Personal and Ubiquitous Computing (2018) 22:825–838
https://doi.org/10.1007/s00779-018-1165-y

ORIGINAL ARTICLE

Visual analysis of risks in peer-to-peer lending market

Zhao Xiao [1] · Yuelei Li [2] · Kang Zhang [3,4]

Received: 17 October 2017 / Accepted: 24 May 2018 / Published online: 6 June 2018

Abstract
Online peer-to-peer (P2P) lending financial products have been developing rapidly in recent years. This investment method is designed for people free of high-rate debts. However, the lending and borrowing affairs between anonymities may potentially produce risks, including wash sale and money laundering. Apart from the well-documented research on the causal factors and economic influence of the P2P lending market, limited attention has been paid to the risk management of individual P2P lending platforms. This study presents a visual analysis method that detects and analyzes risks in P2P lending transactions. Moreover, we evaluate our approach on real-world P2P data sets and report our findings.

Keywords Visual analytics · P2P lending · Financial risk · Clustering · Fluid dynamics simulation

1 Introduction

The Internet has become a powerful force in accelerating financial innovation. Peer-to-peer (P2P) lending market is an Internet-based financial product developed in the era of e-commerce. P2P lending market anonymously matches lenders to borrowers entirely online and provides borrowers with loans of relatively low interests and lenders with relatively high returns. To achieve successful and secured borrowing and lending, P2P lending intermediaries should verify the credit and financial states of lenders based on personal information and payment histories and filter out unqualified borrowers. A P2P lending platform divides a large sum of loan requirements into small loan components and assigns them to lenders who provide loan offers. In general, borrowers can avail of loans at rates lower than those provided by banks. Meanwhile, lenders can obtain higher returns than saving in banks. P2P products considerably stimulate the financial market potential. P2P lending services provide mutual benefits, given that they support ordinary individuals in the community, particularly those who are unqualified for traditional bank loans, free from high-rate debts. It also satisfies those with urgent needs, such as medical treatments, small business, and education.

Zopa, which is the first P2P lending platform, was launched in the UK in 2005. Eventually, its business model inspired many imitations all over the world.

In China, at the time of writing, the outstanding balance in the P2P lending industry has reached 1.29 trillion yuan (≈ 2000 billion dollars) [27]. Most P2P lending companies have had immense success. At present, the initial business model of P2P lending has been modified or reformed by leading companies, such as Prosper (2006, USA), Lending Club (2007, USA), and Paipaidai (2007, China), based on their respective domestic regulations.

However, P2P lending, operated entirely online between anonymities, is favored by financial criminals in conducting Ponzi schemes, credit fraud, and money laundering. These illegal activities have been jeopardizing the P2P lending industry. According to WDZJ (www.wdzj.com), by the end of March 2018, the cumulative number of platforms is 6081, of which 4198 platforms become bankrupt or involved in risky events [27]. Chinese P2P lending markets have, however, made great contributions to solving the difficult problems of

✉ Kang Zhang
kzhang@utdallas.edu

Zhao Xiao
zxiao@tju.edu.cn

Yuelei Li
liyuelei@tju.edu.cn

1 School of Computer Software, Tianjin University, Tianjin 300354, China

2 College of Management and Economics, Tianjin University, Tianjin 300072, China

3 Department of Computer Science, The University of Texas at Dallas, Richardson, TX 75080-3021, USA

4 Faculty of Information Technology, Macau University of Science and Technology, Avenida Wai Long, Taipa, Macau, China

financing for small- or medium-sized enterprises. Therefore, P2P lending industry plays a vital role in the development of China's economy, and thus being able to control the systemic risks of P2P lending industry. How to assist and guide the P2P lending industry is of critical importance.

Economists focus on modeling P2P lending credits and optimal loan rates using traditional statistical methods. Limited academic attention has been paid to detecting the characteristics of illegal behavior. This paper presents a methodology and toolkit of a visual analysis for P2P lending managers to detect abnormal behaviors that may involve illegal activities.

2 Related work

Prior studies have clearly indicated that borrower anonymity could lead to latent trading risks in the P2P lending business. Chen et al. [6] determined that overdue behaviors are likely to show when borrowers have low incomes or live in low economic conditions. The increased borrower overdue instances in P2P lending transactions result in substantially strict scrutiny by systems, such as the Securities Act of 1933. Zhang and Liu [38] reported the rational herding effect in P2P microloan markets. Wei and Lin [35] compared two most popular mechanisms in P2P lending, i.e., auction and posted price, from several profiles such as market participants, transaction outcomes, and social welfare. Gonzalez and Loureiro [13] indicated that P2P loan success is related to personal characteristics of lender and borrower, such as attractiveness, age, and gender.

The evidence of significant racial disparities is also found by Pope and Sydnor [28]. For example, in loans listing with photos, blacks are 25 to 35% less likely to receive funding than the whites with similar credit profiles. Herzenstein et al. [16] studied data from the P2P lending website Prosper.com and reported that borrowers' identity claims in narrative forms influence lender decisions. A recent research by Ge et al. [11] suggested the borrowers' social media information could be used for credit screening, default reduction, and debt collection. However, most scrutiny systems focus on regular financial activities rather than criminal behaviors that are demonstrated intentionally by taking advantage of loopholes in a financial platform. Therefore, these scrutiny systems are hardly capable of prohibiting fraudulent investment activities, such as the Ponzi scheme.

Users in the P2P lending market show herding behaviors. For example, lenders in Prosper.com, an online P2P lending market, are likely to bid on loans with numerous existing bids [15]. Zhang and Liu [38] obtained evidence that lenders tended to make decisions by relying heavily on the previous selections of other investors.

Mainstream research focuses on the effect of borrower and lender behaviors on interest rates and the likelihood of funding [33]. Iyer et al. [19] found that soft information was important for P2P lending in Prosper. They also found that the interest rates bidding by lenders could identify the default more accurately than other certifications, such as credit scores. Hildenbrand et al. [17] reported that the perverse incentives were not fully recognized by the Prosper market and some group leaders wrongly bid, for their leader rewards, resulting in lower interest rates on high-risk loans. Freedman and Jin [9] employed social network analysis to learn how the lenders in Prosper markets respond to a particular type of soft information represented by their social ties. Further studies by Freedman and Jin [9] found that when the group leader rewards were discontinued, the group and related social ties changed as well. Via social connections, Lin et al. [23] reported that the friendships of borrowers could help them not only being funded but also with lower default rates.

Lenders and borrowers have built para-social relationships via a P2P lending platform. Their friendships affect their perceived trustworthiness. Friendship improves the possibility of successful funding and reduces the interest rate on funded loans [23]. Bachmann et al. [3] conducted an empirical research on the credit scores of borrowers and other factors to identify the determinants of interest rates in the P2P lending market in China. Yum et al. [36] studied P2P lending in microfinance for which borrowers were unbankable. Thus, signals on the creditworthiness of new borrowers are extremely limited. Puro et al. [30] introduced a borrower decision aid, facilitating the formalization of the decision-making process of lenders or borrowers. These studies investigate the related factors but do not measure the effects of these factors.

Several P2P lending platforms have even built their internal social networks. Therefore, numerous studies have been focusing on social networks in the P2P lending platforms and the associations of social factors and financial indicators. Those studies employ ordinary least square (OLS) regression analysis in their empirical studies. Berger et al. [2] analyzed the roles of intermediary users and concluded that they would facilitate the reduction of information asymmetries. Greiner and Wang [14] used the elaboration likelihood model (ELM) and suggested a trust-building mechanism among consumers to reduce risks. Matsubara [25] utilized the game theory to propose a protocol for studying last-minute bidding. Freedman et al. [8] determined the similarities and differences on information verification issues between social networks and the P2P lending market.

Kou et al. [22] proposed the multiple criteria decision-making (MCDM) method to evaluate several clustering algorithms of financial data sets. Their experiments confirm the effectiveness of the MCDM method in analyzing real-life credit and bankruptcy risks. Other researchers used mathematical tools. Ceyhan et al. [5] presented a model based on the temporal progression of bidding to predict the success of loan requests. In this model, the authors focus on analyzing numerous features and are concerned on the extent of signals that would affect the market sentiments. To determine the relations between the social behavior and money lending, we also refer to Freedman and Jin [8] on the structure and function of a complex network.

