_j=h)+p_1\Pr(y_j|x_j=l)\bigr] \times\prod_{ij\notin G,\,j\ne i}\bigl(1-\tfrac12\nu(1+p_1)\bigr)\cdot\Pr(y_i|x_i=h)\cdot\tfrac12. \tag{37}$$

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

$$\Pr(x_i=l,\mathbf{y}_i) = \prod_{ij\in G}\nu\bigl[p_1\Pr(y_j|x_j=h)+\Pr(y_j|x_j=l)\bigr] \times\prod_{ij\notin G,\,j\ne i}\bigl(1-\tfrac12\nu(1+p_1)\bigr)\cdot\Pr(y_i|x_i=l)\cdot\tfrac12.$$

Hence we compute the following ratio:

$$\frac{\Pr(x_i=l\mid\mathbf{y}_i)}{\Pr(x_i=h\mid\mathbf{y}_i)} = \left(\frac{\varepsilon}{1-\varepsilon}\right)^{y_i}\left(\frac{\varepsilon p_1+1-\varepsilon}{\varepsilon+p_1-\varepsilon p_1}\right)^{L_i}\left(\frac{p_1-\varepsilon p_1+\varepsilon}{1-\varepsilon+\varepsilon p_1}\right)^{H_i}.$$

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

Proof of Lemma 4. Similar to the proof of Lemma 3, we find the expression $\Pr(x_i=h,\mathbf{y}_i)$ by replacing p_1 in (37) with p_β. The expression for $\Pr(x_i=l,\mathbf{y}_i)$ is

$$\Pr(x_i=l,\mathbf{y}_i) = \prod_{ij\in G}\nu\bigl[p_\beta\Pr(y_j|x_j=h)+p_\theta\Pr(y_j|x_j=l)\bigr] \times\prod_{ij\notin G,\,j\ne i}\bigl(1-\tfrac12\nu(p_\theta+p_\beta)\bigr)\cdot\Pr(y_i|x_i=l)\cdot\tfrac12.$$

So the ratio becomes

$$\frac{\Pr(x_i=l\,|\,\mathbf{y}_i)}{\Pr(x_i=h\,|\,\mathbf{y}_i)}=\left(\frac{1-\nu(p_\theta+p_\beta)/2}{1-\nu(1+p_\beta)/2}\right)^{S-n_i}\cdot\left(\frac{\varepsilon}{1-\varepsilon}\right)^{y_i}\times\left(\frac{\varepsilon p_\beta+p_\theta-\varepsilon p_\theta}{\varepsilon+p_\beta-\varepsilon p_\beta}\right)^{L_i}\left(\frac{p_\beta-\varepsilon p_\beta+\varepsilon p_\theta}{1-\varepsilon+\varepsilon p_\beta}\right)^{H_i}.$$

Taking the limits of S and ν completes the proof. □

Proof of Lemma 5. We want to study the expected log odds as a function of θ. By Lemma 4, we have

$$\begin{aligned}\mathbb{E}[R_i(\theta,\beta)\,|\,l,\theta,\beta]&=(1-2\varepsilon)\log\left(\frac{\varepsilon}{1-\varepsilon}\right)\\&\quad-\tfrac{1}{2}N[\varepsilon p_\beta+(1-\varepsilon)p_\theta]\log\left(\frac{\varepsilon p_\beta+(1-\varepsilon)p_\theta}{\varepsilon+(1-\varepsilon)p_\beta}\right)\\&\quad-\tfrac{1}{2}N[\varepsilon p_\theta+(1-\varepsilon)p_\beta]\log\left(\frac{\varepsilon p_\theta+(1-\varepsilon)p_\beta}{\varepsilon p_\beta+(1-\varepsilon)}\right)\\&\quad-\tfrac{1}{2}N(1-p_\theta).\end{aligned}\tag{38}$$

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

$$\begin{aligned}\frac{\partial\,\mathbb{E}[R_i(\theta,\beta)\,|\,l,\theta,\beta]}{\partial\theta}&=\frac{\partial\,\mathbb{E}[R_i(\theta,\beta)\,|\,x_i=l,\theta,\beta]}{\partial p_\theta}\cdot\frac{\partial p_\theta}{\partial\theta}\\&=\tfrac{1}{2}N\left[(1-\varepsilon)\log\left(\frac{\varepsilon p_\beta+(1-\varepsilon)p_\theta}{\varepsilon+(1-\varepsilon)p_\beta}\right)+\varepsilon\log\left(\frac{\varepsilon p_\theta+(1-\varepsilon)p_\beta}{\varepsilon p_\beta+(1-\varepsilon)}\right)\right]f(\theta).\end{aligned}\tag{39}$$

