

# To Gather Together for a Better World: Understanding and Leveraging Communities in Micro-lending Recommendation

Jaegul Choo  
Georgia Institute of  
Technology  
jaegul.choo@cc.gatech.edu

Daniel Lee  
Georgia Institute of  
Technology  
dlee@cc.gatech.edu

Bistra Dilkina  
Georgia Institute of  
Technology  
bdilkina@cc.gatech.edu

Hongyuan Zha  
Georgia Institute of  
Technology  
zha@cc.gatech.edu

Haesun Park  
Georgia Institute of  
Technology  
hpark@cc.gatech.edu

## ABSTRACT

Micro-finance organizations provide non-profit lending opportunities to mitigate poverty by financially supporting impoverished, yet skilled entrepreneurs who are in desperate need of an institution that lends to them. In Kiva.org, a widely-used crowd-funded micro-financial service, a vast amount of micro-financial activities are done by lending teams, and thus, understanding their diverse characteristics is crucial in maintaining a healthy micro-finance ecosystem. As the first step for this goal, we model different lending teams by using a maximum-entropy distribution approach based on a wealthy set of heterogeneous information regarding micro-financial transactions available at Kiva. Based on this approach, we achieved a competitive performance of 0.84 AUC value in predicting the lending activities for the top 200 teams. Furthermore, we provide deep insight about the characteristics of lending teams by analyzing the resulting team-specific lending models. We found that lending teams are generally more careful in selecting loans by a loan's geo-location, a borrower's gender, a field partner's reliability, etc., when compared to lenders without team affiliations. In addition, we identified interesting lending behaviors of different lending teams based on lenders' background and interest such as their ethnic, religious, linguistic, educational, regional, and occupational aspects. Finally, using our proposed model, we tackled a novel problem of lending team recommendation and showed its promising performance results.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

WWW'14, April 7th, 2014, Seoul, Korea.

Copyright 2014 ACM 978-1-4503-2329-1 ...\$15.00.

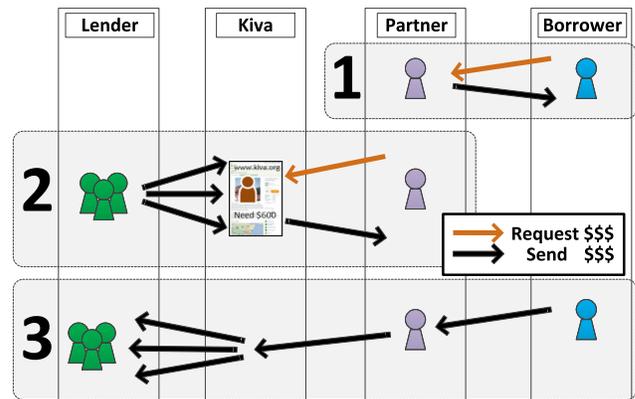


Figure 1: An overview of how Kiva works. 1. A borrower requests a loan to a field partner, and a loan is disbursed. 2. The partner posts a loan request to Kiva, and lenders choose to fund the loan, where upon fulfillment Kiva sends the funds to the field partner. 3. The borrower makes repayments to the field partner, who forwards the funds back to Kiva. Kiva then repays the lenders. The lenders receive the repayment through Kiva and are given the opportunity to recycle their loans.

## Categories and Subject Descriptors

H.2.8 [Artificial Intelligence]: Learning

## General Terms

Algorithms, Performance, Experimentation, Recommendation

## Keywords

Microfinance, Maximum entropy distribution, Heterogeneous feature, Community characteristic

## 1. INTRODUCTION

Micro-finance institutions lend credit to entrepreneurs who have no credit available to them in impoverished countries.

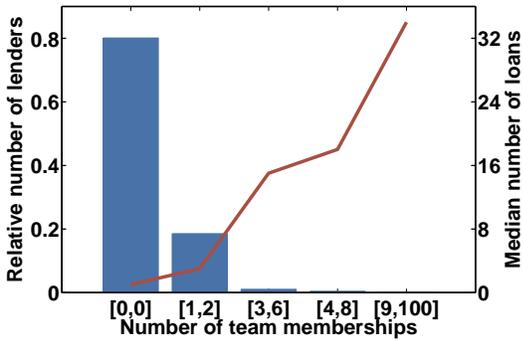


Figure 2: The degree of lending activities depending on lending team involvement.

Its concept was conceived when Muhammad Yunus discovered that the extreme poor barely had enough means with which to sustain themselves, only to use their business sales as repayment for the materials they loaned [37]. By loaning credit without collateral and interest, impoverished entrepreneurs are given the opportunity to overcome the vicious cycle of debt.

Kiva.org, the overall process of which is summarized in Fig. 1, takes the idea of micro-financing and pairs it with crowd-sourcing to provide easy online access for lending micro-credit. Kiva, a non-profit organization, does not collect interest but rather provides an intermediary service which pools together money from its lenders and forwards them to the field partner who then distributes it to the requesting borrower.

Kiva relies heavily on its transparency due to its core values for successful growth [13]. Kiva’s transparency allows open public access to its transactional and entity data which can be downloaded as daily snapshots or through their API. Kiva’s May 2013 data snapshot contained over 1,100,000 lenders, 500,000 loans, and 150,000 journal entries for over four million transactions that resulted in 400 million US dollars of loans issued. There are a variety of data types within the Kiva data including geo-spatial, temporal, categorical, numerical, and free-text unstructured data. The size of the data set, along with its massive set of heterogeneous information makes the Kiva data set a fascinating data set for data mining and social media researchers.

**Impact of lending teams.** Virtual communities thrive when their users are active in their participation, and in this case, when lenders are actively lending. Particularly with Kiva, lending is synonymous with donating due to the lack of monetary gain from lending, thus keeping lenders actively and consistently involved is a critical factor in making Kiva self-sustainable. As one such way, Kiva encourages each of their lenders to join teams, called lending teams, to allow lender collaboration in locating and funding credit requests. Lending teams are primarily formed through a common interest, where one such similarity interest group could contain lenders interested in funding a particular type of business.

The Kiva data reveals that lending teams play a major role in the level of participation for lenders. As seen in Fig. 2, the median number of loans per lender increases quickly with the number of teams the lender is part of (the red line). However, about 80% Kiva lenders are still not affiliated with any lending teams, while most of the remaining 20% of lenders participate in only one or two lending teams.

Overall, lenders affiliated with at least one lending team fulfill about 50% of the total loan activities in Kiva. These statistics suggest that matching lenders with teams can be a key driver for further increasing participation.

## Overview of Our Work

Motivated by such an importance, we study the diverse characteristics of lending teams in a principled manner and show the advantage of leveraging the team information in the context of two important problems: loan recommendation as well as team recommendation for lenders.

Loan recommendation largely differs from the standard recommendation problem due to two attributes: the transient nature of loans and its binary rating structure. Regarding the transient nature of loans, loans are only available until they have been funded, thus they can be seen as a consumable and limited resource. This attribute of loans makes it a more difficult problem when compared to other recommendation systems such as the Netflix recommendation system in which recommendations for a movie previously liked by similar users can be recommended. Secondly, once loans have been funded, they are no longer available. This binary structure complicates the loan-to-lender relationship due to the fact that a lender who has not funded a loan may not have directly rejected it.

