Detecting systemically important platforms in P2P market of China

1st Shan Lu
School of Economics and Management, Beihang University
Beijing, China

2nd Xiaolei Xu
National Computer network Emergency Response technical Team/ Coordination Center of China
Beijing, China

3rd Huiwen Wang
Beijing Advanced Innovation Center for Big Data and Brain Computing; School of Economics and Management,
Beihang University, Beijing, China

4th Jichang Zhao
Beijing Advanced Innovation Center for Big Data and Brain Computing; School of Economics and Management, Beihang University
Beijing, China, jichang@buaa.edu.cn

5th Zhen Wu
National Computer network Emergency Response technical Team/ Coordination Center of China
Beijing, China

***Abstract*—Fueled by advances in technology and shifts in generational values, the market of online peer-to-peer (P2P) platforms is growing prosperously in China. While the development of this rapidly-evolving industry has stimulated the financial market where traditional banks could not reach, collapses of well-known platforms due to malpractice, fraud or business failure are raising great concerns about systemic risk, as troubles of systemically important platforms may trigger wide range of failures because of the network of financial exposures among the industry. However, the network of Chinese P2P lending platforms remains unknown. To fill this vital gap, we performed the first complex network analysis of the Chinese P2P lending platforms using empirical platform-level data. The network analysis enables us to detect the systematically important platforms, through the static importance indicators as well as cascading scheme. A further grouping analysis presents the different characteristic of the platform with different importance levels, which in turn proves the network modeling is able to function effectively in detecting the systemic important platforms.**

***Index Terms*—P2P lending, network modelling, systemic risk**

I. Introduction

Peer-to-peer (P2P) online lending, namely lending through Internet, is a hot topic in China. The convenience of completing all procedures, including the authorization, accounting, clearing, delivery, etc., through Internet has met people's requirements in getting loans. With an estimated two thousands P2P Internet finance platforms (henceforth denoted as 'platform) across the country, online P2P service has successfully connect individual borrowers with lenders. Nearly all market participants can borrow and lend directly on the Internet with few information barriers. The number of investors on online P2P lending platform has increased from 172 thousands in 2014 to 2861 thousands in 2016. While the booming P2P lending service have contributed to market-driven financial liberalization, it also incorporates multiple problems, from managers running away to the collapse of platforms in China's biggest fraud cases. For example, an Internet P2P lending platform called Ezubao attracted funds of about 50 billion yuan (7.6 billion dollars) from 900,000 investors since it set up in 2014, and ceased to trade in December 2015 because the company operated as a Ponzi scheme. To avoid that kind of thing from happening again, China has been exploring the potential of peer-to-peer regulation for years [1]. Furthermore, the collapses of well-known platforms due to malpractice, fraud or business failure [2] are raising great concerns about systemic risk, as troubles of systemically important platforms may trigger wide range of failures because of the network of financial exposures among the industry. However, the network of Chinese P2P lending platforms remains unknown.

To fill this vital gap, we performed the first complex network analysis (to the best of our knowledge) of the Chinese P2P lending platforms using empirical platform-level data. We argue that platforms with overlapped consumers, identical business models or congruent corporate controls would have correlative performance and thus can be connected in the established network of P2P market. Therefore, the connections among the platforms could be their similarity of operating performance. In this paper, eleven indictors of operating performance are carefully collected from P2P lending platforms, such as transaction volume, number of investors, number of borrowers, average interest rate, and so on at monthly frequency. The dataset possesses a sample size of 691, and the sampling period starts from March 2016 to February 2017, which makes it a well-representative sample to investigate the P2P platforms system. We establish the distance matrix of the P2P platform's network by dynamic time warping (DTW), as the observations of each platform varies. Based on the network of P2P platforms, we pick out those that are 'important' to the network system using the language of network science, such as degree centrality, betweenness centrality, etc. [3]. Detecting the systemically important platforms in advance, the regulation department could take precautions against the collapsing of

978-1-5386-5178-0/18/$31.00 ©2018 IEEE

them to avoid tremendous lossesand that's good for investors, borrowers, and the companies themselves.