Note that the derivative is strictly increasing in p_θ, thus strictly decreasing in θ. By the definition we gave to $\bar\theta$, the derivative is zero at $\bar\theta(\beta)$. So we conclude that the expected log odds as a function of θ is quasi-concave with the maximum attained at $\bar\theta(\beta)$. □

Proof of Lemma 6. We want to study the expected log odds as a function of β. First, the expected log odds is expressed as in (38). We take its derivative w.r.t. β

$$\begin{aligned}\frac{\partial\mathbb{E}[R_i(\theta,\beta)|x_i=l,\theta,\beta]}{\partial\beta}&=\tfrac{1}{2}N\left[\varepsilon\log\left(\frac{\varepsilon p_\beta+(1-\varepsilon)p_\theta}{\varepsilon+(1-\varepsilon)p_\beta}\right)+(1-\varepsilon)\log\left(\frac{\varepsilon p_\theta+(1-\varepsilon)p_\beta}{\varepsilon p_\beta+(1-\varepsilon)}\right)\right.\\&\quad\left.+\frac{\varepsilon^2-(1-\varepsilon)^2p_\theta}{\varepsilon+(1-\varepsilon)p_\beta}+\frac{(1-\varepsilon)^2-\varepsilon^2p_\theta}{\varepsilon p_\beta+(1-\varepsilon)}\right]f(\beta).\end{aligned}$$

Since f is positive, we focus on the term within the brackets. Using the inequality $\log(t)<t-1$ except for $t=1$, we have

$$\begin{aligned}[\cdots]&<\frac{\varepsilon^2p_\beta+\varepsilon(1-\varepsilon)p_\theta}{\varepsilon+(1-\varepsilon)p_\beta}+\frac{\varepsilon(1-\varepsilon)p_\theta+(1-\varepsilon)^2p_\beta}{\varepsilon p_\beta+(1-\varepsilon)}\\&\quad-\frac{\varepsilon(1-\varepsilon)p_\beta+(1-\varepsilon)^2p_\theta}{\varepsilon+(1-\varepsilon)p_\beta}-\frac{\varepsilon(1-\varepsilon)p_\beta-\varepsilon^2p_\theta}{\varepsilon p_\beta+(1-\varepsilon)}.\end{aligned}$$

Note that the right side is (linearly) decreasing in p_θ. Recall that the condition of the Lemma is $\theta<\beta$, which implies that $p_\theta>p_\beta$. Hence we replace p_θ by p_β on the right-hand side of the above inequality

$$[\cdots]<\left[\frac{\varepsilon^2-(1-\varepsilon)^2}{\varepsilon+(1-\varepsilon)p_\beta}+\frac{(1-\varepsilon)^2-\varepsilon^2}{\varepsilon p_\beta+(1-\varepsilon)}\right]p_\beta,$$

which is smaller than zero because the denominator in the first term is smaller than that of the second. □

Proof of Proposition 4. For notational simplicity, we omit the (θ,β) in the conditional of any expectation operator. We also use R_i short for $R_i(\theta,\beta)$.

We start with a low-type person. With $P_i(\theta,\beta)$ given by (15), we can easily write down i's expected credit score for any $\theta_i\ge\theta$

$$\begin{aligned}\mathbb{E}(R_i\,|\,l,\theta_i)&=(1-2\varepsilon)\log\left(\frac{\varepsilon}{1-\varepsilon}\right)\\&\quad-\tfrac{1}{2}N\left[\varepsilon p_\beta+(1-\varepsilon)\int_{\theta_i}^{+\infty}f(t)\,dt\right]\log\left(\frac{\varepsilon p_\beta+(1-\varepsilon)p_\theta}{\varepsilon+(1-\varepsilon)p_\beta}\right)\\&\quad-\tfrac{1}{2}N\left[\varepsilon\int_{\theta_i}^{+\infty}f(t)\,dt+(1-\varepsilon)p_\beta\right]\log\left(\frac{\varepsilon p_\theta+(1-\varepsilon)p_\beta}{\varepsilon p_\beta+(1-\varepsilon)}\right)\\&\quad-\tfrac{1}{2}N(1-p_\theta).\end{aligned}$$