**Maximum-entropy distribution modeling.** To address these difficulties of our domain, we treat the lending activity data as presence-only data or one-class data, and apply a maximum-entropy distribution approach (**maxent**). Maxent has been successfully used in various applications such as species distribution modeling [27, 29] and natural language processing [7].

**Loan recommendation.** Based on the maxent approach, we build team-specific models by fully incorporating a wealthy set of heterogeneous information reflecting the lending behavior of each team. We apply our proposed team-specific lending model in the loan recommendation problem and show that our approach performs significantly better than a single aggregate lending model in which the team diversity is ignored.

**In-depth knowledge about team behaviors.** We use our team-specific lending models to gain valuable insights into the team characteristics. We point out that lending teams are generally more careful in selecting loans with respect to a loan’s geo-location, borrower’s gender, field partner’s reliability, etc., when compared to lenders without team affiliations. In addition, we identify interesting lending behaviors of various lending teams based on lenders’ background and interest such as their ethnic, religious, linguistic, educational, regional, and occupational aspects.

**Team recommendation.** To increase lending activity by community building such as lending teams, we propose a team-to-lender recommendation model that leverages our team-specific lending models. For a given lender, we rank potential teams based on how likely his first few loans are under the team-specific model. We show that our approach outperforms the two baseline approaches.

To the best of our knowledge, this is the first work that has analyzed the characteristics of lending communities in a micro-finance domain and recommended the most appropriate communities based on lender-to-team suitability.

The rest of this paper is organized as follows. Section 2 discusses related work. Section 3 describes our basic pre-

processing steps to handle the heterogeneity of Kiva data; in addition, we have made the post-processed data readily available on the web for other researchers. Section 4 describes our main approaches, and Section 5 reports the prediction performances for loan recommendation. Section 6 describes insightful knowledge revealed from our analysis. Section 7 presents the team recommendation application of our proposed model. Section 8 discusses further implications based on our analysis. Finally, Section 9 concludes the paper and discusses future work.

## 2. RELATED WORK

In this section we discuss related work about (1) general recommender systems, (2) opinion-based recommender systems, and (3) micro-finance analysis.

**Recommender systems.** A recommender system, also known as an active information filtering system [6], estimates the utility function for a given user and an item. Typically, there are mainly two different approaches for recommender systems: *content-based filtering methods* and *collaborative filtering methods*. Content-based filtering methods aim at matching users to products by finding similar items which they have liked in the past, while collaborative filtering methods make suggestions based on finding other items that similar users have liked in the past [31, 2]. Within collaborative filtering approaches there are *model-based approaches* which utilize some data mining or machine learning algorithm to find patterns and *memory-based approaches* which typically use user rating data to find recommendations. For a thorough summary of collaborative filtering techniques, please refer to our listed survey articles in the references section .

Although collaborative filtering methods have seen wide use in the design of many recommender systems [34], Kiva’s micro-financing loan data contains three major challenges: the transient nature of loans, the binary rating structure, and the heterogeneity of Kiva data. To address these challenges, our work leans more towards a content-based filtering approach where the proposed lender features represent the user’s profile of preferences while the loan features represent the product content. Traditional content-based approaches focus only on textual information as seen in information retrieval literature [5]. Yet, due to the heterogeneous nature of Kiva data, our approach extends upon the content-based filtering approach and uses *ad-hoc* information retrieval [24] to represent various information as features and predict a relevance score for loans by training a learner model. These types of approaches are widely applicable in various novel applications including online dating systems [12].

**Opinion-based recommender systems.** Sinha and Searingen [32] showed that users’ friends consistently gave higher quality recommendations than those from a recommender system due to the friends’ intimate knowledge of their tastes. The idea is that individuals with similar tastes will form connections and develop a sense of trust within the communities. The following literature on the relation between connections based on shared interests, trust, and agents’ decisions justify the importance of lending teams for member participation in Kiva.

Abdul-Rahman and Hailes [1] claimed this very idea and proposed a trust model for recommender systems that in part showed that agents within a similar context, e.g. professional communities, trusted the opinions of agents with

similar profiles of interest. In virtual communities, trust can be seen as a derivative of both the ability and the combined benevolence and integrity of the agent to be trusted [30]. Ridings et al. [30] hypothesized that this gained trust is positively related to their willingness to give and receive information from within their network. In essence, recommendations from those with similar interests have significant impacts on an agent’s decision because of the strong correlation between trust and interest similarity [38], and furthermore, agents find themselves less vulnerable to risk and are even encouraged to collaborate when trust is present within their network [25]. That is why it is not surprising to see that recommender systems which have incorporated trust models have gained much attention due to their favorable properties for social filtering [26, 21, 16].

In our system, we felt that there were two major dimensions that encouraged participation in a network: interest similarity and civic responsibility. We believe that having like-minded individuals who want to address similar issues of public concern within a team will help foster an environment that is conducive to peer encouragement. Specifically for Kiva, if lending teams are developed around mutual interests, due to the non-profit nature of Kiva participation, lenders are highly likely to trust in the general direction of their lending team network.

**Micro-financial activity analysis.** Technological advancements have reshaped the structure of micro-financing as seen by the effects of the internet on micro-financing [8] and by the transformation of lending transaction behavior caused by peer-to-peer technologies [4]. Studies on micro-finance lending patterns have discovered that lenders choose opportunities based on similarity of interests, emotional responses, and other social biases. Lenders tend to choose borrowers who share similarities to their personal or professional interests, e.g. artists will loan to other artists, and/or trigger an emotional response [3, 14]. Findings specific to Kiva have claimed patterns that show bias within the lending decision process. In particular, women and more physically attractive individuals have a higher probability of receiving support from first-time lenders and lesser-active lenders [18]. Other studies on Kiva have observed the nature of lending behavior by correlating the impact of group dynamics to lending participation [17, 23].

Surprisingly, even with Kiva’s openly available data set, only a handful of research work has used advanced statistical analysis approaches in studying micro-financing. In one study, researchers manually defined a set of categories about the motivation of lending and applied machine learning techniques to train automatic text classifiers using a lender’s *loan\_because* field [23]. Their work only used several simple features such as the loan count and team affiliations to perform regression on lending frequency and amount. They revealed various interesting knowledge about lending behavior, but the used information and techniques are relatively limited compared to our work.

To the best of our knowledge, *our work is the first in-depth study to directly tackle the loan and the lender recommendation problems by actively incorporating the lending team information available from Kiva. As seen in Section 5, we achieve performance viable for practical application and reveal significant finding about lending behavior of teams that has not been discussed in any previous other work.*

### 3. OVERVIEW OF KIVA DATA

The Kiva data set contains a massive set of heterogeneous information about the following types of entities:

- a lender or kiva user  $u$  (1,174,383 in total),
- a lending team  $t$  (25,481 in total),
- a loan  $l$  (564,177 in total),
- a field partner  $p$  (254 in total), and
- a borrower  $b$  (1,099,997 in total).

their transactions (requesting, funding, and paying back loans). In addition to its heterogeneity, the data set contains a complex set of many-to-many relationships, e.g. lenders may concurrently participate in more than one lending team and contribute to multiples loans. There are a variety of data types within the Kiva data including geo-spatial, temporal, categorical, numerical, and free-text unstructured data.