3 P2P lending process and graph representations

Our study focuses on discovering abnormal transactions and users to support risk control of P2P lending platforms by identifying fraud. The business model of the P2P lending market varies slightly in different countries. In countries with established credit systems (e.g., FICO), the P2P lending platforms could identify the risk exposure based on user credit records. In contrast, information asymmetry in countries that lack credit systems makes risk control a challenge.

A few years ago, the Chinese government led the establishment of a credit rating system. Only a few large financial enterprises or state-owned organizations are authorized to access personal credit information. Most P2P lending platforms in China lack the authority to access such information.

Since 2015, the government approved several leading financial enterprises, such as Alipay and Tencent, to start credit investigation. Then, the P2P lending platforms could access the users' credit data via collaboration. However, only 2 years of credit records are far from enough for P2P lending platforms to make informed decisions. Compared with the American and European markets, the P2P lending markets in China access asymmetrical information, thereby involving substantially unpredictable risks.

3.1 Data sources

Over 500 P2P lending platforms have been established in China (CBC News Finance Research Center 2013 [4]). The current study collected data of over 2000 users, 15,000 loan submissions, and 60,000 bid submissions from RenRenDai, one of the largest online P2P lending platforms in China.

A user in RenRenDai can register as a lender and borrower. Borrowers submit their loan requests. The maximum loan amount is determined by the borrower credit points and repayment capability. RenRenDai divides a submitted loan into several small equal portions and invites lenders thereafter to bid for one or more portions of the loan. A loan is successfully completed when all portions are bided before the deadline. Otherwise, the process fails. The process flow of a P2P loan is shown in Fig. 1.

The legitimate lender-borrower relationships are presented in Fig. 2a, where lender A invests a loan to a borrower B. In this figure, a circle represents a lender or a borrower and an arrow indicates the fund direction from lender to borrower. A lender can invest with more than one borrower (see Fig. 2b) and more than one lender can invest with one borrower (Fig. 2c). When borrowers with unpaid debt lend money, the lending relations become a chain structure or a graph structure (see Fig. 2d–i).

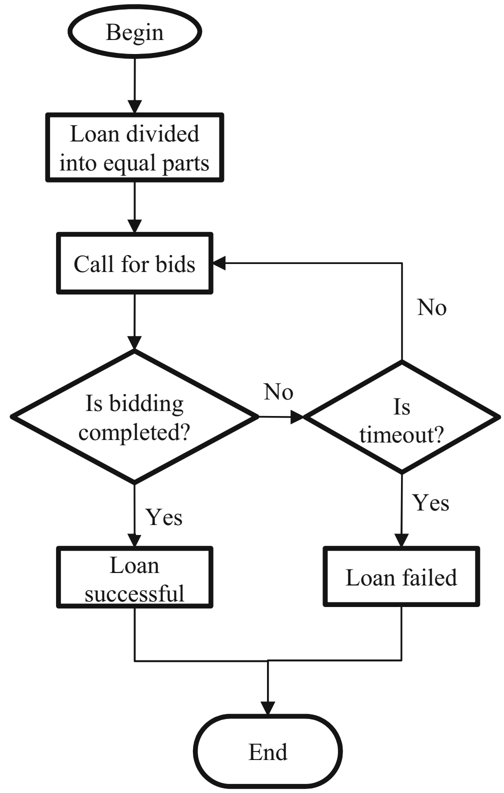


Fig. 1 Flowchart of a loan transaction

We use these loan relationships as basic structures to compose substantially complex structures expressed as graphs.

3.2 Fraud cases visualized

P2P lending platforms should prohibit three types of user activities, namely, Ponzi scheme, credit fraud, and money laundering. Each of these user activities has a specific user relationship structure. A Ponzi scheme is sustained with continuously joining new investors. Welshers attract investors by high profit promises and use new investments to pay interests. Investors should attract numerous new investors to obtain their investments. Thus, a Ponzi scheme structure is similar to a pyramid (see Fig. 3a). Users involved in a Ponzi scheme start from random positions, connect to their nearby users, and gradually move up levels. The Ponzi structure would eventually become a swirl at the top level (see Fig. 3b).

P2P lending is such a new concept that most people do not know the process details. Welshers therefore take advantage of this fact to implement a Ponzi scheme. They absorb money by promising investors with high interests. The payments of

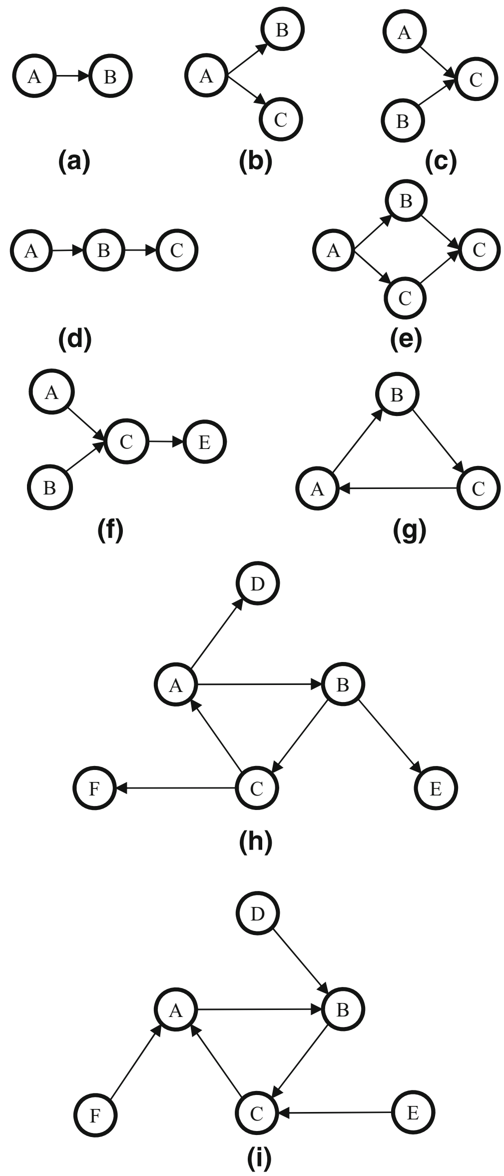
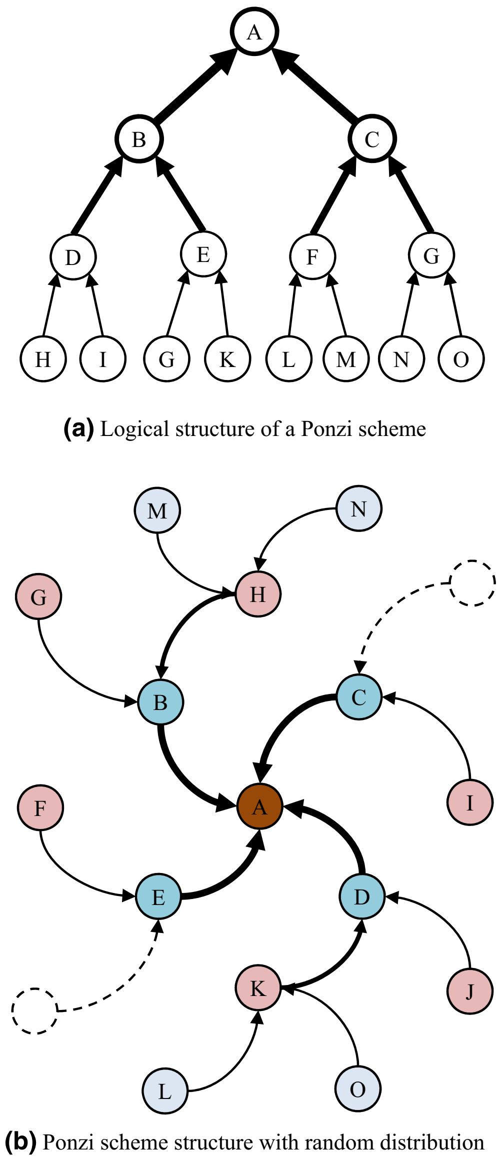


Fig. 2 Basic loan relationships



(a) Logical structure of a Ponzi scheme

(b) Ponzi scheme structure with random distribution

Fig. 3 Ponzi scheme structure and fund flow

interests are actually collected from other investors' lending. Welshers could collect a large amount of money by repeating these operations. Eventually, welshers disappear with money, leaving the investors' P2P lending accounts full of debts.

