Multi-level Modeling of Social Roles in Online Micro-lending Platforms

LU SUN, Human-Computer Interaction Institute, Carnegie Mellon University, USA
ROBERT E. KRAUT, Human-Computer Interaction Institute, Carnegie Mellon University, USA
DIYI YANG, School of Interactive Computing, Georgia Institute of Technology, USA

In many teams, members play distinct roles, from leader to disrupter to social networker. Understanding the roles that contributors enact in micro-lending platforms is both psychologically and socially important for sustaining members' motivation and coordinating their joint efforts. Knowing what roles exist in these communities or which additional ones might be needed can also help teams function more effectively. In this paper, we identify social roles in Kiva.org, a peer-to-peer micro-funding platform, by utilizing members' lending behaviors, social network behaviors and communication behaviors to model their social roles at three levels. At the individual level, this method discovered active lenders who made many loans, early-bird lenders who made loans well before deadlines, and lurkers who rarely lent. At the topical level, our method differentiated those who had broad interests and lent to many causes from those who made loans only to borrowers in certain geographic regions or industry sectors. At the team level, the method revealed eight team-oriented functional roles such as encouragers, reminders, competitors, followers, and welcomers. To demonstrate the utility of the team roles, we used regression analysis to show how the distribution of social roles within teams influences the amount of money teams lent. Implications for identifying roles and understanding their contributions to teams are discussed.

CCS Concepts: • **Human-centered computing → Computer supported cooperative work**; **Empirical studies in collaborative and social computing**.

Additional Key Words and Phrases: Social roles; Online micro-lending platforms; Team success

ACM Reference Format:
Lu Sun, Robert E. Kraut, and Diyi Yang. 2019. Multi-level Modeling of Social Roles in Online Micro-lending Platforms. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 133 (November 2019), 25 pages. https://doi.org/10.1145/3359235

1 INTRODUCTION

Millions of people participate in online communities to collaborate, exchange expertise and persuade others through digital communication. Despite the growth in the number of online groups and communities, making them successful is still challenging. A large body of research studying online communities has already investigated how members lead the group [61], communicate with others through social networks [54] and support others to contribute [60]. However, few studies explored how social roles that users play influence the online communities' success. The concept of *social role*

Authors' addresses: Lu Sun, Human-Computer Interaction Institute, Carnegie Mellon University, Pittsburgh, USA, lus1@andrew.cmu.edu; Robert E. Kraut, Human-Computer Interaction Institute, Carnegie Mellon University, Pittsburgh, USA, robert.kraut@cmu.edu; Diyi Yang, School of Interactive Computing, Georgia Institute of Technology, Atlanta, USA, diyi.yang@cc.gatech.edu.

2573-0142/2019/11-ART133 $15.00
https://doi.org/10.1145/3359235

is a tool for describing common patterns of users' behavior that emerge in a particular context with specific social goals [7, 55]. Social roles can help explain why users perform the work they want [56] and receive the support they need [58]. To identify roles in online communities, like Wikipedia [56], Usenet [22], and online health communities [58], researchers have clustered members' low-level behavior. For example, Welser et al. [53] tied behavioral and social signatures with social roles and used these roles to explain the coordination and users' contribution on Wikipedia. Yang et al. [58] measured support-related actions and language styles in online cancer support groups and algorithmically identified eleven common roles, including "support seeker", "newcomer welcomer", and "story sharer". However, since low-level behaviors differ across online communities, the roles responsible for success in different communities are also likely to vary. The current research explores the social roles and examines the roles associated with success in a micro-lending platform.

Most research examining roles in online communities have been conducted at a macro level by assuming each community has a single set of roles regardless of context [4, 22, 33, 53, 56]. In Wikipedia, for example, role modeling has explored the different kinds of roles involved in writing and editing articles, largely ignoring a different role structure among those who do administrative work or who program bots. It ignores differences in roles when editing articles individually or as part of a Wikipedia project. Although the assumption of a single role structure can capture some of the major roles in a community, it loses more nuanced role differentiation at a micro level. Hence, our paper introduced a multi-level modeling method which takes complicated role structures into account.

The current paper examines roles within peer-to-peer lending sites. Online micro-lending platforms provide opportunities for people to support low-income entrepreneurs around the world who lack access to credit and other financial services. We conduct our research in Kiva.org, the world's first and largest peer-to-peer micro-finance website. As in many other online communities [45], an overall modeling of roles within Kiva might reveal two major roles–lurkers and contributors. But in fact, the role structure is more complex, with different roles emerging when examined users' general behavior, lending behavior and team behavior.

Although Kiva started in 2005 and has grown rapidly, Kiva has struggled to retain members and encourage contributions. Fully 36% of the people who created a Kiva account have never made a loan, and 86% of lenders have only made one [34]. To encourage lending, Kiva introduced lending team in 2008, where people with common interests can come together to lend. Prior research has shown that becoming a member of a team increases lending, and group membership can be leveraged to promote pro-social behavior through increased lending activities and team posts [1, 13]. The current paper extends this research to more completely describe how members' interactions with each other influence team's leading and how teams leverage members' interests to coordinate joint efforts. One goal of the current paper is to understand the social roles that members play in these lending teams and to examine how role composition influences team success.

A role analysis of Kiva must take into account users' general behaviors in the community, their personal contributing behaviors and their behaviors in teams. This multi-level analysis may allow the discovery of finer-grained roles, such as people who lend for different purposes or who perform different types of coordination work in lending teams, such as loan advocates, followers, and networkers [29]. To this end, we propose to model social roles in three distinction levels in Kiva - an individual level, a topical level and a team level. Then, we focus on individuals' differences at each level. Specifically, we first model 502,752 users' individual roles by looking at generic individual behaviors, such as how much they lend and how long they participate in Kiva. We then focus on 403,984 lenders(80.35% of all users) who have made at least one loan and derive their topical roles by clustering their lending interests. For the 28,535 individuals (5.68% of all users) who also participate in Kiva teams, we derive their team roles by clustering their actions within teams. In the end, each participant in Kiva can be represented via a combined representation of these three different role levels, as shown in Figure 1. To the best of our knowledge, this is the first study that models social roles within a single organization across multiple levels.

In this work, we first identify behavioral features associated with each of the three level, including users' individual features for individual roles, topic features for topical roles and team features for team roles. We then use the Gaussian Mixture Model to cluster features in order to extract the latent roles that lenders occupy. In contrast to other methods used in role modeling, the Gaussian Mixture Model method assumes that an individual can occupy multiple roles within a level at the same time, rather than only one role. We then validate the social roles through a series of quantitative and qualitative methods to identify role that are simultaneously accurate (i.e., adequately accounting for the clustering of the low-level features) and interpretable. This method identified three individual roles, seven topical roles and eight team roles. Individual roles captured the representative behaviors of all users in the platform, which benefits our understanding of the overall Kiva community. Topical roles revealed users' lending interests on different loan categories. Team roles captured users representative behaviors in a team.

Understanding the function and influence of team social roles can account for team success. To demonstrate this point, we conducted a linear regression analysis to measure the extent to which the distribution of roles within a team predicts the amount the team lends, a production measure of team success [36]. The results show that adding team roles to a base model improves the ability to predict team contribution and helps us interpret the function of team roles. In particular, teams that have a relatively even distribution of who occupy eight roles and those with members who occupy the "competitor" and "reminder" roles lend more, while teams with more members enacting "follower" and "individualists" roles lend less. These results improve our understanding of how members' leadership behaviors, persuading behaviors and social network positions influence micro-lending teams. To sum up, this multi-level social role modeling approach could contribute to the understanding of user interactions and behaviors in the Kiva micro-lending platform. Meanwhile, multi-level modeling of roles may also be used to develop recommendation systems to automatically suggest teams, loans and activities within teams. Furthermore, the results can also be used to design different strategies to encourage team members.

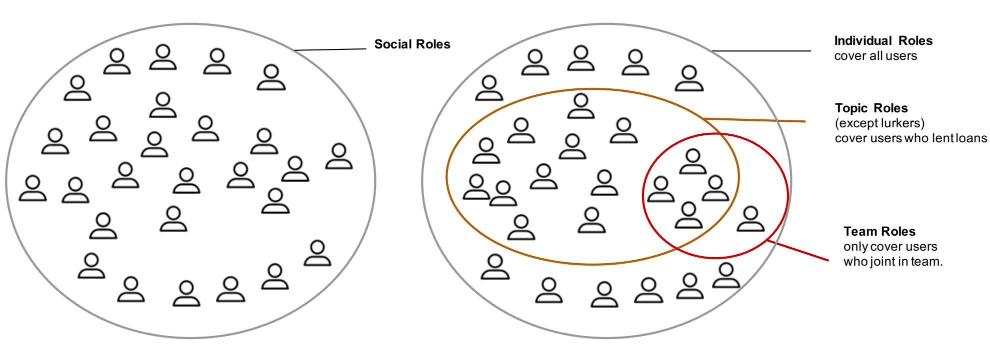

Fig. 1. Social roles in multi-levels vs. Social roles in a single level

2 RELATED WORK

2.1 Social Roles in Conventional Organizations

The concept of social roles has long been used in social sciences to understand people's behavior in conventional organizations [6, 7]. A *role* is generally defined as a cluster of related and goal-directed behaviors characteristic of a person within a specific situation [47]. Sometimes the role-relevant behaviors are informed by well-defined expectations and responsibilities defining the way that a role occupant should behave [27]. Thus, the social science literature distinguishes between in-role behaviors, which are defined by expectations associated with a job description, and extra-role behaviors or organizational citizenship behaviors, which are not defined by a job description, but are still important for an organization [44, 52]. The traditional social science literature has focused on factors associated with a particular social role [16], factors that cause people to perform their roles well and the downstream consequences of roles on organizational performance [15]. In a classic paper in the context of teams, Benne and Sheats [5] classified functional roles occupied by team members into three broad groupings: group task roles that facilitate and coordinate group effort, such as coordinator, orienter and information giver; group building and maintenance roles that maintain the group's way of working and that regulate the group, such as encourager, gate-keeper and follower, and individual roles that target on individual goals instead of group tasks, such as aggressor, recognition-seeker and help-seeker.