Entities of each type contain various information involving both unstructured data, such as image, video, and text, and structured data, such as geo-spatial, numerical, categorical, and ordinal data. For example, lender entities is represented in terms of its essential web profile data, e.g., a profile image, a registration timestamp, a geo-location, a lending count, an occupation, and other fields. Lending team entities also have its own information including a name, a team category (e.g., religious, common interest, etc.), a brief description, and a webpage URL. Finally, loan entities, which have the most rich set of information, are described by a loan description, a loan sector (e.g., agriculture, food, retail, etc.), a list of borrowers requesting the loan, a field partner, a geo-location, a loan amount, and posted/funded/paid timestamps.

In addition, a complex set of many-to-many relationships are available in the data set. For example, lenders may concurrently participate in more than one lending team and contribute to multiples loans. Field partners manage loans within their local region, while borrowers request loans from their local field partners. These relationships can be represented as various graphs between different entities, and the following two important graphs are directly available from the data set:

- a graph between lenders and loans, which indicates who funded which loans (564,177 edges in total), and
- a graph between lenders and lending teams, which indicates the team membership of lenders (313,040 edges in total).

Kiva provides a recent snapshot of this data set in JSON and XML formats.<sup>1</sup> In this work, we used a 2.9 GB JSON snapshot collected on 5/31/2013. We preprocessed it to obtain the numerical representations of each available field. In particular, the preprocessing of temporal, categorical, and textual fields all required a nontrivial amount of work. For temporal data, such as the loan’s *posting date* and lender’s *sign-up date*, we converted them to a serial date number using Matlab’s *datenum* function, which represents the whole and fractional number of days from a fixed preset date of January 0 in the year 0000. For categorical data, such as a loan’s *country code* and a team’s *category*, we used a dummy

<sup>1</sup><http://build.kiva.org/docs/data/snapshots>

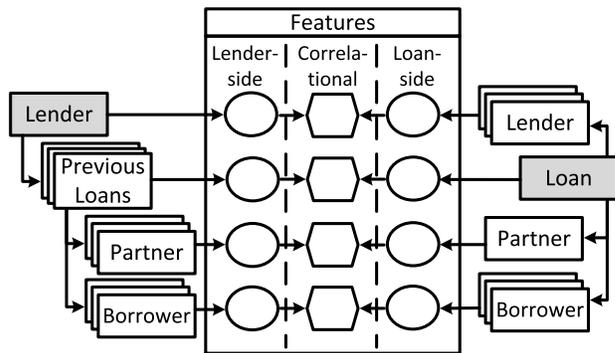


Figure 3: A graph-based feature integration for a lender-loan pair (grey-colored).

encoding scheme that converts a variable with  $m$  categories into an  $m$ -dimensional binary vector where only the values in the corresponding categories are set to ones.

For textual data, we encoded each textual field separately as a bag-of-words vector where an individual dimension corresponds to a unique word. Afterwards we reduced the dimensionality using nonnegative matrix factorization<sup>2</sup> (NMF) [22, 20] to 100 for each textual field. The reason for performing a dimension reduction set is two-fold. First, the vocabulary size, which corresponds to the number of original dimensions, usually reaches up to the hundreds of thousands demanding intensive computation and memory. Second, each of the reduced dimensions improves its semantic meanings by grouping multiple coherent words into a single dimension. These dimension-reduced data can be versatile for both good prediction performance and data/model understanding [10, 36].

As an additional processing step, we created mappings between entities from the different tables. For example, a lender entity found in the table containing metadata for lenders may have a different identifier in another table about the lender-loan graph, and even worse, it may exist in only one table, meaning that some information about it will be completely missing. The mappings we created allow these issues to be handled with ease. Finally, we made the processed data available as a Matlab data file at <http://fodava.gatech.edu/kiva-data-set-preprocessed>.

### 4. MODELING OF LENDING TEAMS

In this section, we describe the process for modeling lending teams in terms of their lending activities in Kiva.

#### 4.1 Feature Integration of Lending Activities

As we briefly highlighted in Section 1, although lending activities are usually performed at an individual lender level, a significant amount of them are driven by lending teams that lenders are affiliated with. Furthermore, different lending teams may have different characteristics in their lending behaviors. Due to these reasons, we intend to model each lending team separately as follows. Basically, we represent each lending team as a set of its lending activities. A lending activity of a lending team  $t$  is described as a pair  $(u, l)$  of a lender  $u$  belonging to team  $t$  and a loan  $l$ . As depicted in

<sup>2</sup>We obtained the code from <http://www.cc.gatech.edu/~hpark/nmfsoftware.php>

Fig. 3, given a lender-loan pair  $(u, l)$ , we retrieve the various entities to which the lender and loan have links to. For a lender, we obtain the list of his/her previously funded loans as well as their associated partners and borrowers. Similarly, for a loan, we obtain its associated partner and borrowers as well as other lenders who funded the given loan.<sup>3</sup>

**Lender- and loan-related features.** By incorporating the information from these linked entities, we form two feature sets for lenders and loans,  $v^l$  and  $v^u$ , respectively (circles in Fig. 3). Note that all these features are numerically represented as described in the basic preprocessing steps in Section. 3.

In order to maintain a fixed number of dimensions for  $v_i^u$  (or  $v_i^l$ ) given a variable number of linked entities across different lender-loan pairs, we treat multiple entities of the same type as a single averaged entity. For instance, if a lender has funded multiple loans in the past, the feature vectors generated from them are averaged into a single vector. Similarly, features about multiple borrowers associated with a single loan, such as their genders, are also averaged. However, in this process, information about the total number of previous loans or the total number of borrowers is lost. To compensate, we encode any potentially lost information as additional features.

**Lender-loan correlation features.** The lender- and loan-related feature sets built in this manner have counterparts of the same entity type, which can be directly compared with each other. In other words, both lenders and loans have all the feature sets representing borrowers, field partners, loans, and lenders. Intuitively, if lenders prefer to fund a particular type of loan, then these two counterparts would have similar values. To directly take into account such correlation information, we compute an element-wise multiplication of a lender-and a loan-related feature vectors, i.e.,

$$v^{ul} = v^u \circ v^l,$$

and include it as an additional feature set (hexagons in Fig. 3). This correlation feature vector  $v^{ul}$  indicates how strongly the values of a particular dimension are represented in both the lender and the loan; this can be considered as the degree of matching at an individual feature level.

**Temporal features.** Each loan contains temporal information such as its *posted\_date*, *funded\_date*, and *paid\_date*. We assume that the relative time difference between two consecutive lending activities could be an important factor for a particular lender, and thus we encode such information as our features. That is, we construct all temporal features in the form of  $x - y$  where  $x$  is one of *posted\_date* and *funded\_date* of a loan of interest and  $y$  is one of *posted\_date*, *funded\_date*, and *paid\_date* of a lender’s most recent loan. Additionally, we encode the time taken for a lender to fund the loan since it has been posted.<sup>4</sup>

Finally, by including all the features encoded in the above-mentioned manner, we construct an overall feature vector  $f(u, l) = [f_1(u, l) \cdots f_m(u, l)]^T$  representing a lender-loan pair  $(u, l)$

<sup>3</sup>Information about when individual lenders funded a particular loan is not available in the data set. Therefore, we randomly selected five other lenders in our experiment.