The rest of this paper is organized as follows. Section II reviews the related works in recent years and further point out our motivations. The data sets are detailedly introduced in Section III. Section IV depicts our presenting approaches for networking method, and the importance indicators including degree centrality, betweenness centrality, closeness centrality, eigenvalue centrality, and a cascading scheme based indicator. Section V shows the systemic important platforms under different important measures, and a further grouping analysis presents the different characteristic of the platforms with different importance level. Finally we briefly conclude the paper and review the limitations in Section VI.

II. Related works

The studies in the field of P2P online lending in China mainly focus on the business model [4], [5], regulation manner [1], the detection of individual fraud cases [6]. There are few that pay attention to the systemic risk among P2P online platforms. Recent studies in banking systems and financial systems take rather different account of systemic risk. Instead of focusing on the individual institution intending to minimize risk and maximize profit, more attention was paid to the potential effects on the system-wide stability [7]–[12]. On one hand, the collapses of well-known platforms are raising great concerns about systemic risk, as troubles of systemically important platforms may trigger wide range of failures because of the network of financial exposures among the industry. On the other hand, detecting the systemically important platforms in advance will benefit the regulation department for taking precautions against the collapsing of them to avoid tremendous losses.

III. Data sets

From the beginning of 2016, we have kept record Internet finance platforms operating performance data from WDZJ (The data source website is http://www.wdzj.com/.), the most powerful third-party information web portals of Internet finance platforms. In this paper, the sampling period starts from March 2016 to February 2017, covering 691 online finance platforms. The collected data contains 11 indicators of operation performance, shown in Table I.

Note that the data is not a complete records of all the investigated Internet finance platforms for the following two reasons: (1) as most of the online finance platform are not listed company, they have no obligation to public consecutive operating performance data. The most extreme case is that some platforms just appear once in the whole dataset, providing only one-month information; (2) some platforms get into trouble and become a so-called 'problematic platform'. The platform will then be closed so the subsequent data are missing. Fig. 1 shows the frequency that platforms appear in our dataset. In addition, it should be noted that there are 5761 records on 634 all-time normal platforms while only 216 records of 58 problematic platforms. In preparation, once the platform become a problematic one, all the records of it will be labeled as problematic. For the convenience to be aware of the difference availability for the normal platforms and problematic platforms, frequency histograms are drawn separately in Fig. 1.

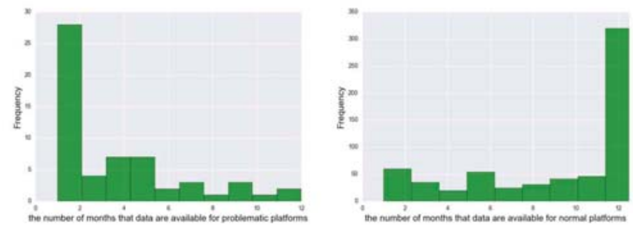

Fig. 1. Histogram of platforms contained in the data set.

As can be seen from Fig. 1, for the normal platforms, there are over 300 platforms that have a full-sample period records. However, a half of the problematic platforms are available only once, either in the month that it is announced as the 'problem or in one of the former months. The above data descriptions stress that the data is severely imbalanced between normal and problematic platforms, and has a limitation in censoring. Modeling with the data should consider both of the concerns.

IV. Modeling method

A. Distance matrix

The whole procedure to create the distance matrix for the P2P platforms is shown in Fig. 2. The details of each step is given below.

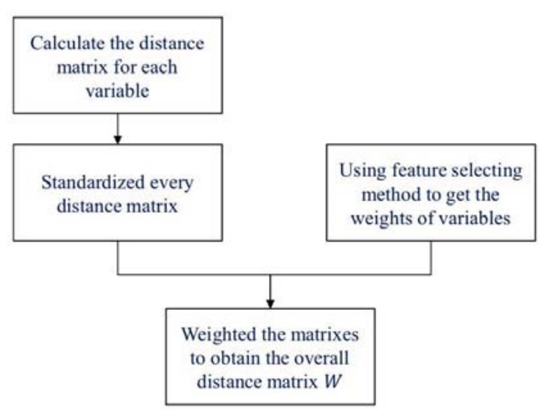

Fig. 2. The procedure to obtain distance matrix.