One problem with traditional role theory [7] is that roles are generally vaguely defined, with substantial disagreements about the roles that constitute a work group. In a review of the literature, Mumford and colleagues [41] identified 120 roles that have been applied to work teams. The contribution of this research was an attempt to collapse them into a set of eight teamwork roles. A related problem is that social sciences have few objective methods to identify the behaviors that comprise a role. One of the reasons for the distinction between in-role behaviors and extra-role ones, even if they are part of a "cluster of related and goal-directed behaviors characteristic of a person within a specific situation", is the lack of a method to behavioral cluster the behaviors associated with a role. Because of the ambiguities in defining roles in social sciences, most recent research on social roles in conventional organizations has focused on leadership roles [8].

2.2 Social Roles in Online Communities

More recently, social roles have been used to understand online communities [4, 22, 33, 53, 56]. In contrast to its use in understanding conventional organizations, much of the research on roles in the social computing literature has been methodological and descriptive, focusing on methods to identify the roles that exist in a community. The standard approach to quantitatively research social role identification is to use some type of unsupervised clustering technique on the content of users' interaction and their behavioral cues. For instance, Yang et al. [58] operationalized a set of support-related actions and language styles in members' messages in online cancer support groups and found eleven roles that members occupy, including "support seekers", "newcomer welcomers", and "story sharers". An analogous line of work used the social interactions that define social networks as the input to clustering. For instance, Welser et al. [54] and Fisher et al. [19] used social network signatures to obtain social roles such as "answer people" and "social networkers".

Hybrid approaches have also been used for role identification. For example, Welser et al. [53] combined Wikipedia editors' edit history and egocentric network features to identify four social roles: "substantive experts", "technical editors", "vandal fighters", and "social networkers". In addition to the use of content of interaction and network structure, existing studies have also relied on users' individual differences and other meta descriptions to better profile users [20, 48]. For instance, Arazy et al. [3] used Wikipedians' access privileges to identify twelve social roles, which aligned well with Wikipedia's organizational structure. Other research looked at roles via users' successive levels of participation, such as "readers" or "collaborators" [45]. Research in organizations proposed that organizational behavior should be captured both in macro and micro perspectives [29]. Studying social roles at the macro level by modeling social roles at one time for all populations, can reveal the overall contextual factors, such as situational constraints and demographics. However, it neglects individual behaviors, perceptions and interactions at the micro level. To this end, we propose to conduct multi-level modeling of social roles where multi-level refers to an individual level, a topical level and a team level.

Specifically, we first model users' individual roles at a community level by looking at lenders' behaviors, such as how much they lent and how long they stayed on Kiva. We then focus on active individual roles and derive topical roles by clustering their lending interests. Furthermore, for individuals who also participate in Kiva teams, we derive their team roles by measuring their specific actions in teams. To identify social roles on each level in a reliable way, we take advantage of social science theories to guide what types of roles to expect and what to measure. In the team roles level, team functional roles specified by Benne and Sheats [5] were used as seeds to make a hypothesis on candidate social roles. In the Kiva context, we expect that some task roles, group-maintenance roles and individual roles would emerge. For example, an earlier study of Kiva by Yang and Kraut [57] found that some users were especially likely to send persuasive messages to persuade others to contribute, analogous to the encourager group-maintenance role identified by Benne and Sheats [5]. In addition, we relied upon prior social science research to identify organizational behaviors which might indicate roles. For example, empirical research from conventional organizations demonstrated the importance of leader and leadership behaviors in helping group members effectively focus on the tasks and achieve a collective outcome [9, 61]. Extending this research, Zhu et al. [61] examined the effectiveness of four types of leadership behaviors in Wikipedia. Results showed that aversive and directive leadership behaviors increased contributions specifically related to a task request, whereas transactional and person-focused leadership behaviors increased more general motivation. Among these leadership behaviors, directive leadership was task-oriented, intended to set goals and direct people to achieve group goals. Person-focused leadership was intended to support the group by maintaining close social relationships with team members. In addition to leadership, social ties and social interactions help people gain support in achieving collaborative goals [19, 40, 53, 54].

3 RESEARCH SITE: KIVA.ORG

Our research is conducted in the context of Kiva.org, an online crowdfunding platform that enables people to lend to individuals or small businesses [35]. Although others have studied challenges in crowdfunding [13, 34, 38], relatively little is known about social roles on crowdfunding platforms. Founded in 2005, Kiva.org is one of the world's largest online peer-to-peer platforms where people can lend money to underserved entrepreneurs across the globe. Loans can be 25 dollars or more and are interest free. To increase lender engagement and contribution [39], Kiva instituted the "lending teams" program in 2008, where lenders can create teams or join existing ones. Lenders who have something in common (a common interest, company, religious affiliation, etc.) can also form a team to lend money together. When teams form, they appear on the Kiva team leaderboard, which ranks teams by the total amount of money contributed by their members. Lenders who have joined one or more teams could assign any loan they make to one of their teams. Lenders can also post messages in team forums to persuade other members to lend to particular loans [57].

Based on collaboration with Kiva, we were provided access to the basic profile information for members and public teams, including an anonymized user id, the names of the teams they joined, their locations, their registration time, and a "I loan because" statement in which users describe their motivation to lend. In addition, the dataset contained the complete lending activities of all lenders and a complete list of messages posted to lending public team message boards. As of Jan 2017 there were a total of 502,752 registered users, and 28,535 users exchanged approximately 493,646 messages in the public discussion forums. We used data from all registered users in the individual role modeling. Among all users, 403,984 users (80.35% of all users) made at least one loan. We used these users who made at least one loan to model topic roles. In order to explore users' interaction within the teams, we created a new user–team corpus where the unit of analysis is one user's behaviors in one team as a data point. This corpus comprises 28,535 users engaged in behaviors in 5879 teams, leading to 52,192 user–team data points.

4 ROLE IDENTIFICATION METHOD

Our role identification approach included defining the levels in which the roles occurred, hypothesizing emergent roles at each level, proposing role-relevant behaviors and clustering them to identify roles, interpreting roles and evaluating their utility. Users may take on different social roles at different levels. Our first level describes all users' general behaviors. After users log in to the Kiva website, they can provide loans. Some users actively return to the platform or enroll in a team. We targeted these generic behaviors of all users on the platform as the first level to derive "individual roles". In addition, since lending is the main task for users, how users lend to different categories of borrowers becomes important. Thus, we modeled "topical roles" for the users who lend loans to describe their specific loan interests and lending habits. For example, some users only lent to requests from Africa and some users were only interested in educational loans [26]. In Kiva, only 5.68% of users participated in a team, a small fraction of the community. Since team members give more money than those not in teams, understanding how users interact within the team matters for the development of the community [61]. Hence, here we separate the third level to differentiate behaviors that happened outside a team or within one. We restrict the sample to users who enrolled in a team and derive their "team roles" within the team based on their social network signature, behavioral patterns, and interaction within their teams.

For each level, we use the same method to identify the roles that members enact: clustering members who show similar patterns of behavior. To this end, we used Gaussian Mixture Model (GMM), which is a statistical model that assumes all data points are generated from a mixture of a finite number of Gaussian distributions with unknown parameters [37], to specify reasonable clusters of people as potential social roles. GMM has been widely applied in empirical research to cluster attributes to understand patterns or dynamics of online communities [51, 58]. In contrast to the traditional clustering method such as k-means clustering, which assumes individuals occur only a single role at a time [2], this mixture model allows each individual to enact multiple roles as the same time. This approach makes the role closer to real online community situation.

A Gaussian mixture model is a weighted sum of M component Gaussian densities as given by the following equation:

$$p(x \mid \lambda) = \sum_{i=1}^{M} w_i g(x \mid \mu_i, \Sigma_i) \tag{1}$$

where X is a set of observable behaviors over D-dimensions, $w_i, i = 1, \ldots, M$, are the mixture weights, and $g(x|\mu_i, \Sigma_i), i = 1, \ldots, M$, are the component Gaussian densities. Each component density is a D-variate Gaussian function with a mean vector μ_i and a covariance matrix Σ_i. The weight or a mixture coefficient represents the degree to which a user was associated with each cluster; that is, each user is modeled as a mixture of roles. The mixing probability $p_1, p_2, \ldots, p_M$, means and covariance matrices of the Gaussian density functions, are learned from data $\{x_i\}_{i=1}^{N}$, where the total number of users in the corpus is N. Given a large corpus of data, we estimate the covariance matrices using "full" matrices which adopted each component to its own general covariance matrix and perform well on large dataset. We first specified the behavior input features X from previous theory and then used the computational approach described in Section 6.1 to select the number of implicit roles K. We describe the complete procedure for role identification in the next two sections.

5 FROM THEORY TO MEASURES

As an initial step in role modeling, we specified features that are associated with user roles at different levels. Then at each level, we used these features as variables to build the Gaussian Mixture Model (GMM) to generate reasonable clusters of people as potential social roles. We specified generic features for all users who registered an account on Kiva to model individual roles. Similarly, for the subset of users who made loans, we used categories of the loans they made to model topic roles. For teams roles, we identified users' behaviors within a team based on previous theory and qualitative methods, such as building affinity diagrams to cluster user's common behaviors. These features are all selected based on previous theories which proposed a series of role related behaviors and factors. This section introduces the features we selected and the methods to calculate these features in detail. Feature statistics are described in the Table 2.