In the case of credit fraud, welshers fake as many good loan records as possible to obtain a maximum loan amount. They organize a group of P2P lending market users to consummate many internal loan transactions (see the red flow in Fig. 4), thereby forming a cycle.

Money laundering occurs when welshers use a P2P lending platform to disperse "dirty" money by investing many lending transactions. They lend their "dirty" money to other accounts under their control with an extremely low rate for which no investor would bid. Thereafter, they clean their money by receiving payments. In this process, money laundering welshers organize many borrowers to borrow from target lenders for extremely short terms. These lenders return the capital and interest on time. Lenders and borrowers act

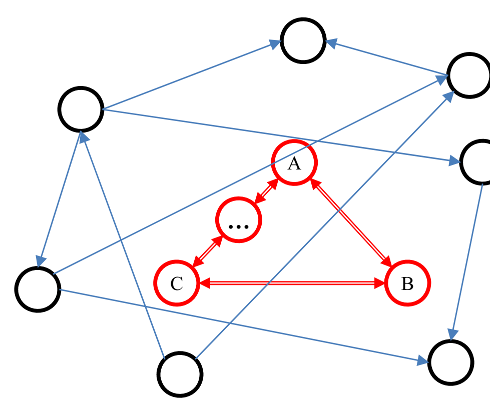


Fig. 4 Fund flow of a typical case with high fraud risk

frequently in the short term. To avoid risks, money laundering welshers act as borrowers to lend money to one another (see Fig. 5). Using the information on the P2P lending platform, we can detect abnormal transactions that are similar to the money laundering structure.

The three types of illegal activities occur in a short period, during which the involved users are generally active.

We infer the features of fund flows, loan relationships, and clustering of certain users that may suggest suspicious transactions by analyzing the pattern of illegal behavior. Therefore, we may detect potential risks as follows:

a. *Find the most active users and their trading groups.*
b. *Find groups with members who are closely related.*
c. *Simulate the fund flows between users.*
d. *Identify the characteristics of suspicious fund flows.*

4 Graph layout and multi-cluster analysis

We develop a spider program to simulate user browsing and collect user information and trading data from RenRenDai over a few months. We have successfully collected over 15,000 user information (including location, income, education, personal assets, lending history, and credit points) and over 60,000 bid data (including total amount, each bid amount, biding status, interest rate, and terms of borrowing). Table 1 shows the details of the data.

We use the force-directed algorithm [29] and K-means clustering algorithm [18] to represent the information clearly and intuitively and to visualize the loan relationships, which bear similarity to social network relationships.

4.1 Layout algorithm

Force-directed graph drawing algorithms assign forces among the set of edges and set of nodes of a graph [10]. Spring-like attractive forces based on Hooke's law are typically used to attract pairs of endpoints of the graph edges toward each other. Repulsive forces, similar to those of electrically charged particles based on Coulomb's law, are simultaneously used to separate all pairs of nodes. After achieving equilibrium states of forces in the system, the edges tend to have uniform lengths (because of spring forces) and unconnected nodes tend to be drawn further apart (because of electrical repulsion). Edge attraction and node repulsion forces can be defined by functions as required, apart from the physical functions of the springs and particles.

For RenRenDai data, nodes represent users $U = \{u_1, u_2, \ldots, u_n\}$, lines $L = \{l_1, l_2, \ldots, l_m\}$ indicate lending relationships between users, and the force v_i ($i = 1..m$) is modeled as the lending amount. Figure 6 shows a P2P lending force-directed graph that reaches the steady condition (i.e., equilibrium state).

Table 1 P2P lending data features

Features	Descriptions
Lender background information	User ID, gender, age, marital status, city of residence, education info, housing situation, vehicle situation, bank debt info
Lender credit information	Credit points, lending history, current debt
Lending transaction information	Total amount, bid amount, time limit, start time, ending time, progress, loan purpose

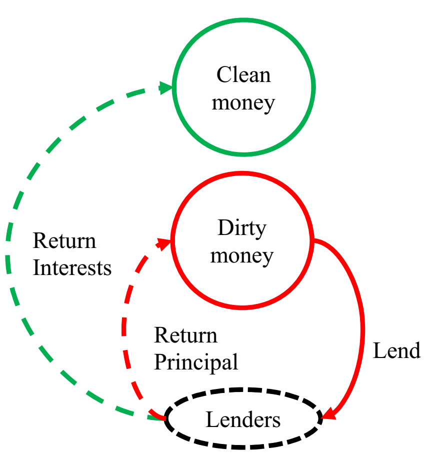


Fig. 5 Fund flow in money laundering

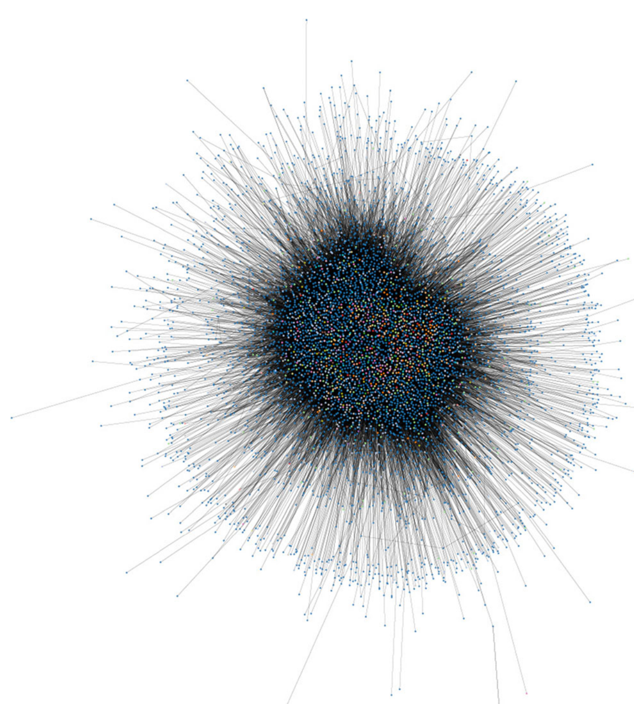


Fig. 6 Force-directed graph of P2P lending data

4.2 Clustering

The graph layout via the force-directed algorithm does not show different user groups. We wish to divide users into five groups according to their lending amounts, to be aligned with the five-star credit ratings of RenRenDai.

Given the above graph layout, we first obtain the center of the entire graph by calculating the average of all the nodes' coordinates in the graph and then cluster users in the following steps:

a. Randomly select five users as five cluster centers.
b. Calculate the distances between each user's lending amount v_i and the five cluster centers.
c. Assign each v_i to the cluster whose center has the shortest distance to v_i.
d. Calculate the center for each cluster as the new center and repeat step b until the clusters become steady.

The above algorithm divides users into five clusters (or groups) at the end, so that we could color them differently in our visualization.

5 Results

The RenRenDai data is clustered and visualized in Fig. 7. The *Y*-axis represents the lending amount, calculated by the lending amount minus the borrowing amount. The *X*-axis represents the distance between the average lending amount and each user's lending amount.

Each point represents an account (a user) on a P2P lending platform. Discrete points in the peripheral belt represent users without loan records (see Fig. 7). The central portion refers to users with loan records. All points are color-coded based on the lending amounts on which credit rates are determined, and thus, five groups are classified. The top 20% credit rated users are purple, followed by yellow, blue, green, and red. The five groups include proportionally 5.5, 3.96, 7.95, 10.43, and 72.15% of users. In general, the user accounts located in the center are more active than those on the edge.

5.1 Summary

The graph layout and clustering algorithms clearly illustrate the user groups and place the most active users in the center. The next section presents our method for visualizing the relationships between users.

6 Fluid model based on visual analysis

P2P lending platforms provide opportunities for users to interact with one another, by reading the profiles, comments, and replies of other users and following them. Such relationships could be represented as a network, where users are connected by bidding loan requests. Similar to predicting information diffusion, lending network could predict risks in the P2P lending market by analyzing loan flows. P2P lending bids are also time-varying activities. Data evolution in a period is important for activity analysis. Dynamic graph or graph animation is one of the appropriate methods to visualize time-varying data. Archambault and Purchase [1] proved that animation could improve task performance when the stability of dynamic graph drawing is high.