Thus for any $\theta_i\ge\theta$,

$$\frac{\partial\,\mathbb{E}(R_i\,|\,l,\theta_i)}{\partial\theta_i}=\tfrac{1}{2}N\left[(1-\varepsilon)\log\left(\frac{\varepsilon p_\beta+(1-\varepsilon)p_\theta}{\varepsilon+(1-\varepsilon)p_\beta}\right)+\varepsilon\log\left(\frac{\varepsilon p_\theta+(1-\varepsilon)p_\beta}{\varepsilon p_\beta+(1-\varepsilon)}\right)\right]f(\theta_i).\tag{40}$$

Next for the social utility $\mathbb{E}(\sum_{ij\in G}m_{ij}-1_{\{x_j=h\}}\,|\,l,\theta_i)$, which we use $\mathbb{E}(u_i\,|\,l,\theta_i)$ as a shorthand for, we have for any $\theta_i\ge\theta$,

$$\mathbb{E}(u_i\,|\,l,\theta_i)=\tfrac{1}{2}N\left(\int_\beta^{+\infty}(t-1)f(t)\,dt+\int_{\theta_i}^{+\infty}tf(t)\,dt\right).$$

Thus for any $\theta_i\ge\theta$,

$$\frac{\partial\,\mathbb{E}(u_i\,|\,l,\theta_i)}{\partial\theta_i}=-\tfrac{1}{2}N\theta_if(\theta_i).\tag{41}$$

First, we show that $\theta^*>0$. Consider the case wherein every low type chooses $\theta=0$. Clearly, $\partial\,\mathbb{E}(u_i\,|\,l,\theta_i)/\partial\theta_i=0$ but $\partial\,\mathbb{E}(R_i\,|\,l,\theta_i)/\partial\theta_i>0$ at $\theta_i=0$ for any $\beta\ge1$. Hence $\partial\,\mathbb{E}(U_i\,|\,l,\theta_i)/\partial\theta_i>0$ and the low type wants to increase θ_i above 0 and be more exclusive towards her fellows. This incentive to deviate means that $\theta=0$ cannot be part of an equilibrium.

Second, to show that $\theta^*<\bar\theta(\beta^*)$, we use our refinement. Suppose (θ^*,β^*) is an equilibrium where $\theta^*\ge\bar\theta(\beta^*)$. Now consider a behavior θ^{**} that is smaller than but sufficiently close to $\bar\theta(\beta^*)$ for the low type. Every low type will be better off in (θ^{**},β^*) than in (θ^*,β^*), because (i) by Lemma 5, we know that $\mathbb{E}(R_i\,|\,l)$ is quasi-concave in θ and differentiably maximized at $\bar\theta(\beta)$; and (ii) the social utility

$$\mathbb{E}(u_i\,|\,l)=\tfrac{1}{2}N\left(\int_\beta^{+\infty}(t-1)f(t)\,dt+\int_\theta^{+\infty}tf(t)\,dt\right)$$

is strictly decreasing in θ

$$\frac{\partial \mathbb{E}(u_i \mid l)}{\partial \theta} = -\tfrac{1}{2} N \theta f(\theta). \tag{42}$$

Thus $\mathbb{E}(U_i \mid l)$ is decreasing in θ on the interval $[\theta^{**}, \theta^*]$. Furthermore, given that every other low type chooses $\theta = \theta^{**}$ and every high type chooses $\beta = \beta^*$, a low-type i has no incentive to increase her criterion θ_i beyond θ^{**}. This can be seen by comparing (39) with (40) and (41) with (42), and from $\partial \mathbb{E}(U_i \mid l, \theta_i)/\partial \theta_i < 0$ for $\theta_i > \theta^{**}$; nor does she have incentive to lower the criterion because doing so changes nothing (since friendships are established mutually). We conclude that (θ^*, β^*) fails the refinement.