5.1 Individual Features

Users' participation in online communities is dependent on the types of people that form them. Some of the online community's members do not actively participate in the community, whereas others participate actively and support others. Although most research on online communities concentrated on people who actively contribute to meeting the community goals or to its discussions, most people in most online communities are not active, either acting as lurkers [42], observing what others had contributed, or dropouts, who leave the community shortly after joining [49]. Instead, our research starts with the general features of *all* registered users to understand what types of users exist over the whole community. We select these general features based on previous literature [1, 24], including users' participation, lending practices, and demographics.

Participation. The degree of participation is an important element in identifying users' social role, both outside and within teams [1, 24]. We calculated "participation" using two measures: participation in lending activities operationalized as their *loan count* and participation in teams operationalized as their *team count*, i.e., the number of teams to which a user belonged. We also captured *activity duration* as one aspect in participation, by calculating the number of days between their first and last loan.

Lending practice. Temporal patterns generally exist in online communities. For example, investments on Kickstarter show temporal effects with a peak of contributions at the beginning of a project and close to the investment deadline [51]. Hence, we measured users' average *time to deadline* as the time interval between when a lender made a loan and the deadline for making loans. Moreover, some lenders always concentrated on a particular class of loans or lenders from a certain continent and repeatedly lent to the same borrower. Here, we measured their "repetition" or "concentration" as *repeated loan count*, i.e. the number of times of users lent to the same borrower.

Lender demographic. Prior literature shows that the location of lenders is an important factor for team joining and pro-social lending behaviors [1]. Hence, we coded the lender's location as from US (1) or outside the US (0) as a *demographic feature.* According to our dataset, 88% of users on the Kiva platform were from US.

5.2 Topic Features

Kiva categorizes each loan into a loan category in order to help lenders select loans that they are interested in. Lenders who have same loan interests could be grouped together into topic roles to contribute. To understand which loan category did lenders lend together, we use loan categories, loan geographical location and lenders' lending activity concentration as topic features to group users who lend loans.

Loan categories. Kiva's loans are spread across fifteen categories: arts, retail, transportation, entertainment, food, clothing, housing, personal use, education, construction, health, wholesale, services, manufacturing, and agriculture. Since making loans is the major task in the community, different social roles might show various lending preferences. We use these fifteen categories as fifteen features to measure lender's lending topics.

Loan geographical location. Kiva borrowers come from different parts of the world. Prior research documented that some lenders discriminate positively towards borrowers who are lighter-skinned [26]. Hence, we consider a borrower's geographical information, which may influence lenders' preference. Since the continents of loans are broadly distributed, based on previous literature [26], we classify borrowers' location into seven continents: Africa, Antarctica, Asia, Europe, North America, Oceania, South America, and then used them as a geographical feature to measure lenders' geographical preferences. Statistics of loan features are described in Table 1.

loan category	**Frequency**	**(%)**	**loan location**	**Frequency**	**(%)**
Arts	22,484	1.94	Africa	319,691	27.52
Retail	241,169	20.76	Antarctica	0	0.00
Transportation	33,564	2.89	Asia	485,849	41.82
Entertainment	1,758	0.15	Europe	10,512	0.90
Food	268,602	23.12	North America	152,374	13.11
Clothing	69,854	6.01	Oceania	13,960	1.20
Housing	48,379	4.16	South and Central America	179,462	15.45
Personal Use	32,364	2.79			
Education	35,237	3.03			
Construction	16,681	1.44			
Health	12,580	1.08			
Wholesale	1,908	0.16			
Services	85,831	7.39			
Manufacturing	13,829	1.19			
Agriculture	277,607	23.89			

Table 1. Descriptive statistics of loan categorical features. Frequency of loan categories and loan location are presented.

Lending Activity Concentration. Some lenders may concentrate on one specific loan category and make most of their loans in that category while others may loan more broadly. Similarly, some may limit their loans to one continent while others may loan more. To capture these dimensions, we calculated lenders' *loan category entropy* and *loan location entropy*. Entropy is a widely used concept to describe the diversity of a probabilistic distribution. Zero entropy mean high concentration, in which lenders place all of their loans into a single category or location. Higher values indicate greater diversity, with lenders distributing their loans more evenly across categories or continents.

5.3 User-Team Features

In order to build a rich understanding of interactions within Kiva teams, we randomly sampled messages from five small groups and five large group discussions and then use these sample data to build an affinity diagram to summarize general behavioral patterns, including social networking, persuading, goal level, coordinating and competing. Here, we summarized users' behaviors from three aspects: lending features, social network features, and linguistic features.

Lending Features. Members' commitment to a team could influence their lending behaviors [1]. When lenders are members of multiple teams, they often have a preference for the team in whose name they would most like to contribute. We measured *team importance* as the proportion of a lender's loans made under a team's name. Higher team importance represents a user contributed to this team more than to other teams.

Social Network Features. Social network signatures could represent users' interaction patterns and reveal team dynamics. Previous research has used network structure and people's relationships with other users to identify social roles, such as distinguishing a small group of "answer people" from the many who ask questions [19, 53, 54]. In our work, we constructed a user-reply network and extracted features through network analysis, where the vertices represent members who have participated in the team, and edges represent direct replies. For each user, we used degree centrality to measure lenders' *centrality* in the whole social network, by calculating the number of other team members they link or link to them. Social network links could also indicate users' influence on others. In particular, some followers emerged in the online community who seemed to go along with the other team member's lending behavior [10]. We defined "following" here as people following others' recommendation on loans. When one user posted a loan information and loan links in the discussion forum, people might follow that user's suggestion and then donate to that loan. We calculated the *following* by computing how many loans they made by following links in the team discussion.

Linguistic Features. Lender's interaction behaviors, such as persuading, competing, goal setting, could be captured through text analysis [13, 57, 60]. We captured lender's linguistic features in terms of the degree to which the words in their messages corresponding to the semantic categories provided by psycho-linguistic lexicon LIWC [43] and the lexical categorizing tool Empath [18]. LIWC dictionaries have been broadly used to detect psychological content of text data in a wide variety of settings, including attentional focus, emotionality, social relationships and thinking styles. Larrimore et al. [31] extracted the linguistic features from a large number of loan requests in a peer-to-peer lending website using LIWC and analyzed the relationship between these language features and funding success. Another lexical categorizing tool we used is Empath, which allows users to define new semantic dictionaries by providing a small set of seed words. In contrast to LIWC, it allows for a richer vocabulary when building text categories. We construct a competing lexicon using Empath to measure users' competitive behavior. We also manually constructed lexicons to capture lenders behaviors under this specific context. Also, users' other behaviors, such as sharing information, are captured by counting the number of messages, the average number of links users sent and the average message length.

Cialdini and Cialdini [14] concluded that the influence of persuasion is based on six key principles: reciprocity, commitment and consistency, social proof, authority, liking, and scarcity. We borrowed previous measurements of persuading and used four related features: (1) The principle of scarcity demonstrated that people automatically assign the rare things more value. For example, users emphasize the limited time available to attract others' attention, such as posting the message "This man is married to a nurse and only has 13 minutes left to get his loan". We measure *explicit*

Features	Min	Max	Mean	Median	Sd
Individual Features					
team count	1.00	$1.34e4$	1.55	1.00	32.59
loan count	0.00	$2.17e5$	32.05	2.00	558.69
repeated loan count	0.00	0.94	0.07	0.00	0.15
activity duration(d)	0.00	1719	674	3.35	970
time to deadline	0.00	1.00	0.56	0.62	0.35
US location(US as 1)	0.00	1.00	0.22	0.00	0.41
Lending Topic Features					
loan category entropy	0.00	2.71	0.66	0.00	0.75
loan location entropy	0.00	1.77	0.46	0.00	0.55
User-Team Features					
team member count	0.00	$1.73e5$	148.4	12.0	2863.41
team loan amount	0.00	$5.33e7$	$86,961$	4325	$1.04e6$
team importance	0.00	1.00	0.37	0.15	0.40
following	0.00	1.00	0.03	0.00	0.11
centrality	0.00	407.00	1.82	0.00	8.30
number of message	1.00	$9.97e3$	9.21	1.00	97.74
number of link	0.00	$1.51e5$	11.76	0.00	885.01
message length	1.00	$8.62e5$	591.95	68.00	7600
mention scarcity	0.00	1.00	0.19	0.00	0.34
we word	0.00	1.00	0.02	0.01	0.03
i word	0.00	1.00	0.10	0.1	0.07
positive emotion	0.00	1.00	0.05	0.04	0.05
negative emotion	0.00	0.50	0.01	0.00	0.02
achievement	0.00	1.00	0.03	0.02	0.04
mention goal	0.00	1.00	0.00	0.00	0.05
mention modal verb	0.00	21.00	0.82	0.25	1.45
competing	0.00	0.33	0.00	0.00	0.01
mention greeting	0.00	1.00	0.03	0.00	0.13

Table 2. Descriptive statistics of individual features, lending topic features and user-team features.

mention scarcity by manually constructing an urgency lexicon that contains words such as “expire”, “remaining”, and “left” and then calculate the frequency of these words. (2) Some prior research in social psychology and related fields has already examined the effectiveness of systematic and heuristic behavioral cues in persuading team members to donate, including social proof, mentioning scarcity and showing social identity [38, 57]. *Social identity* is defined as an individual’s self-concept which derives from his knowledge of a social group (or groups) membership together with the emotional significance attached to that membership [50]. Social identity perspective has been applied to identify factors that influence inter-group behaviors and lead to success in distributed groups [59]. In addition to persuasive tactics [57], social identity is also shown as an important factor in team organization [60], and this feature may also indicate lenders’ intention to coordinate and encourage team members. We measured social identity by the frequency of mentioning the words related to the first-person pronouns plural (*mention we word*) using LIWC, e.g. we, us, and our. (3) In contrast, we distinguish *self identity* by the frequency of *mention “I” word* with the goal of specifying their self-identification as an individual or as a group member. We calculated this variable by calculating the frequency of words related to first-person pronouns singular, e.g. I, me, and my. (4) *negative emotion* and *positive emotion* are usually indicators for interpersonal interaction in text-based Computer-Mediated Communication [23]. Also, emotion factors could influence people choices [21, 46]. One study found that emotion information from weblogs is a good predictor for future stock market prices. To capture emotion information in teams, we measure the frequency of positive and negative emotion words using the positive emotion lexicon (e.g. love, nice, sweet) and negative emotion (e.g. hurt, ugly, nasty) lexicon in LIWC.