Since 1979, the force-directed method [29] adopted geometry and physical models to express network structures, such as graph drawing. Subsequently, scholars developed force-directed models, electric field models, and flow models to analyze graph structures, such as social networks and information diffusion data [31, 34], and predict the trend of development. There are many other information diffusion models, such as tractable models [21] and CTIC and CTLT models [32]. In particular, dynamic physical models can stably present a balanced status after a process of evolution. Choustova [7] proposed a socioeconomic model for financial market based on Schrödinger's theory. Liu et al. [24] presented the Lattice Boltzmann method (LBM) fluid model to analyze and predict the trend of information diffusion of social networks. They mapped the number of retweeting and number of following users into driving forces and particle velocity and established an information diffusion simulation model based on the LBM fluid model. Their model represents users as nodes and information flow as curves between users in a 2D enclosed space. The shapes and locations of nodes and curves may reveal the information diffusion trend.

P2P lending shares similar information diffusion process and user structures as those for social networks, thereby suggesting the feasibility of using the LBM fluid model on the P2P lending analysis.

6.1 LBM-based fluid dynamic model

LBM is a classical fluid dynamic simulation model that originated from the lattice gas automata (LGA) method. LBM divides an enclosure space (2D/3D) into several equal lattices. Liquids in this enclosure space are mapped into nodes scattered in the lattices [37]. The driving force on liquids acts on each node. We can employ the Navier–Stokes equations to calculate the vector sum force of all nodes in one of the lattices to its neighbors. The status of the lattice approximates the physical properties with liquids. The current liquid status could be simulated by those lattice statuses calculated from the previous step.

Different LBM models that consider the directions of lattice and spaces are available. A popular method of classifying the LBM models is *DnQm*. "*Dn*" means *n* dimensions and "*Qm*" represents the *m* speed directions, such as D2Q5, D2Q9, D3Q19,

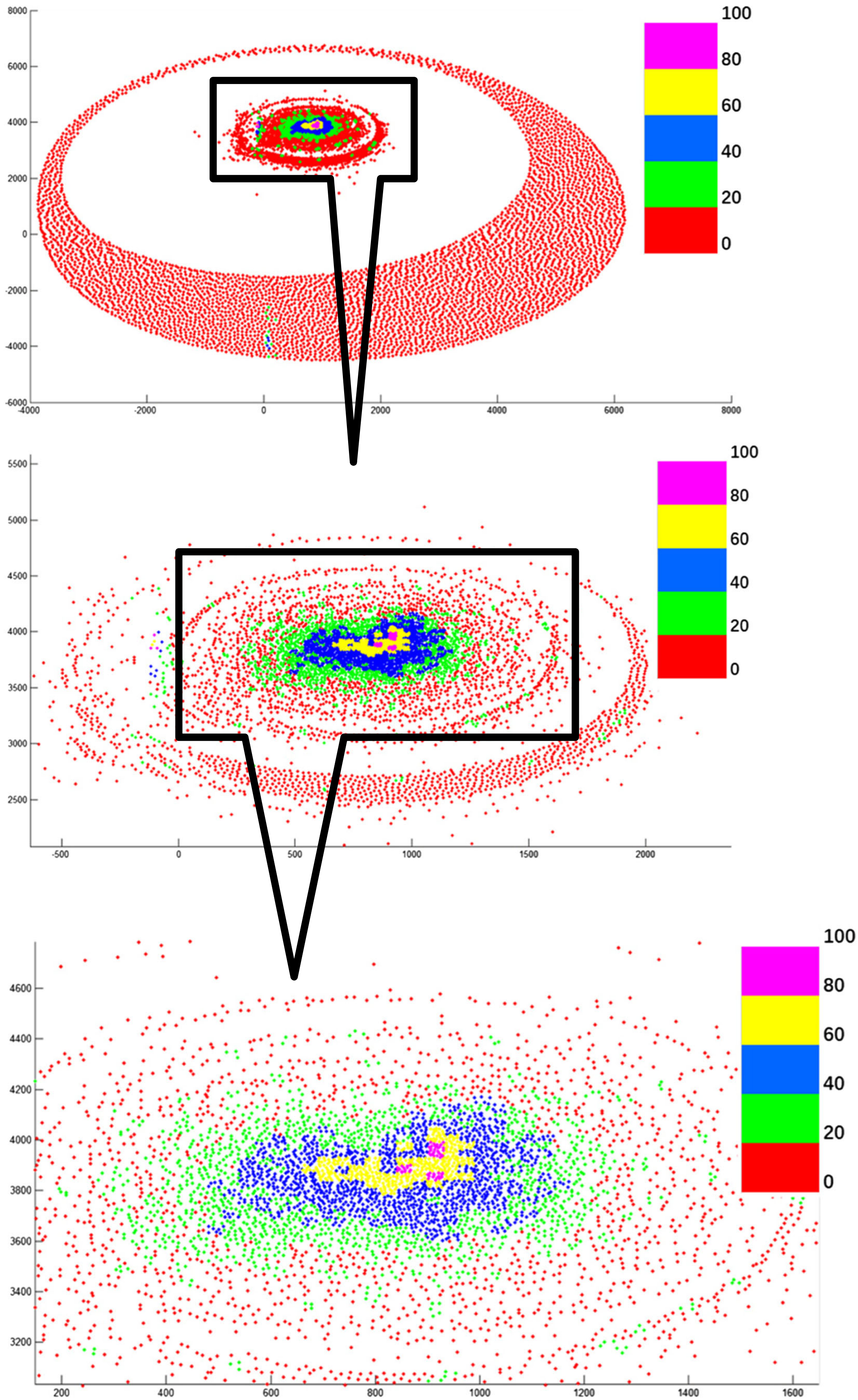


Fig. 7 User clusters

and D3Q27. Figure 8 shows a D2Q9 model that represents a lattice structure with nine speed directions in a 2D space.

Nine arrows indicate nine liquid speed vectors directing to the neighboring lattices. LBM assumes liquid as many microparticles in space. The nine vectors are also contributors to the nine-pointed lattices by microparticle collision between neighboring lattices. The D2Q9 model shows that each node in the lattice has nine particle contributions. Thereafter, a consecutive stream can be calculated by several steps and applied to scattered rules. The main idea of the LBM method is using discrete kinetic equations to represent dynamic Boltzmann. Discrete kinetic equations include the two steps of collision and ballistic streaming [26]. The collision equation is as follows:

$$f_a(r,t+1) = f_a(r-u_a,t) + T_b(\{f_b(r-u_b,t)\}) \quad (1)$$

where u_a is the vector that represents particle contributions to the momentum:

$$u_a = \left[cos\left(\pi\frac{a-1}{3}\right), sin\left(\pi\frac{a-1}{3}\right)\right], a = 1,\cdots,9 \quad (2)$$

where f_a is the distribution function for lattice particles, t is the time step, and r and $r-u_a$ are distances of a lattice and its neighboring nodes, respectively. These distances are inversely proportional to the relaxation time and determined by speed. The streaming equation is defined by:

$$f_i(r+v_i,t+1) = f_i(r,t) \quad (3)$$

where v_i is the velocity vector that points to the neighbors.

Constrained by the bounce-back boundary conditions, the packages, which are going out, reflect opposite momenta.

6.2 Fluid dynamic model for information diffusion

Liu et al. [24] designed a fluid dynamic model to analyze the information diffusion on social networks. They compared the information diffusion and tsunami shock wave traveling and discovered their similarity when they were explained using entropy theory. They adopted the LBM method to simulate information diffusion based on this discovery. This method is combined with the KK algorithm of Kamada and Kawai [20] to improve the speed for drawing large graphs. In fluid simulation, nodes in a lattice represent liquid. Lattice evolution is driven by the liquid forces of the lattice.

The LBM method can also be used to compute the liquid density in fluid simulation, thereby indicating the information diffusion trend between social network users. Liu et al. [24] designed a streamline chart to visualize information density, thereby clearly representing the general information diffusion trend of many social network users (see Fig. 9).