1) Dynamic time warping distance for every indicator: To deal with the 'imbalance' of lengths for the sample platforms' observations that has been discussed in the previous section, we considering using dynamic time warping (DTW) to obtain the distance between platforms. In time series analysis, dynamic time warping (DTW) is one of the algorithms for measuring similarity between two temporal sequences, which may vary in length. The method started being used by the

TABLE I
DESCRIPTION STATISTICS OF INDICATORS.

Indicator	Notation	Description	weight
Average amount of money from borrowers (million)	ave_out	Equal to the total amount of lending divided by the number of borrowers. From the view of a P2P platform, the money flows out form the platform to the borrowers.	0.087
Average amount of money from lenders (million)	ave_in	Equal to the total amount of lending divided by the number of lenders.	0.090
Number of borrowers	out_per	The number of borrowers during the observation period.	0.070
Number of lenders	in_per	The number of lenders during the observation period.	0.094
Number of underlying lending proposals	out_num	The number of lending or borrowing cases claimed on the platform during the observation period.	0.079
Top10 borrowers balance (million)	top10_out	The amount of money to be paid for the top 10 borrowers who borrow the most.	0.090
Top10 lenders balance (million)	top10_in	The amount of money to be receive for the top 10 lenders who borrow the most.	0.116
Average lending term	ave_period	Average lending term of all the subject of lending on the platform.	0.093
Average interest rate	ave_inter	Average annual interest rate of all the subject of borrowing on the platform.	0.121
Transaction volume (million)	volumn	Total amount of money borrowed from the platform during the observation month, reflects the activity level of one platform. The total lending and borrowing amount of money is the same.	0.080
Time of a tender to be fulfilled	full_time	How much time on average for the lending or borrowing cases to be fulfilled. It could reflect both activity and attraction level of one platform.	0.078
Platform Status	status	Whether it is a normal platform or a trouble platform.	/

data mining community to overcome some of the limitations associated with the Euclidean distance [13]. In a word, DTW is a dissimilarity measurement based on the shapes of time series [14]. The sequences are "warped" non-linearly in the time dimension to determine a measure of their similarity independent of certain non-linear variations in the time dimension. Considering x and y as the input series, where x has length of n and y has length of m. The first step in DTW involves creating a local cost matrix $lcm_{n\times m}$, which is created for every pair of distances compared. The elements in $lcm_{n\times m}$ for x and y are obtained by

$$lcm_{ij} = (\sum_{v} |x_i^v - y_j^v|^p)^{1/p}. \tag{1}$$

The second step in DTW involves finding the path that minimizes the alignment between x and y by iteratively stepping through the lcm. In each round of the procedure, the algorithm find the direction that cost increases the least. Denote $\phi = \{(1,1),...,(n,m)\}$ as the set containing all the points that fall on the optimum path, then the final distance would be computed by

$$DTW_p(x,y) = (\sum \frac{m_\phi lcm(k)^p}{M_\phi})^{1/p}, \forall k \in \phi, \tag{2}$$

where m_ϕ is a per-step weighting coefficient and M_ϕ is the corresponding normalization constant. The way that the algorithm traverses through the lcm is primarily dictated by the chosen step pattern. The step pattern is a local constraint that determines which directions are allowed when moving ahead in the lcm as the cost is being aggregated, as well as the associated per-step weights. In this paper, we are using the $TSdist$ package in R to determine the DTW distance among the time series.

2) Standardized distance matrix for every indicator: By applying dynamic time warping method, the distance matrix of each indicator is obtained. That is, we have eleven distance matrices in total. To eliminate the scale effect, we apply standardization for every distance matrix of the eleven indicators as follows. The elements in one weight matrix are scaled by the maximum and minimum of the weight matrix, as shown in equation (3).

$$w_{ij} = \frac{w_{ij} - min(w_{ij})}{max(w_{ij}) - min(w_{ij})}. \tag{3}$$

3) Weighting the distance matrices: As the eleven indicators yield eleven distance matrices, it is necessary to aggregated them in to a single distance matrix according to the importance of indicators. In the dataset, we have $status$ indicating whether the platform is a good platform or a problematic one. Therefore, we could use $status$ as the label of the platform to learn which operating performance indicator is more important to the status of the platform. The issue is then converted to classification modeling. There are lots of classification methods available to deal with our scenario. Specifically, we use the Extra-Tree method (standing for extremely randomized trees) [15] since the algorithm is computational efficiency and of high accuracy.