In many online communities, leaders exist who take responsibility for managing and coordinating team members to contribute. Although Kiva teams do not have an explicit leadership role, the behavior of some lenders reflects an intention to coordinate and motivate team members [17, 28, 30]. These lenders often use goal setting language to coordinate group members and manage their teams, since highlighting important group goals can motivate users to accomplish tasks that are important to the success of the group [60]. We measured *goal setting* through the frequency of explicitly *mentioning a goal*. We calculated this by manually constructing a goal setting lexicon that contains words, such as "our goal" and "leader boards". We also observed that some lenders used language about the team or their own achievement to motivate others. We therefore measured *achievement* by counting the words in the LIWC achievement dictionary, e.g. earn, hero,win. Another mechanism to facilitate team competition and team coordination is through information sharing [60]. We measure *information sharing* using the number of messages, the average number of links and the average message length which lenders sent to the team. Lenders also used modal verbs such as can, should, and will to express requests, suggestions and advice. Hence, we manually constructed a *modal word* lexicon using primary modal verbs in English includes can, could, may, might, must, shall, should, will and would and calculated the frequency of these words.

Empirical studies have shown that messages that promote competition in the form of encouraging members to help the team maintain or boost its ranking on Kiva's team leaderboard [13]. Thus, we measure *competition* by extracting comparative words such as competing, winner, and fight, per message using Empath. Based on our observation, some team members sometimes sent messages to welcome newcomer, such as "Welcome Barbara! Great to have you on the team! Now we just need to talk it up with the rest of the family!". We operationalized *welcoming* by calculating the frequency of mentioning greeting words from Empath, such as "Welcome" or, "Hi everyone".

6 ROLE EVALUATION

After identifying users' actions at the individual, topical and team levels, we used Gaussian Mixture Modeling (GMM) to cluster them into roles [37]. In order to fit the data in GMM, we first specified the appropriate number or roles to extract at each level. GMM results provide us the representative features for each role, and we then used these features to evaluate the quality of the social role identification. In particular, at the team level, we conducted a more detailed qualitative analysis to set the number of team roles. We conducted semi-structured interviews to validate whether these representative behaviors could differentiate team roles. As a next step, we introduced our social role modeling results at individual, topical and team levels. For team roles, we also developed an evaluation method to determine the extent to which the team roles identified using the GMM procedure correspond to human judgments.

6.1 Setting The Number of Roles

To extract the roles that members take on at different levels, the first step was to determine the optimal number of clusters which would fit the data well and be interpretable. To this end, we used the Bayesian Information Criteria (BIC) to identify the numbers of clusters that fit the data well.

From a purely statistical sense, BIC is a criterion for model selection among a finite set of models based on the likelihood function. The model with the lowest BIC value fits the data best. We ran the GMM based clustering algorithm by varying the number of components (M) from 1 to the number of features to get clusters and corresponding BIC scores. We then qualitatively examined social role candidates from models with low BIC scores in terms of interpretability. For individual roles, we set the candidate number of roles to be 3 and 4. At the topical level, we set the candidate number of roles at 7, 8 and 9. At the team level, we found that model with $M = 7, 8, 9$ components fit the data well.

Although BIC score can be used to find the number of clusters that fit the data statistically, the interpretability of the clusters must also be considered in determining the appropriate number of clusters to extract. To interpret the roles, we examined representative features that differentiated a cluster from other clusters derived from a model, defined as features that were either 0.5 standard deviations above or below the feature's center. In order to make social roles interpretation easier, we selected out the models that did not include the same representative feature in two different clusters. The result of this interpretability analysis led to $M = 3$ as the cluster number for individual roles and $M = 7$ as the best fit for topical roles.

At the team level, we conducted a more detailed qualitative validation for identifying the appropriate number of social roles. We used a semi-structured interview with three graduate students with a research background in human computer interaction. We first introduce them to the Kiva.org platform and the definition of social roles, with one social role as an example. Then we showed them the social roles descriptions and three sample messages for each role taken from a team member whose behaviors was closest to the cluster center. In the interview, three students observed that the eight-role solution was easiest to understand. They provided labels and interpretations for each role. They produced consistent labels for six of the eight role clusters, but the labels they applied to the remaining two roles did not overlap. This interview validates that $M = 8$ is a reasonable number of team social roles. Then we used additional validation methods to generate social role labels, which is discussed in Section 6.4. After setting the number of social roles, we discuss below the meaning of the derived individual, topical and team roles.

6.2 Individual Roles

Individual roles are based on the six generic behaviors that all users can perform on the Kiva platform. Table 3 lists the three individual roles our methods derived – "active contributors", "early birds", and "lurkers", descriptions of their representative behaviors and the percentage of users who occupy each role. Since GMM assigns each user a probability on all three roles, we consider that a user occupies a role if he or she has a probability higher than 0.33 for that role. Compared to other roles, active contributors made more loans, participated over a longer period time, loaned repeatedly and joined more teams. "Early birds" made loans that were not affected by the loans' deadlines, contributing well before loans' expiration dates. In contrast to the other roles, lurkers were less involved in the community. They joined fewer teams and made fewer loans.

6.3 Topic Roles

Kiva members who made loans, i.e., the "active contributors" and "early birds", could lend loans to borrowers in different categories, while "lurkers" never lent. We found 403,984 users(80.35% of all users) lent at least one loan. Among these users, we identified topical roles based on their lending preferences. Here, we use 15 loan categories, 7 borrower locations defined as continents and the evenness of lenders' loans across categories and location (i.e., entropy) as the features to be clustered. The GMM extracted seven topic roles. Table 4 shows topic roles with their loan categories and locations, as well as the frequency of each role. As seen in Table 4, loan categories

Individual Roles	Representative behavior	Freq(%)
Active Contributor	lent more loans, lent repeatedly to the same loan multiple times, joined more teams, long participation time, from both US and non-US	47.4
Early Bird	lent when the loans were far from their expiration date	33.0
Lurker	no lending, joined few teams, short participation time	19.6

Table 3. Individual roles, their representative behavior and their frequency in the platform.

were more important than locations in shaping lenders' preferences. One point worth emphasizing is that the "Inactive Lender" in topic roles are different from "lurkers" in individual roles. "Inactive Lender" in topic roles are the lenders who are inactive in lending compared with other lenders. While, "lurkers" in individual roles are inactive not only in lending, but also inactive in joining teams or communicating with others. "Lurkers" in individual roles are not included in the "inactive Lender" in topic roles.

Topic Roles	Lending Topics	Freq(%)
Inactive Lender	Inactive in lending	40.2
Education Lender	Education loans	17.8
Health Lender	Health loans	12.9
Broad Interests Lender	Broad interests in loan categories and locations. High category entropy	12.5
Transportation Lender	Transportation loans	10.0
Oceania Lender	Borrowers from Oceania	2.5
Wholesale Lender	wholesale loans	4.1

Table 4. Topic roles, their representative behavior, and their frequency.

6.4 Team Roles

6.4.1 Generating Social Role Labels. Since team roles were based on both linguistic features and non-linguistic behaviors, it was a challenge to create role labels that were accurate as well as interpretable. After setting the social role number parameter, we used the following procedure to generate accurate and unbiased social role labels for each role cluster. We designed an experiment that asked Amazon Mechanical Turk (MTurk) workers to provide labels for each social role. In order to make crowd workers familiar with Kiva, the crowdfunding context and our labeling task, workers first completed a qualification test which included an introduction about Kiva platform, an explanation of social roles, a detailed example of developing social role labels, and a quiz question testing their understanding. Only U.S workers who passed the qualification test provided labels for the team roles. In the labeling task, crowd workers received information about a social role, including its most representative behaviors and three messages from team members closest to the cluster center. We asked the crowd workers to provide three meaningful names primarily based on the representative behaviors and then use the message examples to evaluate their names. The details about our studies are available online[1]. We received a total of 345 labels from 54 crowd workers for the eight roles. The authors then conducted a card-sorting exercise to group labels for each role, ending up with one label for each role. Table 5 shows the eight social role labels, the most representative behaviors and their frequency. We also provide a more extended description of the eight social roles and a representative message below:

Derived Roles	**Representative Behaviors**	**Frequency(%)**
Self-centered	Used more "I" words; rarely used scarcity words; sent few messages; posted short messages.	52.39
Reminder	Used more scarcity words; made more loans in the name of this team compared with other teams; used fewer modal words.	26.23
Follower	Followed others' loan recommendation; sent more links for loans; built more social connections with other group members in the discussion; posted longer messages.	5.07
Encourager	Used more words related to team goals; used more competitive words; used more modal words; posted longer messages; sent few messages with links to loans	4.07
Competitor	Used more competitive words; used more positive words; mentioned team achievements; used more negative words; used more "we" words; sent short messages; sent fewer messages	3.29
Networker	Sent many more messages; sent longer messages; built more social connections with other group members; sent more messages with links to loans; sent more messages to greet others.	3.32
Welcomer	Posted more messages to greet others; used more "we" words; used more positive words; used more modal words; mentioned goals; sent fewer links to loans.	3.09
Individualist	Donated fewer loans in the name of the team; used more scarcity words; used more modal words; used more competitive words.	2.54