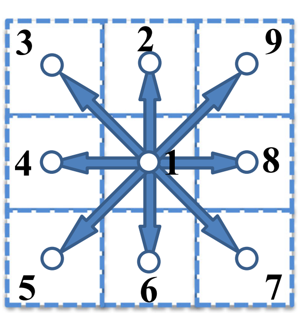


Fig. 8 D2Q9 model

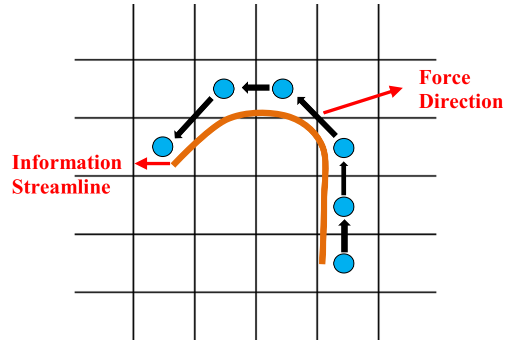


Fig. 9 Generating information streamlines (D2Q9 model)

6.3 Fund flow model on P2P lending

We use the LBM method to simulate P2P lending due to the similarities in user relations and behavior between a social network and P2P lending. A relationship is established when social network users comment or follow others. Similarly, a relationship is established on a P2P lending platform when a requested loan has a successful bid. This structure of users and their relationships can be illustrated as a graph, where nodes represent users and edges represent the relationships between nodes.

The LBM method has two particles in the lattices. We present a fund flow (FF) model, where lattices represent users and driving forces represent the trend of user lending requests. The trend is a vector determined by the user lending directions and amounts (see Fig. 10).

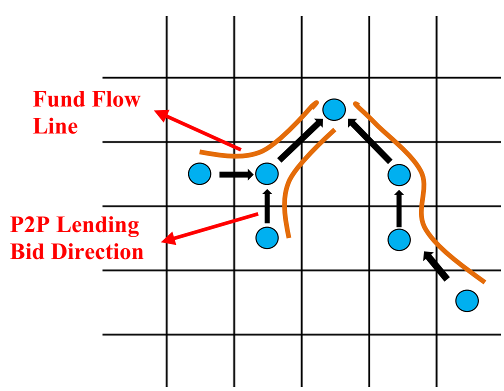


Fig. 10 Generating fund streamline (D2Q9 model)

In Fig. 10, the nodes represent users, the black arrows and linking nodes represent the lending relationships, and the orange lines represent the lending trend of these users. The node that lends money is shown as a field.

$$E = k^{*}M/r^{2} \quad (4)$$

where $\boldsymbol{E}$ is the field strength, M is the amount of lending from the node, r is the distance from a lender node to a borrower node, and k is the scale coefficient.

Table 2 Visual elements in the FF model and their meanings

Features	Implication
Length of segment	Bid amount
Segment color—purple	Trend lines of general transactions
Segment color—orange	Trend lines of huge transactions
Node color—green	Inactive users*
Node color—blue	Active users

*Users who never lend or borrow money

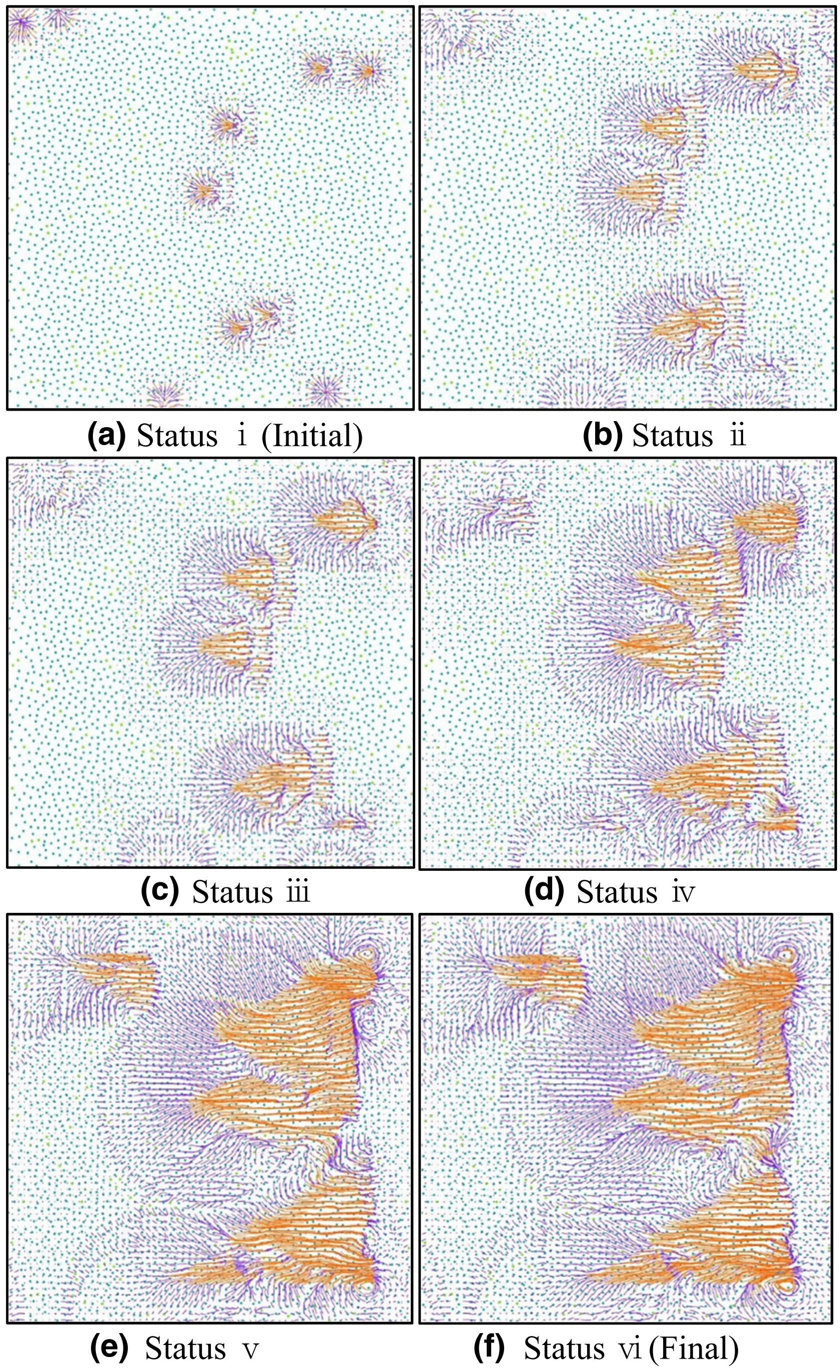
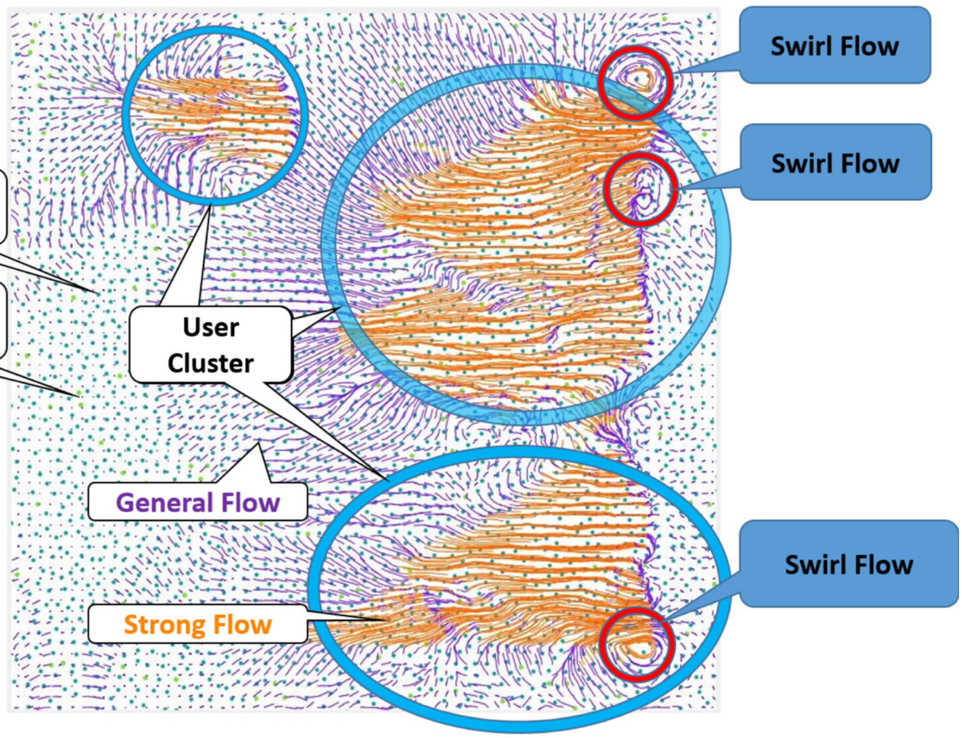