The input features of the Extra-Tree classifier is the eleven indicators of operation performance and the label of the Extra-Tree classifier is $status$, which is a dummy variable, with 1 as problematic platform and 0 as good platform. One of the classifier's outputs that is of our interest is the eleven coefficients for indicators. The coefficients are the feature importance measurement that we are looking for. More algorithm related details of Extra-Tree method could be found in Geurts (2006) [15]. In this paper, we use the $sklearn$ package in $Python$ to conduct the procedure. The feature coefficients (weights for the distance matrices) are shown in the last column of Table I. We then applied weighs to the eleven standardized distance matrices by feature selection coefficients in order to produce accurate estimates of the platforms similarities.

4) The distance distribution: After weighting the eleven distance matrices, we aggregate them into a single distance matrix. The distribution of the aggregated distance is shown in Fig. 3, which exhibits obvious right skewness. This indicates that most platforms are similar in the aspects of operating performances as their distance to each other are low, while some deviate from others as they has long distance from other.

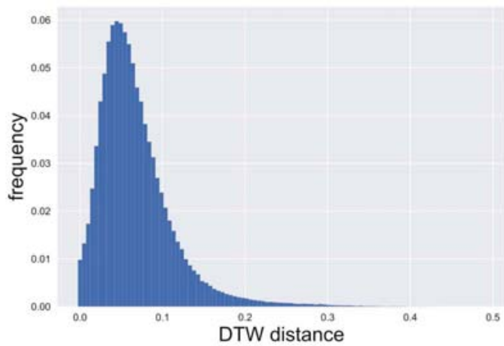

Fig. 3. The distance distribution.

B. Networking method

The network is formed according to the aggregated distance matrix. To neglect those unimportant edges while keep the network connected, we use minimum spanning tree(MST) to extract the bone of the network. This technique has been applied in previous studies and achieve well-defined results [16]–[19]. After the MST extraction, the network possesses 691 nodes and 690 edges, as shown in Fig. 4.

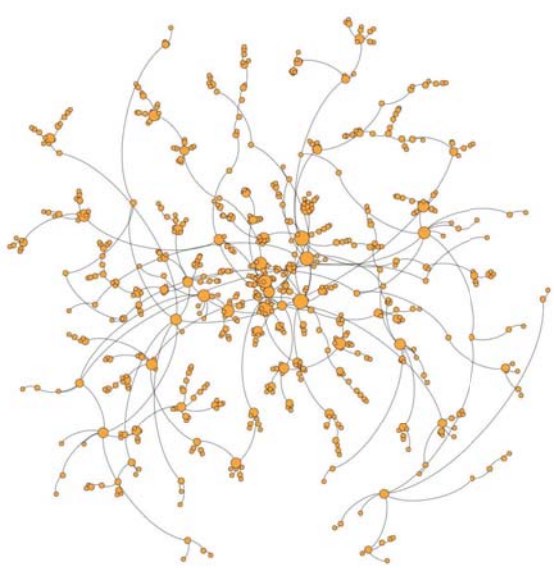

Fig. 4. Network visualization. The size of a node is proportional to its out-degrees.

Usually, for assortative network, nodes of comparable degree tend to link to each other: small-degree nodes to small-degree nodes and hubs to hubs. For disassortative networks, hubs tend to connect to small-degree nodes and small-degree nodes to hubs. The higher the degree assortativity coefficient, the more disassortativity the network is. The present network possesses a degree assortativity coefficient of -0.14, which means the slightly dissasortativity property.

C. Importance measurements

The importance of nodes in a network have various measurement in different aspects.

1) Degree centrality: It is commonly believed that a node is important if it has many neighbors. Intuitively, degree is a simple measure that counts the number of neighbors a node possess in a graph. Degree centrality considers the standardization of degree as

$$DC_i = \frac{k}{N-1}, \tag{4}$$

where k is the degree of node i and N is the number of nodes of the graph. We argue that the collapse of node with high degree will bring about risk to those platforms that connected to the it.