Table 5. Derived team functional roles, their representative behaviors and their frequency

Self-centered user. People who wrote about their lending actions, focusing on themselves using short, brief and straight forward messages. This group of members had overlap with newcomers who also described lending behavior of themselves. For example, one "self-centered user" sent the message, "By the way, I just made a loan to an African lady... Oh well, since I am working in Africa at the moment, I'll just go out and give 25 dollars to someone! (just kidding)"

Reminder. People who reminded other team members of loan deadlines in order to persuade them to lend. They lent in the name of their team. For example, a "reminder" sent this message with the loan's link and it's deadline, "[Borrower name] from Ecuador hopes to buy 2 dairy cows to add to the farm that supports his family. The loan will expire in two days. Thanks in advance! http://www.kiva.org/loan-link "

[1] https://lusun1.github.io/Social-Roles-Kiva/

Follower. People who went along with the loans and recommendations that other team members were making. In particular, they were likely to follow others' loan recommendations. They were also communicative and good listeners. Sometimes they accepted the ideas of others and participated as audience members in group discussions and group decisions. For example, we found one follower followed other member's suggestion on one loan and sent message "@Melissa Just saw NancySomers Post and also read yours. Just made a loan to that Esmerelda Group."

Encourager. People who encouraged the whole team by comparing their team success with that of other teams. They set team goals using terms like "our goals", "team goals", and "leader board"(which represents the rank of the team). They provided suggestions to motivate the team. For example, an "encourager" sent a message, "*sigh* take a deep breath and let it out slowly. Lets get back to loving Seattle and doing our small part to help. This KIVA is such a great opportunity to learn about other people and other parts of the world. Keep up the goodwill Seattle Team!"

Competitor. Active contributors were competitive and opinionated. They seemed to like to compete with other teams as well as motivate team members to win or achieve their team's goals. For example, one "competitor" sent a message comparing his team with another one, "The difference between us and the team [team name] is that they would never lend to an atheist or agnostic group. They lend expecting to get rewarded in the afterlife. We lend because helping your fellow man is just the right thing to do. No reward expected."

Networker. People who were network hubs, frequently interacting with others. They were actively engaged as well as talkative. They sent long, detailed messages to persuade others or coordinate with them. They expressed positive attitudes and greeted team members. For example, one "networker" built connection and sent a message to another member James(name adjusted): "@James Thank you James. He is now fully raised. Wow what a day at 2 pm, he was at 30% , now with 5 team members plus a lot of anonymous individuals etc it's done! "

Welcomer. People who welcomed new members or responded to them after they first posted on Kiva. These active members also provided encouragement. For example, one "welcomer" sent a message to the team, "Hi Mac N Cheese team! I just wanted to say Hello and Welcome to our new members. As a team we have 13 loans. YAY! Nice!"

Individualist. People who lacked engagement and were inactive in the team. They focused on their personal goals instead of team goals. For example, they sent surveys unrelated to the team, asked for favors about personal tasks or advertised for things unrelated to Kiva. For example, one "individualist", who had never made a loan, sent this message to ask for help, "Dear Wales group, I know I only recently joined, but I have a favour to ask: I'm doing an MA in International Journalism at XXX University, and as an assignment for next Monday I am doing a short article about KIVA. Would any one of you be willing to answer with just a short comment about what you think of KIVA, your own experience with it, why you joined, etc.?" It is worth mentioning that "Individualists" are different from "lurkers" in individual roles and topic roles. "Lurkers" are modeled based on all registered Kiva users, while "individualist" described behaviors of users who involved in the team.

6.4.2 Validating the Team Roles. The GMM procedure clustered members' behavior and themes in their messages to identify eight team roles. Do these roles correspond to human judgments of the roles that team members occupy? Previous researchers usually conducted interviews with domain experts or researchers to see whether their opinions were aligned with social roles labels [56, 58]. However, this evaluation method is hard to quantify and scale. To overcome these problems, we developed an evaluation method that could be scaled up for large user studies. To measure the accuracy of social roles, we provided Mechanical Turk workers with samples of team members' behavior and asked them to select the role that was the *worst* fit to the members' behavior as an "intruder" [12, 32]. We asked crowd workers to select the intruder role rather than the most appropriate role because the team members could occupy multiple roles simultaneously, making it difficult for identify the role that best fits the person's behavior. In particular, 14.79% of members occupied more than one role in the team while 85.21% of them occupied only one. We classified users as occupying a role if the predicted probability of their being in the role was greater than 0.125).

In this evaluation exercise, we selected representative team members with three messages, and then selected three role labels based on users' probability on each role. Specifically, in each component, we first calculated the distance between each user's representation and the center of that component. We identified 10 representative team members for each role who were the closest to the component center, comprising a sample of 80 team members. For each member, we calculated their probability distribution among the eight social roles and then selected the role label with the highest probability (top role), the role label with the lowest probability (bottom role), which we call the intruder, and a middle role label whose probability laid between the probability for the top and bottom roles. To evaluate the fit of the automatically generated role labels with human judgments, we provided crowd workers a description of a user's behaviors, three messages sent by the user and three potential role labels for the user (the top role, middle role and intruder role). Each team member is represented as eight probabilities, reflecting the extent to which he or she occupied each social role. For example, a user's top role might be a "networker", bottom role might be "individualist" and the middle role, corresponding to the median of the eight probabilities, might be "reminder".

Crowd workers first read this user's description:

> This user sent many more messages to the discussion forum than the average. This user's messages were much longer than the average. This user built many more social connections with other group members in the discussion. For example, this user replied to another specific member by "@" another member. This user also sent more links of loans to provide the loan's information in the group discussion.

Then, the crowd workers were shown three sample messages from the user and the definition of the "Networker", "Individualist" and "Reminder" roles. The details about our studies are here[2]. Crowd workers ranked the three roles presented for this user from relevant to irrelevant. Our evaluation metrics were the percentage of time that crowd workers ranked the role with the highest probability as the most relevant and the intruder role label as the most irrelevant. Chance is 33%. For each social role, we collected data from 30 crowd workers. Results are shown in Fig 2. The blue bars refer to the percentage of crowd workers who correctly identified the social role with the highest probability as the most relevant role label. The accuracy for all roles was greater than the chance rate of 33%, with the mean accuracy of assigning the most relevant tag to the role with the highest probability being 53.1%. The orange bars indicate the percentage of crowd workers who correctly rejected the intruder role as the appropriate role description. All values were all higher than 60%, with the mean accuracy of correctly rejecting the intruder role being 79.3%, well above chance. However, crowd workers' agreement identifying "individualist" and "self-centered lenders" was less than 60%. One reason might be that these two roles are not easy to identified based on their verbal behaviors, the type of behavior that crowd workers relied upon most. "Self-centered people" sent few messages, and the most typical behavior of "individualist" is that they donated less.

[2]https://lusun1.github.io/Social-Roles-Kiva/

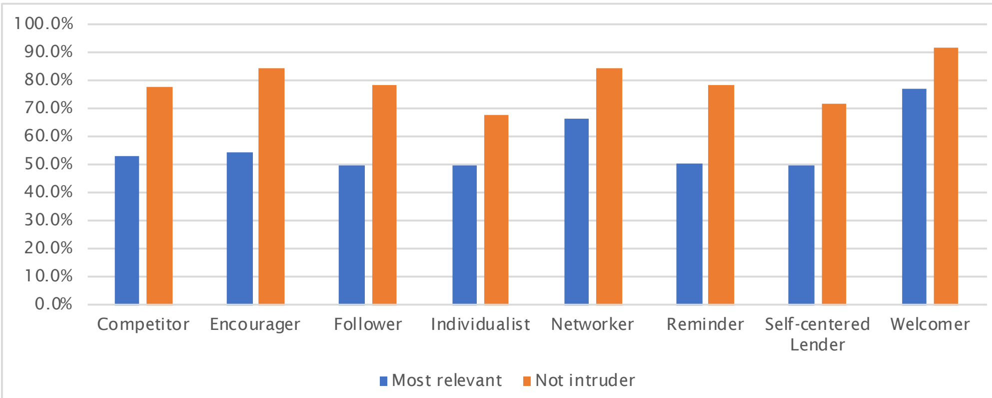

Fig. 2. Intruder test on user role evaluation. X axis shows eight social roles. Y axis represents the percentage of the crowdworkers correctly identified the social role as the most relevant role(blue) and as not intruder(orange). Note that crowd workers' agreement on the "not intruder" of each role is higher than 60%

After the team role validation, we examined the mixture of team roles. Since GMM allows team members to occupy multiple roles simultaneously, we checked the overlap of the eight social roles. In particular, 14.79% of the members occupied more than one role in the team. Among users who played more than one role, the "self-centered lender" and "individualist" roles overlapped the most, representing 20.1% of all overlaps. In addition, among users who played more than one role, 2.8% of users played the "networker" and "welcomer" roles at the same time. These common overlaps probably indicate that "self-centered lenders" and "individualist" roles may be defined by some similar behaviors. Similarly, "networker" and "welcomer" roles may also use similar behaviors to build social connections to newcomers.