Fig. 11 Visualizing evolution

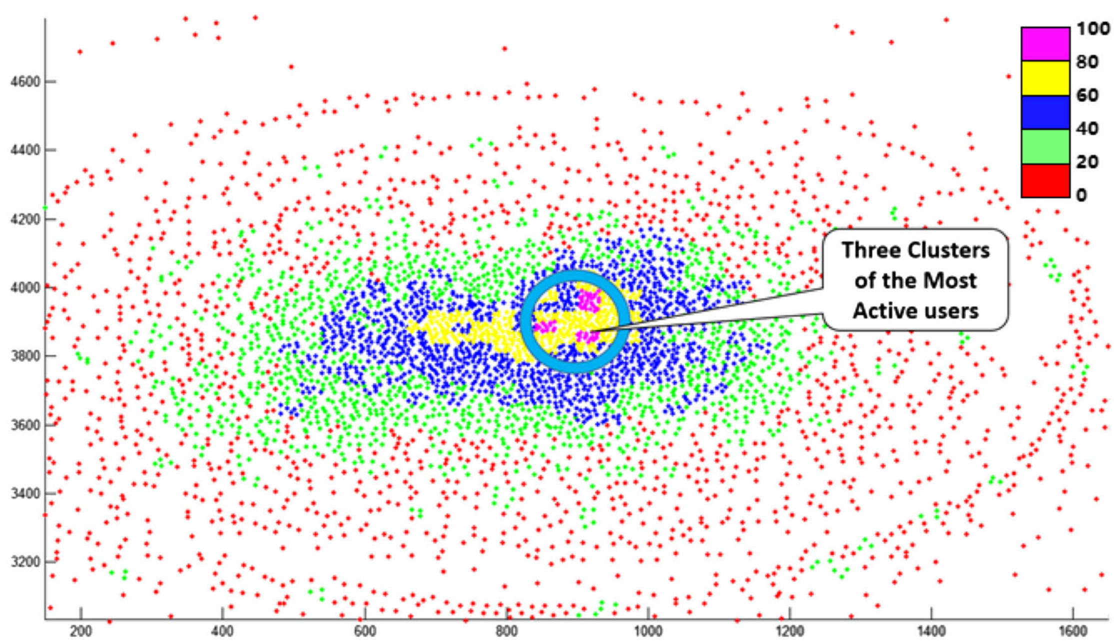
(a) Status i (Initial) **(b)** Status ii **(c)** Status iii **(d)** Status iv **(e)** Status v **(f)** Status vi (Final)

Each node has its own field, the strength of which is determined by all the nearby nodes in the same lattice. The LBM model indicates that every node contributes strength to the field of its lattice. When the lattice edges are short, the corresponding field strength is smooth.

The initial status shown in Fig. 11a is the first state after completing the aforementioned steps. We consider fund change a step rather than a time step in the LBM method. The FF model evolves systematically and the final status represents the FF time sequence of the P2P lending process.



(a) LBM-based FF visualization



(a) P2P lending market multi-cluster visualization

Fig. 12 Visualization of real-world P2P lending data

6.4 Visualization design

In an electric field, lines represent the direction of low potential. No physically visible line exists in any electric field. We use FF to draw field lines to indicate the fund trend among users in the P2P lending market by mimicking the electric field lines.

We use dashed lines rather than solid lines to highlight the evolving processes. Dashed lines extend per segment, with each segment representing the further development from the current status during evolution. Table 2 shows the mapping relationships of the visual elements and semantics in the FF model.

We render the entire evolving process of all transactions, including 33 steps, as shown in Fig. 11. Figure 11a shows the initial status (second step, 1% of all the participants). Figure 11b–e (fifth to tenth steps, 20–99%) shows the intermediate status. Figure 11f shows the final status.

Figure 11 indicates that the users participated in lending at the beginning and over 70% of users were involved in lending thereafter. Dotted purple lines spread and cover an increasing area, thereby indicating an increasing trend of successful lending transactions. Dotted orange lines spread rapidly in the same direction. Initially, all users are in the same status, thereby showing only the basic information of the borrowers. In Fig. 11a, the

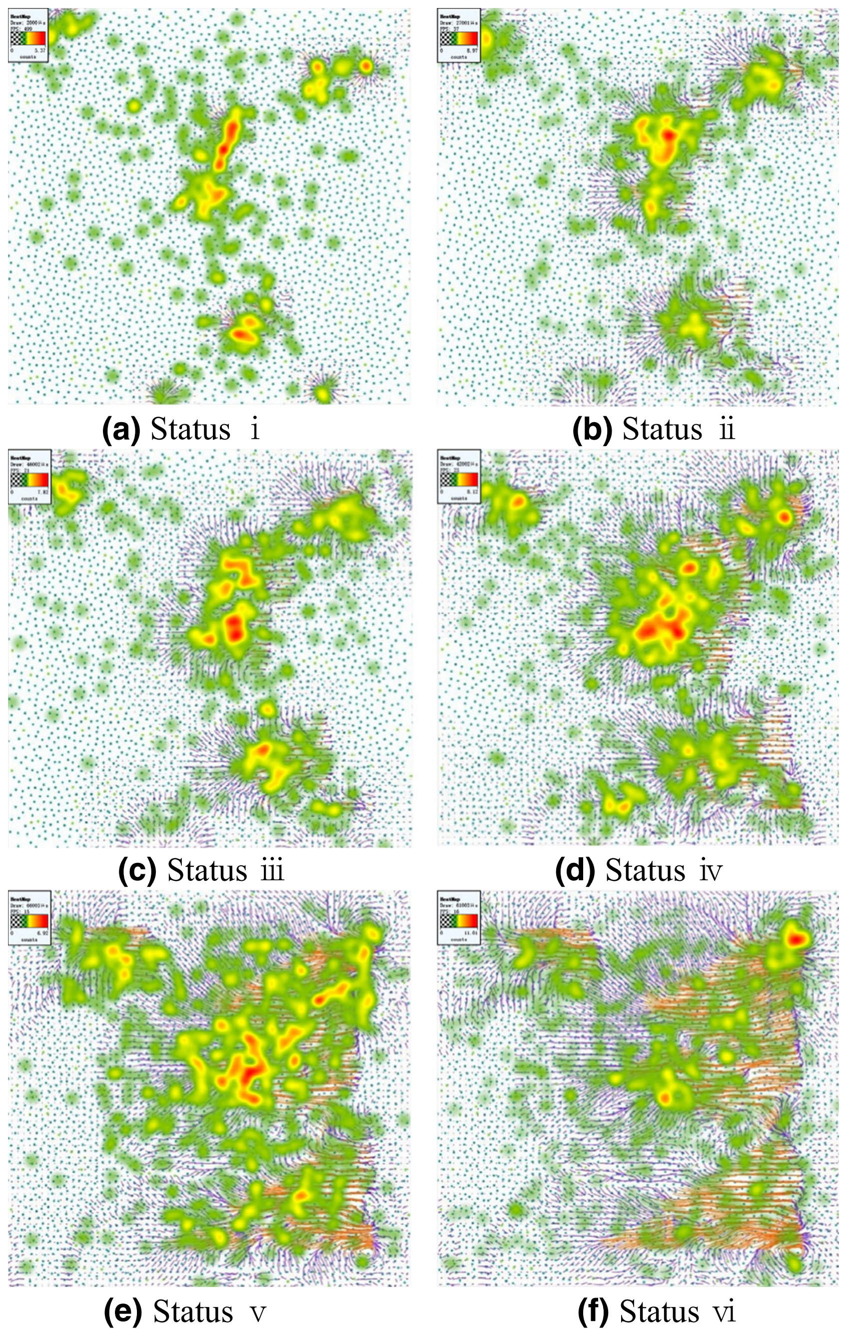


Fig. 13 Heat map of the eye tracking experiment

ten individual active areas indicate that only a few users initially invested. Thereafter, the nearby users become active.