2) Betweeness centrality: Betweenness centrality measures the centrality of a node based on shortest paths [20]. For every pair of vertices in a connected graph, there exists at least one shortest path between the vertices such that either the number of edges that the path passes through (for unweighted graphs) or the sum of the weights of the edges (for weighted graphs) is minimized. The betweenness centrality for each vertex is the number of these shortest paths that pass through the vertex. Denote g_{st} as the number of shortest path from node s to t, among which there are n_{st}^i paths that go through node i. The betweeness centrality is defined as

$$BC_i = \sum_{s \neq i \neq j} \frac{n_{st}^i}{g_{st}}, \tag{5}$$

3) Closeness: In a connected graph, closeness of a node is also a measure of centrality in a network, calculated as the sum of the length of the shortest paths between the node and all other nodes in the graph. Thus the more central a node is, the closer it is to all other nodes. It was defined by [21] as

$$CC_i = \frac{1}{d_i} = \frac{N}{\sum_{j=1}^{N} d_{ij}}, \tag{6}$$

where d_{ij} is the distance between node i and node j, and d_i is the average distance of the paths from node i to other nodes in the network.

4) Eigenvector centrality: Unlike the previous measures of centrality, the eigenvectior centrality accounts for the node's importance as well as its neighbors'. Denote the adjacency matrix of the graph as $A = (a_{ij})$. The relative eigenvector centrality of node i can be defined as

$$EC_i = c \sum_{j=1}^{N} a_{ij} x_j \tag{7}$$

where c is a constance.

One could solve the corresponding eigenvector equation $\mathbf{Ax} = c\mathbf{x}$ to obtain the EC_i. This measure has inspired many follow-up useful indicators such as Google's PageRank and the Katz centrality [22].

5) SIR based importance: An SIR model is an epidemiological model that was proposed to explain the rapid rise and fall in the number of infected patients observed in epidemics [23]. This model is reasonably predictive for infectious diseases which are transmitted from human to human, and where recovery confers lasting resistance. A network-based SIR model assumes that the infection only happen to those who connected to the nodes with disease. For a specific disease in a specific population, these functions may be worked out in order to predict possible outbreaks and bring them under control.

In our scenario, the network of P2P platforms have already demonstrated the interplays among platforms. Our aim is to measure the extent of cascading effect for a given platform's collapse. We design the following simulation procedure the get, what we call it 'SIR based importance'.

- Step1: Initial collapse.
  Denote N as the total number of nodes, S as the number of susceptible nodes, I for the number of infectious nodes, and R for the number of recovered nodes. At the beginning, one node is chose to be infected, or say, one platform is collapsed. Set the time as $t = 0$.
- Step2: Collapse spreading.
  The collapsed node would spread 'disease' to its neighbors. Denote the infection strength as b, where $b \in [0, 1]$. For each neighbor, generate a random number a from the uniform distribution $a \sim U(0, 1)$, and if $a < b$, the neighbor is infected. Otherwise, it is still susceptible. Set the time as $t = t + 1$ after the one-hop cascading.
- Step3: Platform recovering.
  Denote the recovering strength as g. For every infected platform, a specific recovering time r is assigned to it, where r is generated from exponential distribution $r \sim exp(g)$. And when the simulation come to time $t+r$, the infected node will be set as recovered. Once the node is recovered, it would not be infected again.
- Step4: Update S, I, R.
  Count the numbers of susceptible nodes, infected nodes, and recovered nodes. If there is newly infected node, the simulation goes on, that is, we jump back to Step 2 as the newly infected nodes would again spread the 'disease'.
  On the contrary, if there's no more newly infected nodes, we stop the simulation, and get the SIR-based importance measure $importance_{SIR} = I/N$, that is, the final infection ratio of the system. Then, jump back to Step1 to start another round of simulation with the same initial collapse. For every initial collapse, the simulation repeats 100 times to get the mean.

A higher $importance_{SIR}$ indicates the initial collapse of the platform would bring about a more severe systematic failure, thus, the platform is more important.