7 INFLUENCE OF TEAM ROLES

Increasing users' contributions is a key challenge for many online communities, not only in Kiva but also in other micro-funding platforms, since members' contributions explicitly represent the success of the community[13]. While some field experiments have investigated different mechanisms to improve team contribution, such as sending goal-setting messages to motivate inactive teams members, we know of no research that has explored how users' social roles influence team contribution in peer-to-peer lending. The goal of this section is to measure the extent to which the functional roles present in teams predict the amount teams contribute. In addition to leading to a better understanding of online contribution, this analysis will illustrate the utility of each team role. We first explored the correlations among team loan amount, control variables, and social roles and then ran regression analyses to identify the extent to which the distribution of social roles in a team predicted the team's lending. Results show although team size and its overall communication activity were the strongest predictor of the amount it lends, the distribution of roles in a team improved the ability to team loans by a small but reliable amount. Teams with a more even distribution of roles and those with more members playing competitor, encourager, and reminder roles lent more, while those with more members playing the follower and individualist roles lent less.

Variables. The dependent variable for this analysis is the total amount of money loaned by the team as of May 2018. We constructed control variables based on research by Hartley [24]. The most basic variable associated with team success is a team's human capital. To reflect this, we included *team member count* (number of team members) as a control variable. The number of messages posted in the team is an important proxy for a team activity, so here we measured the *message count* as a control variable. Finally, we include whether the team is "open" or "closed", indicating the ease with which members can join. Corporate teams that only allow employees to join are "closed." Here, we consider the *team membership type* as a binary variable: "open" as 1 or "closed" as 0. When we run the analysis, we logged and standarized the continuous variables which have skewed distribution.

Features	**Min**	**Max**	**Mean**	**Median**	**Sd**
Self-centered Lenders	0.000	1.000	0.558	0.556	0.420
Networker	0.000	1.000	0.027	0.000	0.113
Competitor	0.000	1.000	0.019	0.000	0.114
Individualist	0.000	1.000	0.222	0.000	0.356
Welcomer	0.000	1.000	0.068	0.000	0.219
Encourager	0.000	1.000	0.028	0.000	0.129
Reminder	0.000	1.000	0.071	0.000	0.192
Follower	0.000	1.000	0.008	0.000	0.050
Entropy of roles	0.000	2.730	0.487	0.000	0.712

Table 6. Descriptive statistics of independent variables

Independent Variables. The theoretically important variables is the extent to which each of the eight social roles derived from GMM are represented in the team. To operationalize this, we represented each team as a vector of eight with the values being the average role probability across its members. Moreover, we measured the evenness of eight roles in each team using the entropy of roles probabilities of each team, with a higher entropy indicating that the distribution of roles in the team was more even (i.e., the team probably had some people representing each type of role). We logged and standardized the continuous variables which have skewed distribution (except the binary variable group type) before including them in the regression models. The descriptive statistics of the independent variables before standardization are described in Table 6.

Analysis. We modeled 28,535 members' team roles in 5,819 teams using the user–team corpus. We built a baseline regression model including just control variables and a second model including social roles to test their predictive ability. To check for multicolinearity, we calculated the correlations among all independent variables (see Figure 3). We also calculated the variance inflation factor (VIF) score of variables in the regression model. Because the VIF for "self-centered lender" was very high (9.33e+14), we eliminated it for the social role model, leaving seven roles to predict a team's loan amount. The VIF scores for the remaining variables were all less than 2.5.

7.1 Results

Table 7 shows the results of the linear regression models, with a baseline model containing only control variables and a social role model that includes the seven social roles (omitting the "self-centered" role) and the entropy of the team on social roles distribution. The table reports the regression coefficients and the p-value for each variable showing whether the coefficient is reliably different from zero. Because all variables were standardized to have a mean of zero and standard deviation of one, a coefficient indicates the change in the amount the team lent (in standard

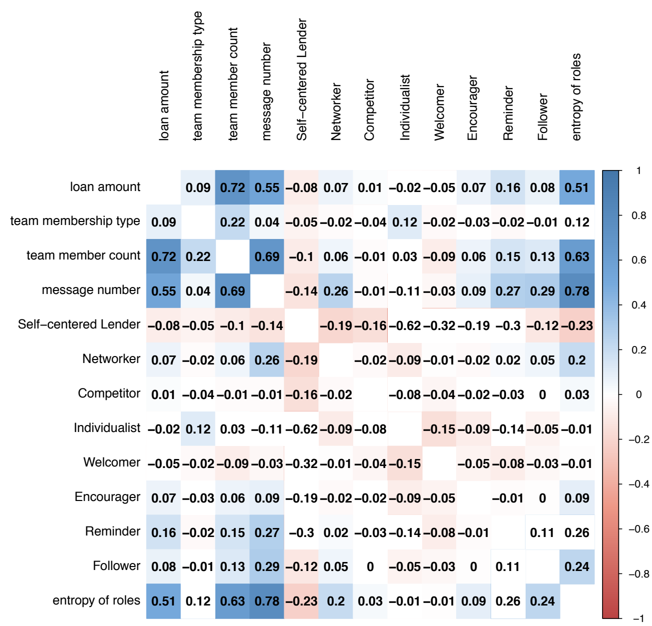

Fig. 3. Correlation among all variables. Cells with a colored background are significant at p < .01.. The color in the cell indicates the correlation coefficient.

deviation units) when a predictor increased by one standard deviation. The baseline model explains approximately 52.7% of the variance in the amount teams give. Team size (i.e., member count) and activity (i.e., message number) are strong positive predictors of team contributions. The coefficient for team type shows that open teams loaned less than closed ones.

An ANOVA test showed that the social role model is a statistically significant better fit to the data than the baseline model (p–value < .001) [11]. The adjusted R-squares of the baseline model and the social role model show that adding the social role related variables improved model fit, boosting the adjusted R-square from .527 to .532. The coefficients indicate that teams with more members who performed the reminder role (0.415) and those with more of the competitor role (0.400) loaned more money (p < 0.01). The welcomer (0.161) role and the encourager (0.270) role also have positive association with team lending. On the other hand, teams with more members enacting the follower role (-2.113) and the individualist (-0.160) role loaned less. Finally, teams with greater role entropy (i.e., more even distribution of the team roles among members) lent more money. This finding suggests that when several major social roles co-existed in the teams, they lent more.

These results are consistent with past research in conventional work groups [5, 8], that the presence in groups of both task roles and group building and maintenance roles is associated with team success. They function by coordinating and motivating group effort, strengthening group connection and maintaining the group way of working. Encouragers stimulate group lending through goal setting. Also, explicitly highlighting group goals can strongly motivate group members to accomplish group goals, especially for members who have voluntarily identified with the group[60]. Competitors also stimulate lending using competitive words, warnings and team achievement to stimulate the team members to focus on the lending task. In addition, competitors' explicit mention

Model	Baseline Model	Social Role Model
(Intercept)	4.361	4.549
Team member count	1.383***	1.366***
Message count	0.198***	0.113***
Open team membership type	- 0.502***	- 0.499***
Networker		0.061
Competitor		0.400*
Individualist		- 0.160*
Welcomer		0.161
Encourager		0.270
Reminder		0.415**
Follower		-2.113***
Entropy of roles		0.169***
Number of teams	5819	5819
Adjusted R-squared	0.527	0.532
Improvement in Adj. R-square		0.005***

Table 7. Team Contribution Prediction Performance of Regression Model. To compare the goodness-of-fit of baseline model and social role model, we used the ANOVA test to compare the Adj. R-square of each model. Note that including social role variables and the entropy of social roles significantly improved the fit of the regression model. P-value significance codes: $< .0001 : ***, < .001, **, < .01 : *, < 0.05 : .$

of opposing teams (i.e., out-groups) can cause team members to identify more strongly with their team (i.e., the in-group [25]) and its goals [60]. From a leadership perspective, people who occupy the reminder, encourager and competitor roles exhibit task-focus leadership behaviors, which increases contribution to a focal task[61].

The networker and welcomer roles are group maintenance roles that probably have their impact on lending by increasing members' connection to the group. Networkers used communication volume, sending long messages and replying to specific other members, to build strong ties in the community. In Kiva they took the advantage of social networks to persuade others to lend, greet to new comers, reply questions, coordinate subgroup users and share ideas with other users. Their behaviors may help the community maintain close social relationships, support group cohesion, and further benefit community-building [8, 61]. Welcomers also actively built social connections to new team members and provided them a sense of belonging to the group. Yang et al. [58] demonstrated that the presence of "newcomer welcomers" in a cancer support group was associated with members staying longer at the platform. In Kiva, these socially positive roles could indicate that the role occupants have already attached them to others or groups as a whole or could have caused non-role occupants to feel more attachment to the group.

In contrast to the group task roles and group maintenance roles, teams with members who occupied "individual roles" focusing on satisfying their personal needs rather than the group task, are typically less successful [5]. For example, in Kiva, they used the group discussion as the platform to send out ads, rather than to advocate for borrowers. Not only were they unlikely to loan themselves, but their selfish behaviors may weaken the cohesion of their teams and demotivate other users. Besides, teams with more members occupying the follower role lent less. Followers tended to passively follow other users' recommendation of loans in the discussion forum instead of actively searching for loans on the platform. One might think that their copycat behavior would be associated with team success, but this is not the case. Even though they were making loans, perhaps because they are not engaged in advocacy or team building activities, they did no lead other members to lend. For example, we observed that some followers responded to others' loans recommendation by sending some messages to show their worries about these loans (see message example in Session 6.4). They may have shown conflicts between team members. Even though they send some loan links to the group and are socially engaged with other users, they were less likely to advocate for new loans and may have impeded the team contribution.