In the later stage, most users are active. Early users generate more transactions than the later ones. Eventually, swirls appear gradually in several areas, thereby suggesting the circulation of FF and suspicious activities. Swirls indicate FF among a small group of users and imply increasing costs because the P2P lending platform charges every transaction. Swirls among small groups reveal an abnormal phenomenon.

Apart from rendering the aforementioned six stages, our approach can also animate along a time line. This animation can be performed by automatically synthesizing all the time step images to construct a movie. The entire evolution can be played at any rate and stopped at any time.

In Fig. 12a, blue nodes indicate regular users and green nodes indicate active users. The three active user clusters with orange lines imply that these users generate a strong FF. Swirl flows circulate within small groups, indicating abnormal phenomena of potentially high risks.

Figure 12b shows the result of the multi-cluster analysis. The most active users are clustered into three pink groups at the center. This result corresponds well with the three orange areas in the visualization of the FF model. The results of the multi-clusters and FCM indicate that the most suspicious fraud activities could be filtered easily. Our case study has confirmed the feasibility and effectiveness of our visual analysis method in predicting risks by analyzing loan relationships, transaction frequencies, and directions of FF. The LBM method enables our FF model to show visually intuitive results for the P2P lending analysis.

Swirl flows could potentially identify intuitively abnormal activities via visualization. However, normal users and risk users are mixed. It is difficult to divide the swirl flows accurately according to the types of users. As a future work, we plan to use a machine learning method to identify potentially risk users.

Our implementation is on a PC with an Intel i7 5600 processor and 8G memory. The FF model on 10 K users and 60 K transactions runs in 5 min. On a faster processor or GPU-enabled PC, it would take much shorter time and thus could support real-time rendering of the visualization. This would allow real-time alert to regulators for any risk activities on such a P2P platform. Without such visualization, it is almost impossible to identify risk activities from the vast amount of P2P transaction data.

7 Evaluation and discussion

In general, line charts are used to visualize financial data, where only two attributes could be represented. However, abnormal features cannot be easily presented in over three dimensions typically found in the P2P lending activities. One of the most important aspects of our study is to evaluate how useful our visualization approach is compared with traditional methods. The most important abnormal feature is swirl flows. It is a qualitative analysis that relies on users' subjective judgments. Eye tracking is therefore considered an effective evaluation for this type of visualization [12]. To analyze the user's foci and attention [34], we conduct an eye tracking experiment to evaluate our approach.

A total of ten college students with different majors volunteered to participate in this experiment. We used a Tobii T60XL eye tracker to track the eye movement of each participant when viewing Fig. 11a–f. They were introduced with the basic design of the FF model. The entire FF model process has 33 steps. The six pictures indicate the different statuses of the entire process from 1 to 99%. Status i is the initial status of 1% and ii to vi are statuses of 20–99%. Each picture is shown for 15 s. The participants were asked to identify any unusual area or features at the end of the viewing. All positions of the participant foci and the time on them were recorded by the eye tracker.

Among various representations of the recorded data, we use a heat map to discover the location of the participants' interest on the pictures (Fig. 13)

Figure 13a–f clearly demonstrates that red points approximately appear in four divided areas. These four areas are marked as hot zones A, B, C, and D in Fig. 14 and blended based on Fig. 13a–f.

In Fig. 14, zones B and C are extremely close; thus, they can be considered one. The three hot zones can be obtained through the eye tracking experiment. This result proves the effectiveness of the visual analysis conclusion of the FF model (Fig. 12a). After the hot zone division was proven, we applied the hot zone division rules to all the images of the eye tracking experiment (see Fig. 15).

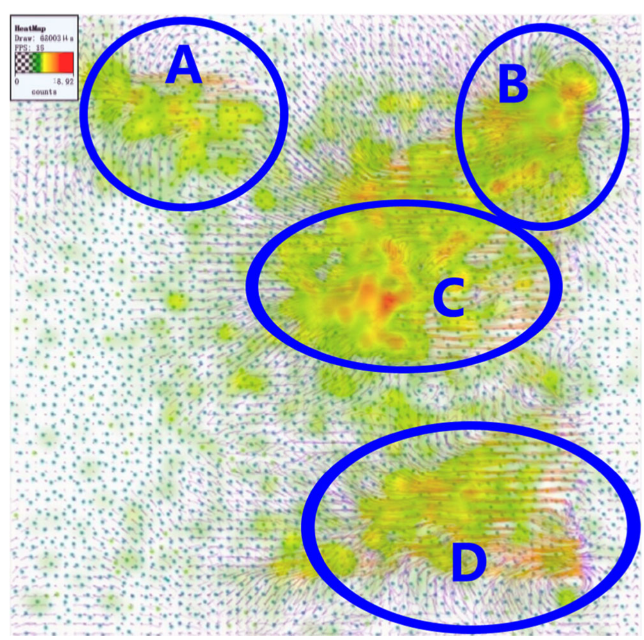


Fig. 14 Hot zones marked on the blending heat map

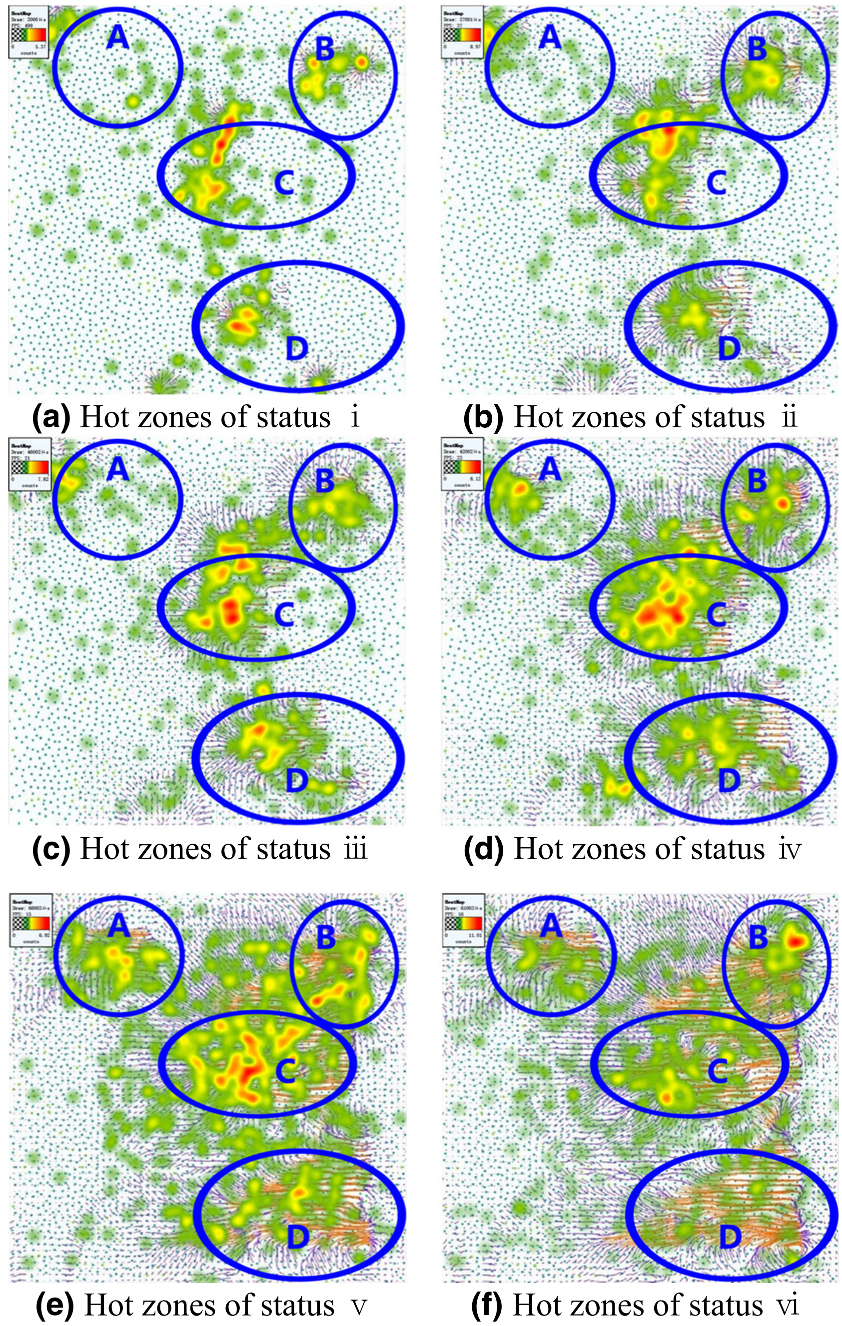


Fig. 15 Hot zones of the different statuses

Hot zone B covers two of the three swirl flow areas in red and zone A covers one area in a cool color. The recognition results of the eye tracker experiment on the divided zones and abnormal features are compared with the FF model visualization (see Table 3).