V. Results

To single out the P2P platforms that are systemically important, four centrality measures DC, BC, CC, EC, that reflect different aspects of importance are computed, as well as one index formed by cascading regime, which we call it $importance_{SIR}$. We rank the platforms by the order from important to unimportant for the five importance measurements, respectively, as shown in Fig. 5.

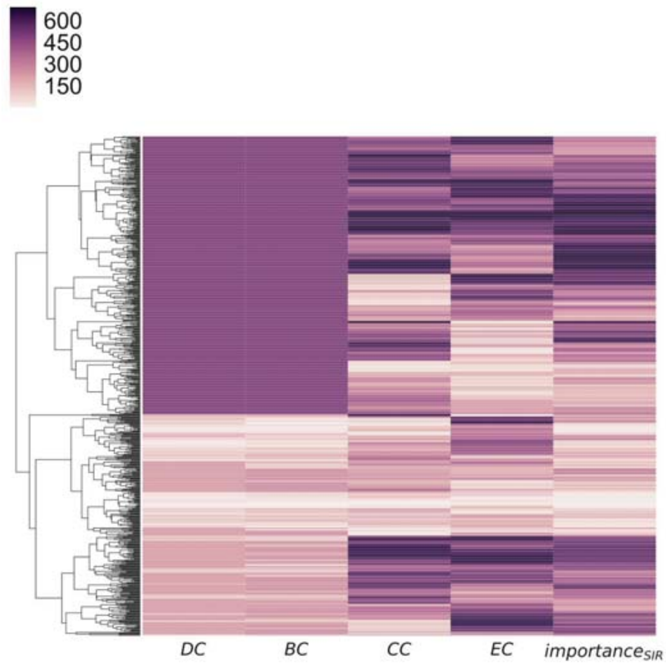

Fig. 5. Nodes importance of online P2P platforms revealed by five indicators.

In Fig 5, each column represents one indicator, as where each row represents one platform. The lighter the color is, the more importance the platform is. The left side of Fig 5 shows the hierarchy clustering results using the Euclidean distance of the five importance measurement. The basic process of hierarchical clustering (defined by S.C. Johnson in 1967 [24]) is that 1) Start by assigning each item to a cluster. We have 691 platforms, so now there are 691 clusters, each containing just one item. 2) Find the closest (the least Euclidean distance) pair of clusters and merge them into a single cluster, so that there are one cluster less. 3) Compute the Euclidean distances between the new cluster and each of the old clusters. 4) Repeat steps 2) and 3) until all items are clustered into a single cluster of size 691. To make the clustering results clear, we extract the clustering results and present it in Fig. 6.

Without the loss of generality, we cut the clustering tree with groups to be five. The five groups are exhibited in different colors. From left to right in Fig. 6, we call them 'green group', 'red group', 'blue group', 'purple group', and 'yellow group'. The operating performances of the four groups are shown in Fig. 7 (the blue group is rather small), with color being accordance with the names of the groups.

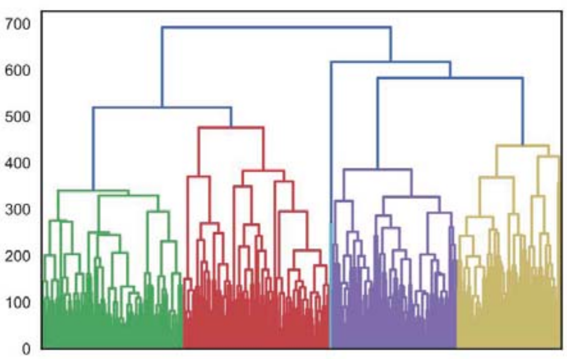

Fig. 6. Cluster.

As can be seen in Fig. 6 and Fig. 5, the green group are those with least importance (rank over 400) in every aspect of importance measurements. The green group possesses 138 platforms, with 10 of them are problematic. In Fig. 7(d), we find that these platforms act at the average level in most operating dimensions, while possess higher average amount of money from borrowers (ave_out) and average amount of money from lenders (ave_in).