<sup>4</sup>Due to the lack of the temporal information about individual lenders’ activities, we assume all lenders funded a particular loan at the same time as a loan’s *funded\_date*.

## 4.2 Maximum-Entropy Distribution Model

Due to the nature of lending activities, lenders who have not chosen to fund a loan have not necessarily opted against funding it. It is often the case that the lender never knew about it. This type of data is known as presence-only or one-class data. In order to properly address this issue, we propose to apply a maximum-entropy distribution model (**maxent**) of a lending team’s lending activity data.

**Formulation.** To model the lending activity, we want to use maxent to estimate the density  $\pi^t(u, l)$ , which indicates how likely a lender  $u$  will fund a loan  $l$  as a member of a lending team  $t$ . The goal of maxent is to maximize the entropy, or *uncertainty*, of an estimated density  $\hat{\pi}^t(u, l)$ , subject to the constraint that the expected value of each feature  $f_i(u, l)$  under  $\hat{\pi}^t(u, l)$  should be the same as that of  $f_i(u, l)$  under the empirical distribution  $\tilde{\pi}^t(u, l)$ . The main idea of maxent is, given presence-only data, to estimate the target density as uniform as possible by assigning the most probability evenly to unseen parts of the space while keeping the same expected values of individual features as those from the observed data.

It has been shown that solving the maxent distribution can also be converted to solving a maximum likelihood of  $\hat{\pi}^t(u, l)$  in the form of a Gibbs distribution [11, 29], i.e.,

$$\hat{\pi}^t(u, l) = q_{\lambda^t}(u, l) \propto \exp\left(\sum_{j=1}^m \lambda_j^t f_j(u, l)\right),$$

whose probability is represented as a log-linear model. With a relaxation on the above constraints, the maxent distribution is solved by maximizing a penalized log-likelihood of these presence data, i.e.,

$$\max_{\lambda^t} \sum_{i=1}^{n^t} \log q_{\lambda^t}(u_i, l_i) - \sum_{j=1}^m \beta_j |\lambda_j^t|,$$

where a lender-loan pair  $(u_i, l_i)$  is the  $i$ -th lending activity ( $n^t$  activities in total) in team  $t$ , and  $\beta_j$ ’s are regularization parameters.<sup>5</sup> Similar to many other approaches such as the lasso in least squares [35], the  $l_1$ -norm regularization on  $\lambda_j^t$ ’s gives a sparse representation, which is robust against overfitting and deals with potential multi-colinearity among different features.

The algorithm for solving this formulation follows a coordinate-wise descent procedure, and in our work, we use the implementation available at <http://www.cs.princeton.edu/~schapire/maxent/>. In this implementation, in order to overcome the limitation of the original log-linear model, various additional features derived from original features, such as quadratic, threshold, and hinge features, are internally generated and used so that it can handle nonlinear responses of original features.

## 5. LOAN RECOMMENDATION

### 5.1 Experimental Setup

**Lending teams and activity data selection.** To begin, we chose the top 200 lending teams with the highest number of lending activities, totaling 70% of the total lending amount made by teams. From each of these lending

<sup>5</sup>For the information about how to set the regularization parameters, we refer the readers to [28].

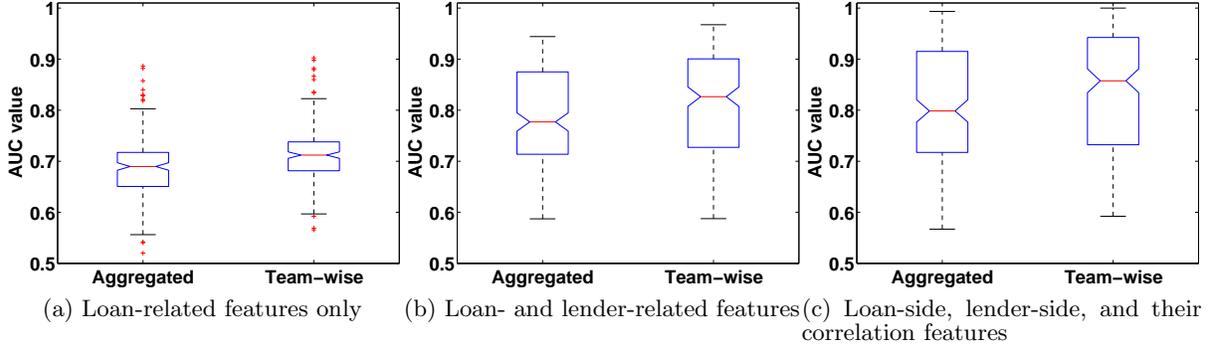


Figure 4: The AUC values for the aggregated and the team-specific models across 200 lending teams, depending on various feature sets used.

teams, we randomly selected 5,000 lender-loan pairs in which the funding occurred. Additionally, maxent requires background or pseudo-negative data instances that properly reflect the overall distribution of the data instances. Therefore, we also randomly selected 5,000 lender-loan pairs where the funding did not occur. These samples serve as our training and test sets under a 5-fold cross-validation setup.

**Feature groups.** Each lender-loan pair  $(u, l)$  generated in this manner (10,000 in total for each team) is then encoded as a feature vector, as presented in Section 4.1. The constructed features can be categorized as follows:

1. Textual information (600 dimensions): reduced-dimensional textual features from a lender’s *loan\_because* and a loan’s *loan\_description*.
2. Loan sector (45 dimensions): features about the industry of a loan, e.g., agriculture, food, retail, etc.
3. Geo-location (228 dimensions): features about the country of a loan and/or a lender.
4. Loan delinquency (13 dimensions): features indicating how many previous loans of a lender  $u$  have been defaulted or delinquent.
5. Partner (33 dimensions): features about field partners in terms of their loan amount, rating, delinquency rate, etc.
6. Borrower (12 dimensions): features about borrowers, e.g., a borrower’s gender and whether he/she has a picture.
7. Temporal information (7 dimensions): Relative time differences between a loan  $l$  and a lender’s most recently funded loan as well as the time taken for a lender  $u$  to fund a loan  $l$  since it has been posted (Section 4.1).

Eventually, the overall feature vector  $f(u, l)$  is represented as a 938-dimensional vector.

**Performance measure.** The resulting maxent model  $\hat{\pi}^t(u, l)$  trained from these feature vectors for each of the 200 lending teams gives the probability that a funding will happen for any given lender-loan pair  $(u, l)$  in case the lender is a member of the lending team  $t$ . It allows us to rank the most appropriate loans for a particular lender as well as the most appropriate lenders for a particular loan. Therefore, we are interested in the quality in terms of the resulting ranking of a given test set of lender-loan pairs. In this respect,

we report the area under the receiver operating characteristic curve (AUC) value, which measures how well the data samples with funding are ranked higher than background samples.