Table 3 shows that the participants can effectively notice the active areas and primary abnormal features of the FF model without any other indications. Financial risks constantly exist in any financial activity. Our study provides an intuitive and easy approach to identify risks by determining the abnormal activities in the P2P lending market.

Table 3 Comparison of the FF model and eye tracking experiment

Type	FF model	Eye tracking experiment	Discrimination rate (%)
User clusters	3	3 (4*)	100
Abnormal features (swirl flow)	3	2	66.7

*Four zones recognized, two of them combined to one

8 Conclusion

Although the FF model has been proven beneficial, one of its limitations is that it still requires users to manually identify abnormal features through our visualization. In future studies, an automatic or semi-automatic detection mechanism should be investigated, with interactive tools added into our visualization.

Acknowledgments We would like to thank Jie Xu, who assisted in analyzing data using MatLab. We also thank Yuhua Liu, who assisted us with the FF model.

Funding information The current study is partially supported by the National Nature Science Foundation of China (No. 61572348, 71201113, 71671120, and 71661137001) and National High Technology Research and Development Program (863) of China (No. 2015AA020506) and China Scholarship Council (No. 201406250112).

References

1. Archambault D, Purchase HC (2016) Can animation support the visualisation of dynamic graphs? Inf Sci 330:495–509
2. Berger SC, Gleisner F (2009) Emergence of financial intermediaries in electronic markets: the case of online P2P lending. BuR-Bus Res 2(1):39–65
3. Bachmann A, Becker A, Buerckner D, Hilker M, Kock M, Lehmann M, Tiburtius P (2011) Online peer-to-peer lending—a literature review. J Int Bank Commer 16(2):1–18
4. CBN News Finance Research Center (2013) China P2P lending service white paper. China Economic Publishing House, Beijing
5. Ceyhan, S., Shi, X., & Leskovec, J. (2011) Dynamics of bidding in a P2P lending service: effects of herding and predicting loan success. In Proceedings of the 20th international conference on World Wide Web (pp. 547–556). ACM
6. Chen X, Ding XY, Wang BF (2013) A study of the overdue behaviors in private borrowing—empirical analysis based on P2P network borrowing and lending. In Finance Forum 11:010
7. Choustova O (2009) Quantum probability and financial market. Inf Sci 179(5):478–484
8. Freedman, S., & Jin, G. Z. (2008) Do social networks solve information problems for peer-to-peer lending? Evidence from prosper.com
9. Freedman S, Jin GZ (2017) The information value of online social networks: lessons from peer-to-peer lending. Int J Ind Organ 51: 185–222
10. Fruchterman TM, Reingold EM (1991) Graph drawing by force-directed placement. Software: Pract Exp 21(11):1129–1164
11. Ge R, Feng J, Gu B, Zhang P (2017) Predicting and deterring default with social media information in peer-to-peer lending. J Manag Inf Syst 34(2):401–424
12. Goldberg J, Helfman J (2011) Eye tracking for visualization evaluation: reading values on linear versus radial graphs. Inf Vis 10(3): 182–195
13. Gonzalez L, Loureiro YK (2014) When can a photo increase credit? The impact of lender and borrower profiles on online peer-to-peer loans. J Behav Exp Financ 2:44–58
14. Greiner ME, Wang H (2010) Building consumer-to-consumer trust in e-finance marketplaces: an empirical analysis. Int J Electron Commer 15(2):105–136
15. Herzenstein M, Dholakia UM, Andrews RL (2011) Strategic herding behavior in peer-to-peer loan auctions. J Interact Mark 25(1):27–36
16. Herzenstein M, Sonenshein S, Dholakia UM (2011) Tell me a good story and I may lend you money: the role of narratives in peer-to-peer lending decisions. J Mark Res 48(SPL):S138–S149
17. Hildebrand T, Puri M, Rocholl J (2016) Adverse incentives in crowdfunding. Manag Sci 63(3):587–608
18. Ilango MR, Mohan V (2010) A survey of grid based clustering algorithms. Int J Eng Sci Technol 2(8):3441–3446
19. Iyer R, Khwaja AI, Luttmer EF, Shue K (2015) Screening peers softly: inferring the quality of small borrowers. Manag Sci 62(6):1554–1577
20. Kamada T, Kawai S (1989) An algorithm for drawing general undirected graphs. Inf Process Lett 31(1):7–15
21. Kimura M, Saito K (2006) Tractable models for information diffusion in social networks. In European Conference on Principles of Data Mining and Knowledge Discovery. Springer, Berlin, Heidelberg, pp 259–271
22. Kou G, Peng Y, Wang G (2014) Evaluation of clustering algorithms for financial risk analysis using MCDM methods. Inf Sci 275:1–12
23. Lin M, Prabhala NR, Viswanathan S (2013) Judging borrowers by the company they keep: friendship networks and information asymmetry in online peer-to-peer lending. Manag Sci 59(1):17–35
24. Liu Y, Wang C, Ye P, Zhang K (2015) Analysis of micro-blog diffusion using a dynamic fluid model. J Vis 18(2):201–219
25. Matsubara, S. (2001) Accelerating information revelation in ascending-bid auctions: avoiding last minute bidding. In Proceedings of the 3rd ACM Conference on Electronic Commerce (pp. 29–37). ACM
26. McNamara GR, Zanetti G (1988) Use of the Boltzmann equation to simulate lattice-gas automata. Phys Rev Lett 61(20):2332–2335
27. P2P Lending Industry Data (2018) WDZJ website. Retrieved May 4, 2018, from https://shuju.wdzj.com/industry-list.html
28. Pope DG, Sydnor JR (2011) What's in a picture? Evidence of discrimination from Prosper.com. J Hum Resour 46(1):53–92
29. Quinn NR, Breuer MA (1979) A force directed component placement procedure for printed circuit boards. IEEE Trans Circuits Syst CAS-26(6):377–388
30. Puro L, Teich JE, Wallenius H, Wallenius J (2010) Borrower decision aid for people-to-people lending. Decis Support Syst 49(1):52–60
31. Romero, D. M., Meeder, B., & Kleinberg, J. (2011) Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter. In Proceedings of the 20th international conference on World Wide Web (pp. 695–704). ACM
32. Saito K, Kimura M, Ohara K, Motoda H (2010) Behavioral analyses of information diffusion models by observed data of social network. In International Conference on Social Computing, Behavioral Modeling, and Prediction. Springer, Berlin, Heidelberg, pp 149–158
33. Tengvall, Marcus, and Gustav Claesson. (2015) Peer-to-peer lending: the effects of institutional involvement social lending. (Master's thesis, Jönköping International Business School, Jönköping University). Retrieved from https://pdfs.semanticscholar.org/4403/d14fc618de64b5f7486f8bd73996057c8922.pdf
34. Van der Lans R, Pieters R, Wedel M (2008) Eye-movement analysis of search effectiveness. J Am Stat Assoc 103(482):452–461
35. Wei Z, Lin M (2016) Market mechanisms in online peer-to-peer lending. Manag Sci 63(12):4236–4257
36. Yum H, Lee B, Chae M (2012) From the wisdom of crowds to my own judgment in microfinance through online peer-to-peer lending platforms. Electron Commer Res Appl 11(5):469–483
37. Zemla A (2003) LGA: a method for finding 3D similarities in protein structures. Nucleic Acids Res 31(13):3370–3374
38. Zhang J, Liu P (2012) Rational herding in microloan markets. Manag Sci 58(5):892–912