8 CONCLUSION AND DISCUSSION

This paper investigated how to measure social roles in a micro-lending platform in multiple levels within a single online community and then explored how team roles predicted team success. Specifically, we first introduced a social role modeling methodology to discover emergent social roles in three-levels: individual roles, topic roles, and team functional roles. These levels cover the entire spectrum of users in Kiva, from users who concentrate on lending individually to users who are core to a lending team. And then we used Gaussian Mixture Modeling to cluster the behaviors at each level. After quantitative and qualitative evaluation, we discovered three individual roles, seven topic roles, and eight team roles. This approach of identifying roles across multiple levels of analysis provides a holistic account of the roles that a single person might play in an online community. For example, in Kiva compared to the user Ann who might be a "Lurker", rarely making a loan, the user John might be classified as an "Active Contributor" for his multiple loans. Because he concentrates those loans in the health area, we can classify him as a "Health Lender". Finally, John is a member of the "Crazy Canucks" and "Nerdfighters" teams. He performed as an "networker" in the former team where he tried to connect with other team members and "encourager" in the latter team where he motivated the team to lend. This approach increases scope and depth of social role modeling compared to one that treats Kiva as a whole as a unitary level and assumes that people occupy only a single role.

Compared with prior research on social roles, which typically used interviews or qualitative analysis to validate social roles, we also developed a scalable validation approach in which human judges identify an "intruder" role label (i.e., the label that least applies to a representative role occupant). First, to generate accurate and interpretable labels for team roles, we provided multiple crowd workers with representative behaviors and messages associated with a role and asked them to generate role labels. Then the researchers used a card sorting process to group labels into eight team roles. Moreover, we designed an intruder validation experiment to validate the accuracy of human labels and eliminate bias. In the experiment, we provided crowd workers with a description of three representative role occupants and three role labels, one for the top role for the user, one for the bottom (intruder) role and one for a middle role. Considering that team members could occupy several roles simultaneously, we asked the crowd workers to select the intruder role–the one which was the worst fit to the user's description. We used the validation results to justify whether team roles are reasonable. This systematic approach of validation could apply to other unsupervised approaches to identify social roles in online communities.

Finally, we conducted an analysis to determine whether the role composition of teams predicted team success–the amount of money the team lend. We examined the function of team roles with the improvement of team contribution. Encouraging more members to occupy "competitor" and "reminder" or eliminating members to perform as "follower" and "individualist" could improve the team loan amount.

8.1 Implications

This research makes both theoretical and practical contributions. At the theoretical level, our goal was to build on previous role theories and taxonomies of task-oriented and group maintenance roles and associated role-related behaviors and then use unsupervised statistical clustering techniques to make the vague term of "social roles" more precise. Prior social science research on functional roles in groups does not provide detailed descriptions of roles' actions and their functions, so it is difficult for researchers to agree on which roles are actually present in a group and how these roles operate. Compared with previous research, our work defines roles with users' task actions and social interactions. The inductive, computational role identification method used in this research, coupled with our role-label generation and validation methods, found eight team social roles in Kiva. Team roles correspond to roles identified in prior research on functional roles in offline work groups [5], such as "encourager", "follower", "networker" and "individualist". In particular, the "networker" role in Kiva is similar to the "social networkers" in Wikipedia, who built strong ties with other editors through discussion [53] . In addition, the "expediter" and the "encourager" roles identified in the previous theory and empirical research in off-line settings by Benne and Sheats [5] correspond to the "encourager" role in Kiva, who mentioned more modal ("should") words and words related to team goals encouraging other team members to give. The "follower" role in Kiva is similar to followers described by Benne and Sheats [5] in off-line groups, who go along with the movement of the team and more or less passively accept ideas from other team members. In our context, followers lent based on other people's recommendations. However, in contrast to the followers described by Benne and Sheats [5] who are audience in the team, followers in Kiva built social connections with other members and doesn't merely serve as an audience. Similarly, Kiva's "individualist" and "self centered lender" roles correspond to Benne and Sheats' individual roles. One aspect of individual roles in off-line groups is that they focus on personal goals rather than the group task or to the functioning of the group. "Individualists" in Kiva do not lend in the name of the team and those occupying the "self-centered" role used "I" words and rarely communicated with other group members. Besides, the "welcomer" role we identified in Kiva is very similar to the same "welcomer" role in health care online community [58], where users are likely to talk to members of the community and show their warmth.

In terms of methods, we developed a method to model social roles in multi-levels in order to understand users more comprehensively, which involves defining the role levels, proposing role-relevant behaviors, automated processing roles and evaluating social roles. Specifically, in contrast to most computational role-modeling research, we explored a method to quantify the social role evaluation process. Previous research only evaluated roles in terms of such quantitative measures like consistency between training and hold-out samples or interviewing experts to justify the labels. While we designed a set of human annotation tasks based on an "intruder" intuition [12], which largely quantifies the social role evaluation process and avoid the expert blind spot.

At the practical level, we foresee our social role results can be applied in several ways. The social roles in three levels could be useful in downstream interventions such as recommender systems, which guide users to appropriate loans or teams. For example, the recommender system could recommend loans to early birds when the loans are first announced. It could recommend networkers join teams that are not currently cohesive and could use help getting teammates to socialize with others. Team administrators could use our functional roles and insights from our team composition analyses to diagnose teams and provide feedback for teams that are less likely to succeed. Of course, these practical implications must be tempered by the weak, although reliable, associations between role composition and team success.

8.2 Limitation

Our research also has a set of limitations. First, our operationalizations of features used to define roles, although based on social science theory and qualitative empirical description of users behavior supplemented by affinity diagrams, still require more carefully constructed features to capture other aspects of user behaviors in this peer-to-peer lending context. Second, we utilized a priori heuristics to capture specific types of behaviors, such as the number of "welcome" words. Future work could improve feature construction by systematically annotating messages and then training machine learning models to more accurately capture such behaviors. A richer set of behaviors types, including negotiating and question asking, could also lead to better derivations of social role identification.

Moreover, since our research uses unsupervised clustering procedures to capture behavioral regularities, we are not able to determine whether roles derived from this procedure align well with role holders' perception of the roles they occupy. Although the concept of social role is useful for simplifying the analysis of a social setting composed of many discrete behaviors and for drawing analogies between social setting, it is not clear whether the role concept is empirically necessary. For example, it might be possible to more accurately predict the amount of team loans based on the discrete behaviors that define roles rather than the roles themselves.

Last but not least, our regression models of the influence of team roles are correlational, and do not provide evidence on whether roles cause behaviors rather than simply reflect them. Without rigorous random assignment experiments, we could not guarantee that recruiting/training people for particular roles can lead to group success.

9 ACKNOWLEDGMENTS

This work was supported by National Science Foundation IIS-1628319: Developing, Testing, and Designing from a Computational Theory of Online Communities. Diyi Yang was supported by a Facebook Fellowship. We thank Akanksha Kartik for help with programming and statistical analysis. The authors would like to thank the reviewers for their helpful feedback. The authors thank all co-workers' helpful feedbacks on the paper.

REFERENCES

[1] Wei Ai, Roy Chen, Yan Chen, Qiaozhu Mei, and Webb Phillips. 2016. Recommending teams promotes prosocial lending in online microfinance. *Proceedings of the National Academy of Sciences* 113, 52 (2016), 14944–14948.

[2] Ofer Arazy, Johannes Daxenberger, Hila Lifshitz-Assaf, Oded Nov, and Iryna Gurevych. 2016. Turbulent stability of emergent roles: The dualistic nature of self-organizing knowledge coproduction. *Information Systems Research* 27, 4 (2016), 792–812.

[3] Ofer Arazy, Oded Nov, and Felipe Ortega. 2014. The [Wikipedia] World is Not Flat: on the organizational structure of online production communities. (2014).

[4] Ofer Arazy, Felipe Ortega, Oded Nov, Lisa Yeo, and Adam Balila. 2015. Functional roles and career paths in Wikipedia. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. ACM, 1092–1105.

[5] Kenneth D Benne and Paul Sheats. 1948. Functional roles of group members. *Journal of social issues* 4, 2 (1948), 41–49.

[6] Bruce J Biddle. 1986. Recent developments in role theory. *Annual review of sociology* (1986), 67–92.

[7] Bruce J Biddle. 2013. *Role theory: Expectations, identities, and behaviors.* Academic Press.

[8] C. Shawn Burke, Kevin C Stagl, Cameron Klein, Gerald F Goodwin, Eduardo Salas, and Stanley M Halpin. 2006. What type of leadership behaviors are functional in teams? A meta-analysis. *The Leadership Quarterly* 17, 3 (2006), 288–307.

[9] C Shawn Burke, Kevin C Stagl, Cameron Klein, Gerald F Goodwin, Eduardo Salas, and Stanley M Halpin. 2006. What type of leadership behaviors are functional in teams? A meta-analysis. *The leadership quarterly* 17, 3 (2006), 288–307.

[10] Meeyoung Cha, Hamed Haddadi, Fabricio Benevenuto, P Krishna Gummadi, et al. 2010. Measuring user influence in twitter: The million follower fallacy. *Icwsm* 10, 10-17 (2010), 30.

[11] John M Chambers, Trevor J Hastie, et al. 1992. *Statistical models in S.* Vol. 251. Wadsworth & Brooks/Cole Advanced Books & Software Pacific Grove, CA.

[12] Jonathan Chang, Sean Gerrish, Chong Wang, Jordan L Boyd-Graber, and David M Blei. 2009. Reading tea leaves: How humans interpret topic models. In *Advances in neural information processing systems*. 288–296.

[13] Roy Chen, Yan Chen, Yang Liu, and Qiaozhu Mei. 2017. Does team competition increase pro-social lending? Evidence from online microfinance. *Games and Economic Behavior* 101 (2017), 311–333.