## 5.2 Recommendation Performance

**Baseline approach.** To evaluate our proposed personalized team-specific model, we designed a baseline approach to compare our method against, as follows: In the baseline approach, we aggregate all data instances from different teams into a single data set and train a single aggregated maxent model  $\hat{\pi}^a(u, l, t)$  on these aggregated data. To make the comparison fair, we still incorporate team information in this case, such as a team’s loan amount, member count, and category (e.g., common interest, religious, etc. ), associated with the lender-loan pairs as additional 60-dimensional features into the aggregated model, as seen as its new input argument  $t$  in  $\hat{\pi}^a(u, l, t)$ . In this manner, our baseline method still uses the same amount of information as in the proposed team-specific models but does not distinguish between the lending characteristics from different lending teams. That is, the distinction is only made at a feature level, but not at a model level.

**Performance comparison.** For our proposed team-specific model and the baseline method, we compared the AUC values of 200 teams on a 5-cross-validation setup when using either (1) loan-related features only, (2) loan- and lender-related features, and (3) lone-related, lender-related, and their correlation features. As can be seen in Fig. 4, in all cases, the team-specific model showed nontrivial performance increase over the baseline method, indicating that the diversity of lending behaviors across different teams is indeed substantial, which cannot be fully handled at a feature level.

Furthermore, beyond just the information about loans of interest, as we incorporate additional features discussed in Section 4.1, the performance was shown to improve significantly. This indicates that the information about lenders and their past lending activities is critical in predicting his/her next loan activities. Particularly, the fact that such performance increase due to involving lender information is significant even for our team-specific models implies that the diversity does exist within each lending team.

Overall, the highest AUC value we could achieve by using all the proposed features under team-specific model was 0.84 on average across teams, which seems useful in practice in recommending loans to lending teams as well as their

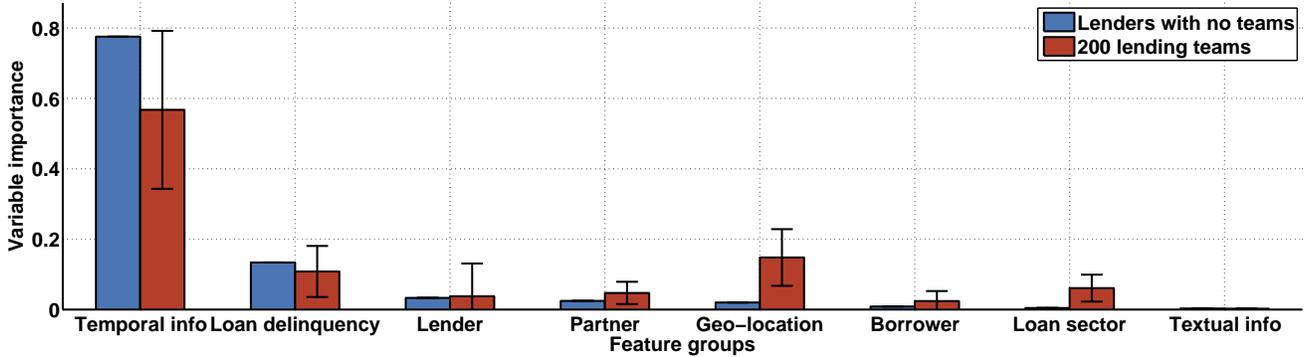


Figure 5: The variable importance scores due to each feature group. The values are sorted in a decreasing order with respect to a no-team model.

lenders.

## 6. EXPLORATORY ANALYSIS

In this section, by analyzing the team-specific maxent models, we present our in-depth analysis on diverse behaviors among different lending teams.

**Interpretation of maxent models.** Our strategy to analyze each team’s lending behaviors is to compute the variable importance scores in its maxent model. As a way to compute these scores, we chose to use the permutation importance score, often used in various machine learning methods [33]. The importance score of each variable is determined by randomly permuting the values of that variable among the training data items and measuring the resulting decrease in the AUC value. A large decrease in the AUC value indicates that the model depends heavily on that variable. After computing the permutation importance of each variable, the values are normalized to give relative percentage values.

### *What do lending teams care about? (lending teams vs. lenders with no teams)*

Before exploring the individual team level, we tried to identify the critical factors that had the greatest influence on the activities of general lending teams. To this end, we collected another set of lender-loan data purely from lenders who had no affiliations with lending teams. Then, we modeled an additional maxent model using this data set, which we refer to as the *no-team* model, in the same way as described in Sec.4. Fig. 5 shows the comparison between the 200 team-specific models and the *no-team* model.

**Commonalities.** Temporal information is shown to be the most important feature in both cases, which is consistent with the findings discussed in [9]. That is, once lenders begin lending, they tend to either keep funding other loans continuously or lose interest drastically as time goes on. Next, a negative experience which involves a loan delinquency also significantly impacts the next lending activities for both cases. Information about lenders, such as the number of previous loans and their characteristics, were shown to be moderately important factors, but the importance of lender information varied highly across different teams, as shown by the relatively large error bar corresponding to the ‘lender’ feature group. Finally, the influence of textual information was shown to be minimal compared to other information because of the significant noise and the information sparsity

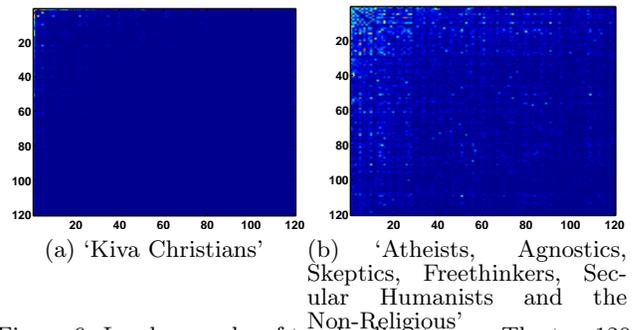


Figure 6: Lender graphs of two lending teams. The top 120 lenders with the most number of loans have been chosen from each team. The value at each cell represents the number of common loans between two lenders within a particular team.

throughout free-text data.

**Contrarities.** At the same time, Fig. 5 highlights interesting distinctions of lending teams compared to lenders with no teams. Specifically, information about a loan’s geo-location and the loan sector were the two most critical factors in lending teams’ activities. It indicates that *lenders in lending teams actually care more about the location and the purpose of the requested loan* than lenders without team affiliations. Furthermore, it was also shown that lending teams care who the borrowers and the associated partner are, as shown in the relatively higher importance of the corresponding variables compared to the *no-team* model.

## 6.1 Individual Team Characteristics

We felt that it was important to explore deeper to discover the diverse behaviors present at a team-specific level. This motivation came as a result of noticing how lending teams had such a large influence on the activity of its members. We decided to rank lending teams based on their variable importance scores for different feature groups. In doing so, we were interested in which teams cared the most (or the least) for each aspect, and in the fundamental characteristics of each team that drove these behaviors. We approached this problem from the perspective of a lender as well as a loan.

### 6.1.1 Lender Feature

As mentioned above, the dependency on lender information varied highly across lending teams. Although it was not reported in this paper, the most influential feature about a

Table 1: The top five teams influenced the most by each of the feature groups.