The red group is similar with the green group in DC and BC, implying that they both have low degrees and connections in the platforms network. However, the red group is of higher importance in aspects of CC, EC, $importance_{SIR}$. This indicates that those platforms may connected to those of high importance. The red group possesses 194 platforms, with 24 of them are problematic. Fig. 7(b) shows that the blue group of platforms has obvious higher number of underlying lending proposals(out_num) and number of borrowers(out_per) than the average level.

Moving on to the purple groups, it is shown in Fig. 5 that the group is of high importance in every aspect of importance measurement, whereas the radar plot of Fig. 7(c) tells that the purple group has lower values in average amount of money from lenders (ave_in), average amount of money from borrowers(ave_out), number of underlying lending proposals(out_num), number of borrowers(out_per). The purple group possesses 166 platforms, with 22 of them are problematic.

Last but not the least, the yellow group which has high importance evaluations in DC and BC, however, has low importance evaluations in CC, EC, $importance_{SIR}$. The yellow group (Fig. 7(a)) behaves abnormal in transaction volume (volumn), number of borrowers(out_per), number of underlying lending proposals(out_num). It possesses 138 platforms, with 10 of them are problematic.

On the top of the four groups, the purple group is the most importance one, given the leading status in the importance measurements of the platform network. Tracing back to the operating performance indicators, it is further shown that the purple group has the most abnormal behaviors in several operating indicators. Specifically, it possesses lower 'energy' in lending and borrowing activities going on the group's platforms, implying these platforms may suffer from ill-run problems so that few borrowers seeking money and few lending proposal could not be fulfilled. The further investigation of the proportion of problematic platforms in every group also shows that the purple group has the highest ratio. This in turn has proved that the capacity of the five importance indicators in detecting the systematic important and problematic platforms.

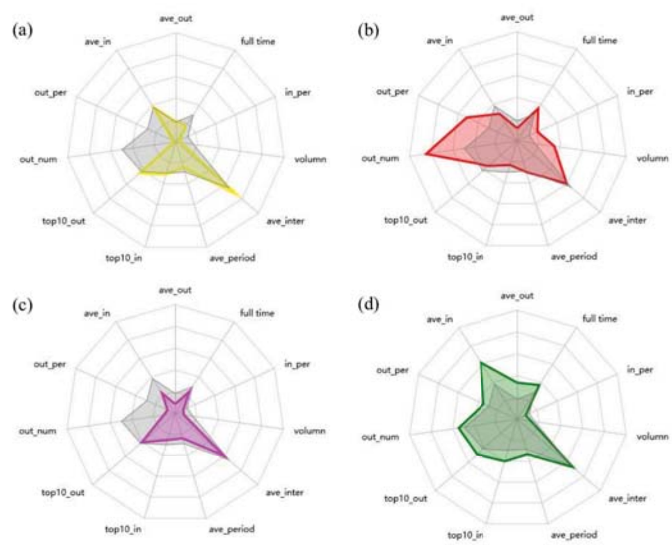

Fig. 7. Characteristic of four groups of platforms. The gray line is the average level of 691 platforms.

VI. Conclusion

Our approach of similarity measures plus network analysis provides a mean of detecting 'systemically important P2P platforms. Overall, two main findings are disclosed in our results. Firstly, a small group of platforms have priority in every angle of importance, and some only domain in one or two indicators and, secondly, when looking close to those 'important nodes, it is shown that they have abnormal operation performance.

This study inevitably has limitations, for example, whether the systemic important measurements could work in prediction is still not investigated because the limitation of longitude of observation period. Nevertheless, our results could help to offer investing advice to investors and regulating priority for macro prudential supervision.

Acknowledgement

This research was financially supported by National Natural Science Foundation of China (Grant Nos. 71420107025 and 71501005).

References

[1] W. Xiaoguang and C. Yi, "The regulation strengthening of the p2p online lending platform," *South China Finance*, vol. 4, p. 009, 2011.

[2] Z. Tania, R. E.J., L. Annie, Z. Bryan, S. K. Randall, and G. Kieran, "The 2017 americas alternative finance industry survey," 2017.

[3] A.-L. Barabási, *Network science*. Cambridge university press, 2016.