[14] Robert B Cialdini and Robert B Cialdini. 2007. *Influence: The psychology of persuasion*. Collins New York.

[15] Uldarico Rex Dumdum, Kevin B Lowe, and Bruce J Avolio. 2013. *A meta-analysis of transformational and transactional leadership correlates of effectiveness and satisfaction: An update and extension*. Emerald Group Publishing Limited, 39–70.

[16] Alice H Eagly, Mary C Johannesen-Schmidt, and Marloes L Van Engen. 2003. Transformational, transactional, and laissez-faire leadership styles: a meta-analysis comparing women and men. *Psychological bulletin* 129, 4 (2003), 569.

[17] Alberto Espinosa, Robert Kraut, Sandra Slaughter, Javier Lerch, James Herbsleb, and Audris Mockus. 2002. Shared mental models, familiarity, and coordination: A multi-method study of distributed software teams. *ICIS 2002 Proceedings* (2002), 39.

[18] Ethan Fast, Binbin Chen, and Michael S Bernstein. 2016. Empath: Understanding topic signals in large-scale text. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 4647–4657.

[19] Danyel Fisher, Marc Smith, and Howard T Welser. 2006. You are who you talk to: Detecting roles in usenet newsgroups. In *System Sciences, 2006. HICSS'06. Proceedings of the 39th Annual Hawaii International Conference on*, Vol. 3. IEEE, 59b–59b.

[20] Andrea Forte, Vanesa Larco, and Amy Bruckman. 2009. Decentralization in Wikipedia governance. *Journal of Management Information Systems* 26, 1 (2009), 49–72.

[21] Eric Gilbert and Karrie Karahalios. 2010. Widespread worry and the stock market. In *Fourth International AAAI Conference on Weblogs and Social Media*.

[22] Eric Gleave, Howard T Welser, Thomas M Lento, and Marc A Smith. 2009. A conceptual and operational definition of' social role in online community. In *System Sciences, 2009. HICSS'09. 42nd Hawaii International Conference on*. IEEE, 1–11.

[23] Jeffrey T Hancock, Christopher Landrigan, and Courtney Silver. 2007. Expressing emotion in text-based communication. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 929–932.

[24] Scott Hartley. 2010. Kiva. org: Crowd-sourced microfinance and cooperation in group lending. (2010).

[25] Matthew J Hornsey. 2008. Social identity theory and selfâĂŘcategorization theory: A historical review. *Social and Personality Psychology Compass* 2, 1 (2008), 204–222.

[26] Christina Jenq, Jessica Pan, and Walter Theseira. 2015. Beauty, weight, and skin color in charitable giving. *Journal of Economic Behavior & Organization* 119 (2015), 234–253.

[27] Daniel Katz and Robert Kahn. 1978. *The social psychology of organizations*. Wiley, New York. TY - BOOK.

[28] Aniket Kittur and Robert E Kraut. 2010. Beyond Wikipedia: coordination and conflict in online production groups. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*. ACM, 215–224.

[29] Steve WJ Kozlowski and Katherine J Klein. 2000. A multilevel approach to theory and research in organizations: Contextual, temporal, and emergent processes. (2000).

[30] Michel Krieger, Emily Margarete Stark, and Scott R Klemmer. 2009. Coordinating tasks on the commons: designing for personal goals, expertise and serendipity. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1485–1494.

[31] Laura Larrimore, Li Jiang, Jeff Larrimore, David Markowitz, and Scott Gorski. 2011. Peer to peer lending: The relationship between language features, trustworthiness, and persuasion success. *Journal of Applied Communication Research* 39, 1 (2011), 19–37.

[32] Jey Han Lau, David Newman, Sarvnaz Karimi, and Timothy Baldwin. 2010. Best topic word selection for topic labelling. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*. Association for Computational Linguistics, 605–613.

[33] Jun Liu and Sudha Ram. 2011. Who does what: Collaboration patterns in the wikipedia and their impact on article quality. *ACM Transactions on Management Information Systems (TMIS)* 2, 2 (2011), 11.

[34] Yang Liu, Roy Chen, Yan Chen, Qiaozhu Mei, and Suzy Salib. 2012. I loan because...: Understanding motivations for pro-social lending. In *Proceedings of the fifth ACM international conference on Web search and data mining*. ACM, 503–512.

[35] ALEXANDRA Mateescu. 2015. Peer-to-Peer lending. *Data & Society Research Institute* (2015), 19–25.

[36] J. McGrath. 1984. *Groups: Interaction and performance*. Prentice Hall, Englewood Cliffs, NJ. TY - BOOK primary.

[37] Geoffrey J McLachlan and Kaye E Basford. 1988. *Mixture models: Inference and applications to clustering*. Vol. 84. Marcel Dekker.

[38] Tanushree Mitra and Eric Gilbert. 2014. The language that gets people to give: Phrases that predict success on kickstarter. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*.

ACM, 49–61.
[39] Melina Moleskis and Miguel Ángel Canela. 2016. Crowdfunding success: The case of Kiva. org. (2016).
[40] Michael Muller, Mary Keough, John Wafer, Werner Geyer, Alberto Alvarez Saez, David Leip, and Cara Viktorov. 2016. Social ties in organizational crowdfunding: benefits of team-authored proposals. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. ACM, 1246–1259.
[41] T. V. Mumford, M. A. Campion, and F. P. Morgeson. 2006. *Situational judgment in work teams: A team role typology.* Erlbaum, Mahwah, NJ: Erlbaum., 319âĂŞ343.
[42] Blair Nonnecke and Jenny Preece. 2000. Lurker demographics: Counting the silent. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. ACM, 73–80.
[43] James W Pennebaker, Martha E Francis, and Roger J Booth. 2001. Linguistic inquiry and word count: LIWC 2001. *Mahway: Lawrence Erlbaum Associates* 71, 2001 (2001), 2001.
[44] Philip M Podsakoff, Scott B MacKenzie, Julie Beth Paine, and Daniel G Bachrach. 2000. Organizational citizenship behaviors: A critical review of the theoretical and empirical literature and suggestions for future research. *Journal of Management* 26, 3 (2000), 513–563.
[45] Jennifer Preece and Ben Shneiderman. 2009. The reader-to-leader framework: Motivating technology-mediated social participation. *AIS transactions on human-computer interaction* 1, 1 (2009), 13–32.
[46] Alan G Sanfey, James K Rilling, Jessica A Aronson, Leigh E Nystrom, and Jonathan D Cohen. 2003. The neural basis of economic decision-making in the ultimatum game. *Science* 300, 5626 (2003), 1755–1758.
[47] Greg L Stewart, Charles C Manz, and H. P. Sims. 1999. *Team work and group dynamics*. Wiley, New York.
[48] Besiki Stvilia, Michael B Twidale, Linda C Smith, and Les Gasser. 2008. Information quality work organization in Wikipedia. *Journal of the American society for information science and technology* 59, 6 (2008), 983–1001.
[49] Na Sun, Patrick Pei-Luen Rau, and Liang Ma. 2014. Understanding lurkers in online communities: A literature review. *Computers in Human Behavior* 38 (2014), 110–117.
[50] Henri Tajfel. 1974. Social identity and intergroup behaviour. *Information (International Social Science Council)* 13, 2 (1974), 65–93.
[51] Thanh Tran, Madhavi R Dontham, Jinwook Chung, and Kyumin Lee. 2016. How to succeed in crowdfunding: a long-term study in kickstarter. *arXiv preprint arXiv:1607.06839* (2016).
[52] Linn Van Dyne and Jeffrey A LePine. 1998. Helping and voice extra-role behaviors: Evidence of construct and predictive validity. *Academy of Management Journal* 41, 1 (1998), 108–119.
[53] Howard T Welser, Dan Cosley, Gueorgi Kossinets, Austin Lin, Fedor Dokshin, Geri Gay, and Marc Smith. 2011. Finding social roles in Wikipedia. In *Proceedings of the 2011 iConference*. ACM, 122–129.
[54] Howard T Welser, Eric Gleave, Danyel Fisher, and Marc Smith. 2007. Visualizing the signatures of social roles in online discussion groups. *Journal of social structure* 8, 2 (2007), 1–32.
[55] Diyi Yang. 2018. *Computational Social Roles: Identify, Recommend and Configure Emergent Social Roles in Online Communities*. Ph.D. Dissertation. Stanford University.
[56] Diyi Yang, Aaron Halfaker, Robert E Kraut, and Eduard H Hovy. 2016. Who Did What: Editor Role Identification in Wikipedia.. In *ICWSM*. 446–455.
[57] Diyi Yang and Robert E Kraut. 2017. Persuading teammates to give: Systematic versus heuristic cues for soliciting loans. *Proceedings of the ACM on Human-Computer Interaction* 1 (2017), 114.
[58] Diyi Yang, Robert E. Kraut, Tenbroeck Smith, Elijah Mayfield, and Dan Juafsky. 2019. *Seekers, providers, welcomers, and storytellers: Modeling social roles in online health communities*. ACM Press, New York, NY.
[59] Bowen Yu, Yuqing Ren, Loren Terveen, and Haiyi Zhu. 2017. Predicting member productivity and withdrawal from pre-joining attachments in online production groups. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. ACM, 1775–1784.
[60] Haiyi Zhu, Robert Kraut, and Aniket Kittur. 2012. Organizing without formal organization: group identification, goal setting and social modeling in directing online production. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*. ACM, 935–944.
[61] Haiyi Zhu, Robert E Kraut, and Aniket Kittur. 2013. Effectiveness of shared leadership in Wikipedia. *Human factors* 55, 6 (2013), 1021–1043.

Received April 2019; revised June 2019; accepted August 2019