Features	Lending teams
Industry	‘KivaFriends - Agriculture Loans’, ‘Ravelry.com’, ‘101 Cookbooks’, ‘Give Green - Environmental Loans’, ‘Thailand’
Geo-location	‘Para México’, ‘Philippines’, ‘Kiva Muslims’, ‘Kiva Detroit’, ‘Portugal’,
Field partner	‘Amici di Raffaele (Raphael’s Friends)’, ‘Woodlands’, ‘Compadres’, ‘Lauren Avezzie’, ‘Kiva Jews’
Borrower	‘women empowering women’, ‘HALF THE SKY: Empowering Women’, ‘Georgia Southern Alumni’, ‘www.idu.cc’, ‘Tareto Maa’

lender was shown to be the number of loans he/she had previously funded. As we analyzed the teams that were most (or least) influenced by lender features, we found out that dependency on lender information is inversely correlated with the diversity of team leaders.

For instance, Fig. 6 shows a comparison of the lender graph between two groups of teams - ‘Kiva Christians’ and ‘Atheists, Agnostics, ...’. The ‘Kiva Christians’ group was one of the teams that were most heavily influenced by lender features. In other words, the lending activities in ‘Kiva Christians’ were significantly influenced by the lender feature, mainly by the number of his/her previous loans. This implies that lending activities are mainly dominated by only a small number of highly active lenders. On the other hand, ‘Atheists, Agnostics, ...’ was one of the least influenced teams, indicating that lending activities are more evenly distributed over lenders with various numbers of previous loans. In our ‘Kiva Christians’ graph (Fig. 6(a)), you will notice that one person, as seen in the first row/column, showed significant overlap in their lending activities with all the other members. On the contrary, in our ‘Atheists, Agnostics, ...’ graph, commonly funded loans were found amongst a large number of different lenders. It is clear from this observation that the latter case could be led by many lenders with a different degree of activities while in the former case, lending teams are mostly led by a few leading lenders.

### 6.1.2 Loan Feature

Now, we analyze the most influenced teams, i.e. special-focus teams, by different feature groups from a loan perspective. Table 1 presents the top five teams corresponding to each feature group.

**Industry.** The ‘KivaFriends - Agriculture Loans’ team funded 77% of its total loans to the agriculture industry, while only 21% of the entire set of loans belongs to this category. The teams ‘101 Cookbooks’ and ‘Give Green - Environmental Loans’ made strong contributions to the agriculture industry while also making significant contributions to the food industry. The ‘Ravelry.com’ team, whose website focuses on knitting and crocheting, funded 23% of its total loans to the art industry, while only 2% of the entire set of loans belongs to this category. These interesting relationships can be expressed as a *homophily*, as observed by the fact that *people tend to fund loans similar to what they like*.

Finally, although we did not find a reasonable explanation for it, we noticed that the ‘Thailand’ team funded 79% of its total loans to the food industry, while only 26% of the entire set of loans belongs to this category.

**Geo-location.** The teams ‘Para México’ and ‘Philippines’, as their team names imply, funded 49% and 91% of the total loans to Mexico and Philippines, respectively. These teams were strongly dedicated to their designated countries especially considering that the total percentage of

loan requests from Mexico and Philippines were only 2% and 15%, respectively.

‘Kiva Muslims’ made major loan contributions to Palestine (12%), Pakistan (10%), Tajikistan (8%), Lebanon (8%), etc., where the dominant religion was Islam. The total percentages of the loans to these countries were only 6%, 2%, 3%, and 1%, respectively.

For the team ‘Kiva Detroit’, the top lending country was shown to be the USA, which holds 11% of this team’s total loans while the total percentage of USA loans were minimal at 0.2%. Interestingly, we found that Kiva recently started supporting a local small business specifically in the USA under the name, Kiva City,<sup>6</sup> and Detroit was selected as the first Kiva City. Finally, the ‘Portugal’ team exhibited a unique behavior. It made 6% of their total loans to Mozambique, as compared to the percentage of total loans to this country, 0.5%. The most probable reason is as follows. Historically, Mozambique was one of the few former colonies of Portugal in Africa, and the official language is still Portuguese. Looking into the languages in which the loan description was written, 92% of the loans from Mozambique were described in Portuguese while only 0.5% of the entire loans were available in Portuguese.

**Field partner.** Features about the field partner generally represent the reliability and credibility as represented by the total loan count/amount, a default/delinquency rate, currency exchange loss rate, etc. From a lender’s or a lending team’s viewpoint, choosing an appropriate partner is critical in minimizing the risk of losing money. By looking at the top ranked teams for this feature group, we found that they are mostly composed of a relatively small number of people with a large number of loans per member. For example, the teams ‘Amici di Raffaele (Raphael’s Friends)’, ‘Woodlands’, ‘Compadres’, and ‘Lauren Avezzie’, each of which had 10, 45, 2, and 17 members in total, funded 305.3, 303, 2627, and 263.2 loans per member, respectively, which are significantly higher than 30.5, the average number of loans per member among the entire 200 teams. This seems reasonable in that highly active lenders are more likely to manage their funds more carefully so that they can support a larger volume of non-profit activities for a long period of time.

On the other hand, the other top ranked team, ‘Kiva Jews’, which had only 14.2 loans per member, did not have as many highly active lenders. Instead, our results suggest that members of this team are more wary of the risk of lending.

**Borrower.** Borrower information is composed of (1) a borrower’s gender and (2) a borrower’s picture availability. Among the top ranked teams for borrower features, most teams were found to be mainly influenced by the gender, usually in favor of women. It is easily understood that the teams ‘women empowering women’ and ‘HALF THE

<sup>6</sup><http://www.kiva.org/kivacity>

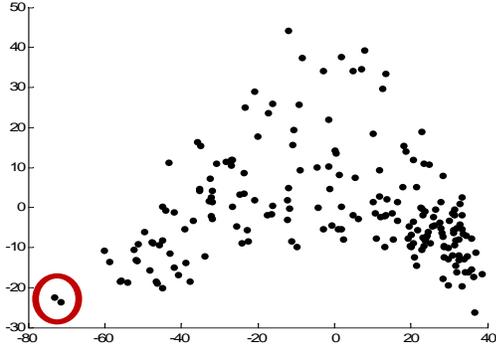


Figure 7: A PCA visualization of variable importance vectors corresponding to 200 lending teams. The two points shown in a red ellipse correspond to the teams ‘Expired Loans’ and ‘Late Loaning Lenders’, respectively.

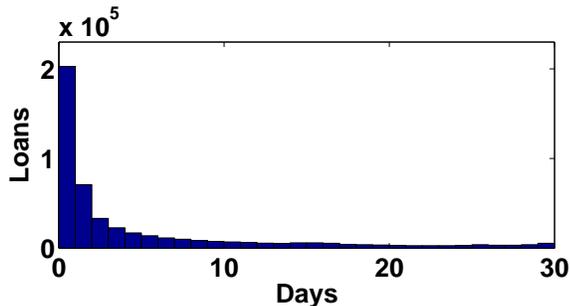


Figure 8: A distribution of the time taken for a loan to be fully funded.