Last, we turn to the high types. We show that $\beta = 1$ cannot be part of an equilibrium. The argument is similar to that for the low types. Briefly, consider a symmetric profile $(\theta, \beta = 1)$. To be an equilibrium, it must be that $\theta < \bar{\theta}(1)$. This implies that $\partial \mathbb{E}(R_i \mid h, \beta_i)/\partial \beta_i > 0$ at $\beta_i = 1$. But $\partial \mathbb{E}(u_i \mid h, \beta_i)/\partial \beta_i = 0$ at $\beta_i = 1$. Hence a high type wants to raise her β_i above 1. This incentive to deviate means that $\beta = 1$ cannot be part of an equilibrium. □

Proof of Proposition 5. Let Θ be the set of equilibria without refinement. Given that $\alpha_h = 0$ and $\beta = 1$, we are effectively looking for the point(s) in Θ that maximizes the expected total utility of low types.

Using (39) with $\beta = 1$, we see that $(\partial \mathbb{E}(R_i \mid l)/\partial \theta)/f(\theta)$ is strictly decreasing in θ and equals 0 at $\theta = \bar{\theta}$. Using (42), we see $(\partial \mathbb{E}(u_i \mid l)/\partial \theta)/f(\theta)$ is strictly decreasing in θ and equals 0 at $\theta = 0$. These imply that

$$\frac{\partial \mathbb{E}(u_i \mid l)}{\partial \theta} + \alpha_l \frac{\partial \mathbb{E}(R_i \mid l)}{\partial \theta}$$

is strictly decreasing and has a single point within $(0, \bar{\theta})$ where it is zero. This is also where $\mathbb{E}(U_i \mid l)$ is maximized. Denote this point by θ^*. We would be done if θ^* is shown to be an equilibrium without refinement. By comparing (39) with (40) and (42) with (41), clearly at $\theta_i = \theta^*$ and $\theta_i \geq \theta^*$,

$$\frac{\partial \mathbb{E}(u_i \mid l, \theta_i)}{\partial \theta_i} + \alpha_l \frac{\partial \mathbb{E}(R_i \mid l, \theta_i)}{\partial \theta_i} \leq 0,$$

and the inequality is strict for $\theta_i > \theta^*$. This implies that a low-type i has no incentive to deviate if every other low type chooses θ^*. □

Proof of Lemma 7. Mathematically it is a special case of Lemma 4, with $p_\beta = p_\theta = p_1$. □

Proof of Proposition 6. The arguments closely resemble those of the proof of Proposition 4. Here we discuss them briefly.

First, consider a candidate profile (θ, β) with $\theta \leq \bar{\theta}(\beta)$. One can show that a low type has incentive to increase her criterion because doing so increases both her social utility and credit score. So it cannot be an equilibrium.

Second, we filter $\theta^* = 1$ with refinement. Suppose that (θ^*, β^*) is an equilibrium. Compare it with (θ^{**}, β^*) where θ^{**} is smaller but sufficiently close to 1. One can show that low types are better off under θ^{**} and that no single low type has incentive to increase criterion under θ^{**}.

Last, consider a candidate profile (θ, β) with $\beta = 1$. From the above we know that, for it to be an equilibrium, it must be that $\theta > \bar{\theta}(\beta)$, which says that removal of one low-type friend strictly increases one's expected credit score. Hence a high type has incentive to raise her criterion above 1. So it cannot be an equilibrium. □

Proof of Proposition 7. First, we look at the case of discrimination. Taking first-order condition (FOC) w.r.t. e_i, we have for each i

$$e_i^* = ab^{-1} + \phi\Bigg(H_i + \sum_{ij \in G, x_j = l} e_j^*\Bigg),$$

which we write in the matrix form

$$\mathbf{e}^* = ab^{-1} + \phi \mathbf{H} + \phi \mathbf{A}_l \mathbf{e}^*.$$

This implies that

$$(\mathbf{I} - \phi \mathbf{A}_l)\mathbf{e}^* = ab^{-1} + \phi \mathbf{H}.$$

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

The case of homophily can be similarly proved. In particular, the FOC is

$$e_i^* = (H_i - L_i) b^{-1} + 2b^{-1} \sum_{ij \in G,\, x_j = l} e_j^* + ab^{-1} + \phi\Bigg(H_i + \sum_{ij \in G,\, x_j = l} e_j^*\Bigg).$$