[4] Q. J.-y. Y. Fei, "The development status and prospects of chinese p2p network lending," in *Finance Forum*, vol. 1, 2012, p. 009.

[5] D. Chen and C. Han, "A comparative study of online p2p lending in the usa and china," *Journal of Internet Banking and Commerce*, vol. 17, no. 2, p. 1, 2012.

[6] R. Emekter, Y. Tu, B. Jirasakuldech, and M. Lu, "Evaluating credit risk and loan performance in online peer-to-peer (p2p) lending," *Applied Economics*, vol. 47, no. 1, pp. 54–70, 2015.

[7] F. Schweitzer, G. Fagiolo, D. Sornette, F. Vega-Redondo, A. Vespignani, and D. R. White, "Economic networks: The new challenges," *Science*, vol. 325, no. 5939, pp. 422–425, 2009.

[8] A. G. Haldane and R. M. May, "Systemic risk in banking ecosystems," *Nature*, vol. 469, no. 7330, pp. 351–355, 2011.

[9] X. Zhang, L. Feng, Y. Berman, N. Hu, and H. E. Stanley, "Exacerbated vulnerability of coupled socio-economic risk in complex networks," *EPL (Europhysics Letters)*, vol. 116, no. 1, p. 18001, 2016.

[10] M. Bardoscia, S. Battiston, F. Caccioli, and G. Caldarelli, "Pathways towards instability in financial networks," *Nature Communications*, vol. 8, p. 14416, 2017.

[11] M. Billio, M. Getmansky, A. W. Lo, and L. Pelizzon, "Econometric measures of connectedness and systemic risk in the finance and insurance sectors," *Journal of Financial Economics*, vol. 104, no. 3, pp. 535 – 559, 2012.

[12] T. Roukny, H. Bersini, H. Pirotte, G. Caldarelli, and S. Battiston, "Default cascades in complex networks: Topology and systemic risk," *Scientific Reports*, vol. 3, pp. 2759 EP –, 2013.

[13] D. J. Berndt and J. Clifford, "Using dynamic time warping to find patterns in time series." in *KDD workshop*, vol. 10, no. 16. Seattle, WA, 1994, pp. 359–370.

[14] S. Aghabozorgi, A. S. Shirkhorshidi, and T. Y. Wah, "Time-series clustering–a decade review," *Information Systems*, vol. 53, pp. 16–38, 2015.

[15] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Machine learning*, vol. 63, no. 1, pp. 3–42, 2006.

[16] R. N. Mantegna, "Hierarchical structure in financial markets," *The European Physical Journal B - Condensed Matter and Complex Systems*, vol. 11, no. 1, pp. 193–197, 1999.

[17] G. Bonanno, G. Caldarelli, F. Lillo, and R. N. Mantegna, "Topology of correlation-based minimal spanning trees in real and model markets," *Physical Review E*, vol. 68, no. 4, p. 046130, 2003.

[18] H.-J. Kim, Y. Lee, B. Kahng, and I.-m. Kim, "Weighted scale-free network in financial correlations," *Journal of the Physical Society of Japan*, vol. 71, no. 9, pp. 2133–2136, 2002.

[19] H. Kim, I. Kim, Y. Lee, and B. Kahng, "Scale-free network in stock markets," *Journal-Korean Physical Society*, vol. 40, pp. 1105–1108, 2002.

[20] L. C. Freeman, "A set of measures of centrality based on betweenness," *Sociometry*, pp. 35–41, 1977.

[21] A. Bavelas, "Communication patterns in task-oriented groups," *The Journal of the Acoustical Society of America*, vol. 22, no. 6, pp. 725–730, 1950.

[22] M. E. Newman, "The mathematics of networks," *The new palgrave encyclopedia of economics*, vol. 2, no. 2008, pp. 1–12, 2008.

[23] M. LeChevallier, W. Norton, and R. Lee, "A contribution to the mathematical theory of epidemics," in *R. Soc. Lond. Proc. Ser. A Math. Phys. Eng. Sci*, vol. 115, 1995, pp. 48–86.

[24] S. C. Johnson, "Hierarchical clustering schemes," *Psychometrika*, vol. 32, no. 3, pp. 241–254, 1967.