SKY: Empowering Women’ funded 95% and 98% of their total loans to female borrowers, respectively. However, it is surprising that the team ‘Georgia Southern Alumni’ funded 98% of their total loans to female borrowers.

The two other teams ‘www.idu.cc’ and ‘Tareto Maa’ also funded 85% of their loans to female borrowers. For the former, the variable importance about the borrower’s picture availability was significantly higher than any other teams, indicating that lenders in this team are highly unlikely to fund loan requests that do not provide a borrower’s picture. Finally, the team ‘Tareto Maa’, translated as “help for the Massai”, was founded in order to eradicate the tradition of female circumcision and child marriages within the Massai people. We find this to be the most likely reason as to why they focus on female borrowers. Additionally, this team was also highly ranked as the 29th in terms of the geo-location feature group, and we found that 25% of their total loans were requested from Kenya, where the territory of the Masai is located.

### 6.1.3 Outlier Teams from Visualization

So far, we looked into various team characteristics due to each of the feature groups. However, in this approach, it may be difficult to pinpoint those lending teams that are moderately different from usual teams from a particular aspect but are significantly different when incorporating entire features altogether. For this reason, we generated the variance importance vector corresponding to the entire features for each lending team. Then, we applied principal component analysis (PCA) [19] in these vectors in order to visually represent teams in a 2D space. The visualization result shown

in Fig. 7 clearly reveals two outlier teams, ‘Expired Loans’ and ‘Late Loaning Lenders’.

Most loans in Kiva have a 30-day period of expiration for its fundraising. Fig. 8, which shows the distribution of the time taken for a particular loan to be fully funded, indicates that most loans are fully funded within a few days but some loans take much longer even possibly failing to be fully funded. These expired or soon-to-be expired loans can be thought of as relatively unpopular ones within the Kiva lending community.

These two lending teams are unique in that their mission is to fund unpopular loans in order to avoid their expiration. We found that their lending behaviors were different from other teams from many aspects. For instance, both teams funded more loan requests from males than from females, e.g., 60% and 58%, respectively. Their top lending countries included Tajikistan, Bolivia, Lebanon, Azerbaijan, Jordan, and El Salvador, all of which were not actively funded by other teams.

## 7. TEAM RECOMMENDATION

In this section, we utilize our team-specific maxent model for team recommendation for lenders who are not yet affiliated with any teams.

### 7.1 Team Model-based Approach

To perform this task, we assume that a lender  $u$  and his/her first  $c$  loans  $l_i^u$ ’s ( $i = 1, \dots, c$ ) are available and that a lender  $u$  did not join any teams while funding these first  $c$  loans. Then given a lender-team pair  $(u, t)$ , we compute the likelihood function  $L(u, t)$  as

$$L(u, t) = \frac{1}{c} \sum_{i=1}^c \hat{\pi}^t(u, l_i^u). \quad (1)$$

Eq. (1) computes an average likelihood value for  $c$  loans when a lender  $u$  funded them as a member of the team  $t$ . By comparing these values from all the 200 teams, we determine the most appropriate team  $t_u$  for a lender  $u$  as

$$t_u = \arg \max_t L(u, t) = \arg \max_t \frac{1}{c} \sum_{i=1}^c \hat{\pi}^t(u, l_i^u).$$

### 7.2 Recommendation Performance

**Baseline approach.** For our baseline approach, we extend the aggregated model  $\hat{\pi}^a(u, l, t)$  described in Section 5.2. That is, the team recommendation for a lender  $u$  based on the aggregated model is performed as

$$t_u^a = \arg \max_t L_a(u, t) = \arg \max_t \frac{1}{c} \sum_{i=1}^c \hat{\pi}^a(u, l_i^u, t).$$

**Performance measure.** From each team  $t_r$ , we randomly selected  $k$  lenders, e.g.,  $k = 500$  in our experiment, from each team along with their first five loans, i.e.,  $c = 5$ . Then, among the values of  $L(u, t)$  for 200 teams for a lender  $u_i$ , we computed the mean reciprocal rank of the correct team  $t_r$  as

$$MRR = \frac{1}{k} \sum_{i=1}^k \frac{1}{r(u_i, t_r)}$$

where  $r(u_i, t_r)$  is the rank of  $L(u_i, t_r)$  among  $L(u_i, t)$ ’s. The maximum value of this measure is one, and a higher

Table 2: A comparison of the reciprocal rank values for team recommendation. The values represent an average value over lenders from 200 teams while those in parentheses represent the variance.

	Mean reciprocal rank
Team model-based approach	.0851 (.0365)
Aggregated approach	.0548 (.0210)
Random assignment	.0294

value of this measure indicates that the correct team ranked higher than other teams.

**Comparison Result.** Table 2 shows the comparison of the team recommendation results. Our team-specific model-based approach clearly shows a better performance than the aggregated approach as well as random assignment. The main reason for this performance improvement is because our team-specific model-based approach allows each team’s model to treat different aspects of a lender and a loan in their own way. That is, some teams may put more emphasis on a particular aspect such as a loan’s geo-location or industry while other teams do not.

## 8. IMPLICATIONS

Lenders are often motivated by their strong preference to address needs that they feel connected to, whether it is to a loan industry or a geographical location or even to a particular feature in the borrower, such as their gender or situation. As presented in Section 6, these preferences are often rooted in the lender’s ethnic, religious, linguistic, educational, regional, and occupational background.

Our comprehensive analysis on diverse lending characteristics of lending teams basically reveals a meaningful yet distinct set of preferences and their connections to the underlying human factors. Based on our findings, we believe that Kiva could drastically improve their practice at both a lending team and an individual lender level.

**Team-level approach.** Kiva would strongly benefit from continuously guiding each team to the appropriate loans. Currently, most teams have a small number of leading lenders that drive their team members’ activities. Identifying the team leaders along with a deep understanding about their interest is crucial in keeping each team as active as possible by providing them with loans they feel are within their interest. Furthermore, Kiva should also encourage each lending team to expand its interest. As discussed in Section 6.1.3, less popular loans could be easily funded with more effort at the team level. Kiva could even incentivize those teams that try to expand their interest. However, such strategies requires the ability to properly identify the lending activities outside of a team’s original interest, the clue of which we provided in this paper,

**Lender-level approach.** Encouraging lenders to join teams that have similar interests based on their background would allow diverse communities in Kiva to thrive. In our team recommendation task, we indirectly addressed this issue by utilizing one’s previous loans, e.g., whether he/she funded loan requests from a particular gender, country, etc. However, the quality of our recommendation would significantly benefit if it also took a lender’s background, e.g., ethnic, religious, educational, regional, and occupational information, into account when making team recommendations. By proactively collecting such additional information

from lenders, rather than just collecting the lender’s current location and occupation, Kiva would be able to accurately suggest the best lending teams to each lender, ultimately to increase the average level of participation for their lenders.