One can again write it into matrix form and solve for $\mathbf{e}^*$. □

Proof of Lemma 8. Using

$$\begin{aligned}
\Pr(x_i = h \mid \mathbf{y}_i) &= \prod_{ij \in G} \nu\big[q_s \Pr(y_j \mid x_j = h) + q_d \Pr(y_j \mid x_j = l)\big] \\
&\quad \times \prod_{ij \notin G,\, j \neq i} \big(1 - \tfrac{1}{2}\nu(q_s + q_d)\big) \cdot \Pr(y_i \mid x_i = h) \cdot \tfrac{1}{2}, \\
\Pr(x_i = l \mid \mathbf{y}_i) &= \prod_{ij \in G} \nu\big[q_d \Pr(y_j \mid x_j = h) + q_s \Pr(y_j \mid x_j = l)\big] \\
&\quad \times \prod_{ij \notin G,\, j \neq i} \big(1 - \tfrac{1}{2}\nu(q_s + q_d)\big) \cdot \Pr(y_i \mid x_i = l) \cdot \tfrac{1}{2}.
\end{aligned}$$

So the ratio is

$$\begin{aligned}
&\frac{\Pr(x_i = l \mid \mathbf{y}_i)}{\Pr(x_i = h \mid \mathbf{y}_i)} \\
&\quad = \left(\frac{\varepsilon}{1 - \varepsilon}\right)^{y_i} \left(\frac{\varepsilon q_d + q_s - \varepsilon q_s}{\varepsilon q_s + q_d - \varepsilon q_d}\right)^{L_i} \left(\frac{q_d - \varepsilon q_d + \varepsilon q_s}{q_s - \varepsilon q_s + \varepsilon q_d}\right)^{H_i},
\end{aligned}$$

which, together with $\Pr(x_i = l \mid \mathbf{y}_i) + \Pr(x_i = l \mid \mathbf{y}_i) = 1$, gives us the result. □

Proof of Proposition 8. Again for notational simplicity, all expectation operators in this proof are conditional on the symmetric profile (e_l, e_h). So, for example, $\mathbb{E}(U_i \mid l, e_i)$ actually refers to $\mathbb{E}(U_i \mid l, e_l, e_h, e_i)$, which is the expected utility of a low type when she chooses e_i, while everyone else follows (e_l, e_h).

Given (13), we see that for any individual i, an additional high-type friend increases (and an additional low-type friend decreases) the expected utility from credit score by

$$D_{x_i} = \alpha_{x_i}\left[\varepsilon \log\left(\frac{\varepsilon+(1-\varepsilon)p_1}{\varepsilon p_1+(1-\varepsilon)}\right)+(1-\varepsilon)\log\left(\frac{\varepsilon p_1+(1-\varepsilon)}{\varepsilon+(1-\varepsilon)p_1}\right)\right].$$

Clearly, the friendship formation criteria will be: A low type accepts another low type, iff, $m_{ij} > D_l$, a high type accepts a low type, iff, $m_{ij} > 1 + D_h$, a low type accepts a high type, iff, $m_{ij} > 1 - D_l$, and a high type accepts another high type, iff, $m_{ij} > -D_h$ (which always holds). Using these criteria and the requirement that a tie is formed upon mutual acceptance, the expected utility of a low type when she chooses e_i is

$$\mathbb{E}(U_i \mid l, e_i) = (e_i+e_l)M\int_{D_l}^{\infty}(t-D_l)f(t)\,dt + (e_i+e_h)M\int_{1+D_h}^{\infty}(t-1+D_l)f(t)\,dt,$$

and the expected utility of a high type when she chooses e_i is

$$\mathbb{E}(U_i \mid h, e_i) = (e_i+e_l)M\int_{1+D_h}^{\infty}(t-1-D_h)f(t)\,dt + (e_i+e_h)M\int_{0}^{\infty}(t+D_h)f(t)\,dt.$$

Using FOCs, we have, in an equilibrium,

$$e_l^* = M\int_{D_l}^{\infty}(t-D_l)f(t)\,dt + M\int_{1+D_h}^{\infty}(t-1+D_l)f(t)\,dt, \quad (43)$$

and

$$e_h^* = M\int_{1+D_h}^{\infty}(t-1-D_h)f(t)\,dt + M\int_{0}^{\infty}(t+D_h)f(t)\,dt. \quad (44)$$

Comparing (43), (44), and (24), one can verify that: (i) $e_l^* < e^* < e_h^*$ for $\alpha_l = \alpha_h > 0$, (ii) $e_l^* > e_h^*$, iff

$$\int_{1+D_h}^{D_l}(D_l-t)f(t)\,dt > \int_0^{1+D_h} D_h f(t)\,dt + \int_0^{1+D_h} tf(t)\,dt,$$

which holds for sufficiently large D_l. □

References

Ambrus A, Mobius M, Szeidl A (2014) Consumption risk-sharing in social networks. *Amer. Econom. Rev.* 104(1):149–182.

Armour S (2014) Borrowers hit social-media hurdles. *Wall Street Journal* (January 8). Accessed June 30, 2014, http://www.wsj.com/articles/SB10001424052702304773104579266423512930050.

Arrow KJ (1998) What has economics to say about racial discrimination? *J. Econom. Perspectives* 12(2):91–100.

Bagherjeiran A, Bhatt RP, Parekh R, Chaoji V (2010) Online advertising in social networks. Furht B, ed. *Handbook of Social Network Technologies and Applications* (Springer, New York), 651–689.

Bakshy E, Eckles D, Yan R, Rosenn I (2012) Social influence in social advertising: Evidence from field experiments. *Proc. 13th ACM Conf. Electronic Commerce (EC'12)* (ACM, New York), 146–161.

Ball S, Eckel C, Grossman P, Zame W (2001) Status in markets. *Quart. J. Econom.* 116(1):161–188.

Ballester C, Calvo-Armengol A, Zenou Y (2006) Who's who in networks. Wanted: The key player. *Econometrica* 74(5):1403–1417.

Becker GS (1971) *The Economics of Discrimination*, 2nd ed. (University of Chicago Press, Chicago).

Benoit DF, Van den Poel D (2012) Improving customer retention in financial services using kinship network information. *Expert Systems Appl.* 39(13):11435–11442.

Bonacich P (1987) Power and centrality: A family of measures. *Amer. J. Sociol.* 92(5):1170–1182.

Bramoullé Y, Kranton R (2007a) Risk sharing across communities. *Amer. Econom. Rev.* 97(2):70–74.

Bramoullé Y, Kranton R (2007b) Risk-sharing networks. *J. Econom. Behav. Organ.* 64(3–4):275–294.

Bursztyn L, Ederer F, Ferman B, Yuchtman N (2014) Understanding mechanisms underlying peer effects: Evidence from a field experiment on financial decisions. *Econometrica* 82(4):1273–1301.

Castilla EJ (2005) Social networks and employee performance in a call center. *Amer. J. Sociol.* 110(5):1243–1283.

Chui M (2013) Social media to boost financial services. *InFinance* 127(2):34–36.

Clover C (2015) Backlash in China over wechats targeted adverts. *Financial Times* (January 28). Accessed February 22, 2015, http://www.ft.com/intl/cms/s/0/0fd1abc2-a610-11e4-abe9-00144feab7de.html.

Demirgüç-Kunt A, Levine R (2009) Finance and inequality: Theory and evidence. *Ann. Rev. Financial Econom.* 1(1):287–318.

Dixon P, Gelman R (2014) The scoring of America: How secret consumer scores threaten your privacy and your future. Accessed February 22, 2015, http://worldprivacyforum.org/2014/04/wpf-report-the-scoring-of-america-how-secret-consumer-scores-threaten-your-privacy-and-your-future/.

Feigenberg B, Field EM, Pande R (2013) The economic returns to social interaction: Experimental evidence from microfinance. *Rev. Econom. Stud.* 80(4):1459–1483.

Galak J, Small D, Stephen AT (2011) Micro-finance decision making: A field study of prosocial lending. *J. Marketing Res.* 48(SPL):S130–S137.

Goel S, Goldstein DG (2013) Predicting individual behavior with social networks. *Marketing Sci.* 33(1):82–93.

Guseva A, Rona-Tas A (2001) Uncertainty, risk, and trust: Russian and American credit card markets compared. *Amer. Sociol. Rev.* 66(5):623–646.

Haenlein M (2011) A social network analysis of customer-level revenue distribution. *Marketing Lett.* 22(1):15–29.

Haenlein M, Libai B (2013) Targeting revenue leaders for a new product. *J. Marketing* 77(3):65–80.

Herzenstein M, Sonenshein S, Dholakia UM (2011) Tell me a good story and I may lend you money: The role of narratives in peer-to-peer lending decisions. *J. Marketing Res.* 48(SPL):S138–S149.