## 9. CONCLUSION AND FUTURE WORK

In this paper, we studied diverse characteristics of lending teams in a widely-used micro-finance service, Kiva.org. By treating lending activities as presence-only data and by fully incorporating the rich set of data available in Kiva, we modeled each lending team as a maxent distribution and achieved a competitive performance for the AUC value of 0.84 in loan recommendation applications for the top 200 lending teams. In addition, we discovered diverse lending behaviors by interpreting the resulting maxent models and enlightened the underlying social aspects that support these findings. Finally, we applied our team-specific models in the team recommendation application, showing promising results for matching lenders to appropriate lending teams.

The importance of our work and the information-rich nature of the Kiva data open up various future research possibilities. We describe a few of them in the following.

**Social influence in lending teams.** One promising direction is to further study the influence team members have on one another. As briefly seen in the lender graph in Fig. 6, most teams have a small number of key members that direct the entire team’s activities. Analyzing such processes in the context of peer pressure and information diffusion [15] would provide a deeper insight into how the lending team influences their nonprofit activities.

**Evolution of lending teams.** We are also interested in the process in which teams emerge and decline over time. Kiva runs a team leaderboard where they show the ten most active teams by the amount of loans and new lenders on a monthly basis. These leading teams change frequently over time, and it would be also interesting to study the cause of their rise and fall.

## Acknowledgments

The work of these authors was supported in part by the Defense Advanced Research Projects Agency (DARPA) XDATA program grant FA8750-12-2-0309. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of DARPA.

## 10. REFERENCES

- [1] A. Abdul-Rahman and S. Hailes. Supporting trust in virtual communities. In *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences*, page 9, 2000.
- [2] G. Adomavicius and A. Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6):734–749, 2005.
- [3] J. Andreoni. Impure altruism and donations to public goods: A theory of warm-glow giving. *The Economic Journal*, 100(401):464–477, 1990.
- [4] A. Ashta and D. Assadi. Do social cause and social technology meet? impact of web 2.0 technologies on

- peer-to-peer lending transactions. *Cahiers du CEREN*, 29:177–192, 2009.
- [5] R. Baeza-Yates and B. Ribeiro-Neto. *Modern information retrieval*. Addison-Wesley, 1999.
- [6] N. J. Belkin and W. B. Croft. Information filtering and information retrieval: two sides of the same coin? *Communications of the ACM*, 35(12):29–38, 1992.
- [7] A. L. Berger, V. J. D. Pietra, and S. A. D. Pietra. A maximum entropy approach to natural language processing. *Computational linguistics*, 22(1), 1996.
- [8] T. Bruett. Cows, kiva, and prosper. com: How disintermediation and the internet are changing microfinance. *Community Development Investment Review*, 3(2):44–50, 2007.
- [9] J. Choo, C. Lee, D. Lee, H. Zha, and H. Park. A better world for all: Understanding and promoting micro-finance activities in kiva.org. Technical report, Georgia Institute of Technology, 2013. Available at <http://hdl.handle.net/1853/49182>.
- [10] S. Deerwester, S. Dumais, G. Furnas, T. Landauer, and R. Harshman. Indexing by latent semantic analysis. *Journal of the Society for Information Science*, 41:391–407, 1990.
- [11] S. Della Pietra, V. Della Pietra, and J. Lafferty. Inducing features of random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(4):380–393, 1997.
- [12] F. Diaz, D. Metzler, and S. Amer-Yahia. Relevance and ranking in online dating systems. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, pages 66–73, 2010.
- [13] M. Flannery. Kiva and the birth of person-to-person microfinance. *Innovations*, 2(1-2):31–56, 2007.
- [14] J. Galak, D. Small, and A. T. Stephen. Micro-finance decision making: A field study of prosocial lending. *Journal of Marketing Research*, 48(SPL):S130–S137, 2011.
- [15] D. Gruhl, R. Guha, D. Liben-Nowell, and A. Tomkins. Information diffusion through blogspace. In *Proceedings of the 13th international conference on World Wide Web*, pages 491–501, 2004.
- [16] R. Guha. Open rating systems. 2003.
- [17] S. Hartley. Kiva. org: Crowd-sourced microfinance & cooperation in group lending. 2010.
- [18] C. Jenq, J. Pan, and W. Theseira. What do donors discriminate on? evidence from kiva.org. 2012.
- [19] I. Jolliffe. *Principal component analysis*. Wiley Online Library, 2005.
- [20] H. Kim and H. Park. Sparse non-negative matrix factorizations via alternating non-negativity-constrained least squares for microarray data analysis. *Bioinformatics*, 23(12):1495–1502, 2007.
- [21] M. Kinateder and K. Rothermel. Architecture and algorithms for a distributed reputation system. In *Trust Management*, pages 1–16. 2003.
- [22] D. D. Lee and H. S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401:788–791, 1999.
- [23] Y. Liu, R. Chen, Y. Chen, Q. Mei, and S. Salib. I loan because...: Understanding motivations for pro-social lending. In *Proceedings of the fifth ACM International Conference on Web Search and Data Mining*, pages 503–512, 2012.
- [24] C. D. Manning, P. Raghavan, and H. Schütze. *Introduction to information retrieval*, volume 1. Cambridge University Press Cambridge, 2008.
- [25] S. Marsh. *Formalising Trust as a Computational Concept*. PhD thesis, Department of Mathematics and Computer Science, University of Stirling, 1994.
- [26] M. Montaner, B. López, and J. L. de la Rosa. Opinion-based filtering through trust. In *Cooperative Information Agents VI*, pages 164–178. 2002.
- [27] S. J. Phillips, R. P. Anderson, and R. E. Schapire. Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190(3):231–259, 2006.
- [28] S. J. Phillips and M. Dudík. Modeling of species distributions with maxent: new extensions and a comprehensive evaluation. *Ecography*, 31(2):161–175, 2008.
- [29] S. J. Phillips, M. Dudík, and R. E. Schapire. A maximum entropy approach to species distribution modeling. In *Proceedings of the twenty-first international conference on Machine learning*, pages 83–90, 2004.
- [30] C. M. Ridings, D. Gefen, and B. Arinze. Some antecedents and effects of trust in virtual communities. *The Journal of Strategic Information Systems*, 11(3):271–295, 2002.
- [31] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl. Analysis of recommendation algorithms for e-commerce. In *Proceedings of the 2nd ACM conference on Electronic commerce*, pages 158–167, 2000.
- [32] R. R. Sinha and K. Swearingen. Comparing recommendations made by online systems and friends. In *DELOS workshop: personalisation and recommender systems in digital libraries*, volume 106, 2001.
- [33] C. Strobl, A.-L. Boulesteix, A. Zeileis, and T. Hothorn. Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC bioinformatics*, 8(1):25, 2007.
- [34] X. Su and T. M. Khoshgoftaar. A survey of collaborative filtering techniques. *Advances in Artificial Intelligence*, 2009:4:2–4:2, Jan. 2009.
- [35] R. Tibshirani. Regression shrinkage and selection via LASSO. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58:267–288, 1996.
- [36] W. Xu, X. Liu, and Y. Gong. Document clustering based on non-negative matrix factorization. In *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*, pages 267–273, 2003.
- [37] M. Yunus. *Banker to the Poor*. Penguin Books India, 1998.
- [38] C.-N. Ziegler and J. Golbeck. Investigating interactions of trust and interest similarity. *Decision Support Systems*, 43(2):460–475, 2007.