Hill S, Provost F, Volinsky C (2006) Network-based marketing: Identifying likely adopters via consumer networks. *Statistical Sci.* 21(2):256–276.

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

Hu Y, Van den Bulte C (2014) Nonmonotonic status effects in new product adoption. *Marketing Sci.* 33(4):509–533.

Iyengar R, Van den Bulte C, Lee JY (2015) Social contagion in new product trial and repeat. *Marketing Sci.* 34(3):408–429.

Jenkins P (2014) Big data lends new zest to banks' credit judgments. *Financial Times*. Accessed June 24, 2014, http://www.ft.com/intl/cms/s/0/dfe64c0c-fadd-11e3-8959-00144feab7de.html.

Kornish LJ, Li Q (2010) Optimal referral bonuses with asymmetric information: Firm-offered and interpersonal incentives. *Marketing Sci.* 29(1):108–121.

Lin M, Prabhala NR, Viswanathan S (2013) Judging borrowers by the company they keep: Friendship networks and information asymmetry in online peer-to-peer lending. *Management Sci.* 59(1):17–35.

Liu K, Tang L (2011) Large-scale behavioral targeting with a social twist. *Proc. 20th ACM Internat. Conf. Inform. Knowledge Management* (ACM, New York), 1815–1824.

Lohr S (2015) Banking start-ups adopt new tools for lending. *New York Times* (January 18). Accessed February 22, 2015, http://www.nytimes.com/2015/01/19/technology/banking-start-ups-adopt-new-tools-for-lending.html?_r=0.

Lu Y, Jerath K, Singh PV (2013) The emergence of opinion leaders in a networked online community: A dyadic model with time dynamics and a heuristic for fast estimation. *Management Sci.* 59(8):1783–1799.

MacKenzie D (2006) *An Engine, Not a Camera: How Financial Models Shape Markets* (MIT Press, Cambridge, MA).

MacKenzie D, Millo Y (2003) Constructing a market, performing theory: The historical sociology of a financial derivatives exchange. *Amer. J. Sociol.* 109(1):107–145.

National Consumer Law Center (2014) Big data: A big disappointment for scoring consumer credit risk. NCLC, Boston.

Pasquale F (2015) *The Black Box Society: The Secret Algorithms That Control Money and Information* (Harvard University Press, Cambridge, MA).

Phelps ES (1972) The statistical theory of racism and sexism. *Amer. Econom. Rev.* 62(4):659–661.

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

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

Rosenblat TS, Mobius M (2004) Getting closer or drifting apart? *Quart. J. Econom.* 119(3):971–1009.

Roth PL, Bobko P, Van Iddekinge CH, Thatcher JB (2016) Social media in employee-selection-related decisions: A research agenda for uncharted territory. *J. Management* 42(1):269–288.

Rusli EM (2013) Bad credit? Start tweeting. *Wall Street Journal* (April 1) http://www.wsj.com/articles/SB10001424127887324883604578396852612756398.

Schmitt P, Skiera B, Van den Bulte C (2011) Referral programs and customer value. *J. Marketing* 75(1):46–59.

Sousa MR, Gama J, Brandao E (2013) Introducing time-changing economics into credit scoring. Technical Report 513, University of Porto, Porto, Portugal.

Stiglitz JE (1990) Peer monitoring and credit markets. *World Bank Econom. Rev.* 4(3):351–366.

Sudhir K, Priester J, Shum M, Atkin D, Foster A, Iyer G, Jin G, et al. (2015) Research opportunities in emerging markets: An inter-disciplinary perspective from marketing, economics, and psychology. *Cust. Need. Solut.* 2(4):264–276.

Toubia O, Stephen AT (2013) Intrinsic vs. image-related utility in social media: Why do people contribute content to Twitter? *Marketing Sci.* 32(3):368–392.

Townsend RM (1994) Risk and insurance in village India. *Econometrica* 62(3):539–591.

van Alstyne M, Brynjolfsson E (2005) Global village or cyber-balkans? Modeling and measuring the integration of electronic communities. *Management Sci.* 51(6):851–868.

Van den Bulte C, Wuyts S (2007) *Social Networks and Marketing* (Marketing Science Institute, Cambridge, MA).

Zeng Z, Xie Y (2008) A preference-opportunity-choice framework with applications to intergroup friendship. *Amer. J. Sociol.* 114(3):615–648.