

## Article

# Unveiling Latent Structure of Venture Capital Syndication Networks

Weiwei Gu, Ao Yang <sup>\*</sup>, Lingyun Lu and Ruiqi Li <sup>\*</sup>

UrbanNet Lab, College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China

<sup>\*</sup> Correspondence: ya@buct.edu.cn (A.Y.); lir@buct.edu.cn (R.L.)

**Abstract:** Venture capital (VC) is a form of private equity financing provided by VC institutions to startups with high growth potential due to innovative technology or novel business models but also high risks. To against uncertainties and benefit from mutual complementarity and sharing resources and information, making joint-investments with other VC institutions on the same startup are pervasive, which forms an ever-growing complex syndication network. Attaining objective classifications of VC institutions and revealing the latent structure of joint-investment behaviors between them can deepen our understanding of the VC industry and boost the healthy development of the market and economy. In this work, we devise an iterative Loubar method based on the Lorenz curve to make objective classification of VC institutions automatically, which does not require setting arbitrary thresholds and the number of categories. We further reveal distinct investment behaviors across categories, where the top-ranked group enters more industries and investment stages with a better performance. Through network embedding of joint investment relations, we unveil the existence of possible territories of top-ranked VC institutions, and the hidden structure of relations between VC institutions.

**Keywords:** syndication network; venture capital; network embedding; investment behaviors; objective classification



**Citation:** Gu, W.; Yang, A.; Lu, L.; Li, R. Unveiling Latent Structure of Venture Capital Syndication Networks. *Entropy* **2022**, *24*, 1506. <https://doi.org/10.3390/e24101506>

Academic Editors: Xiangyun Gao, Feng An, Qingru Sun and Stanisław Drożdż

Received: 28 August 2022

Accepted: 11 October 2022

Published: 21 October 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Venture capital (VC) is typically defined as the investment made by professional institutions in unquoted startup companies with long-term financial growth potential, high risks but also possible enormous returns [1]. Over the past 30 years, VC has been an important financing source for early-stage and emerging companies with innovative technology or business models. Besides financial support, VC institutions can also provide a startup or young business with valuable sources of guidance, consultation, and bring tremendous benefits by tapping into domestic business communities via knowledge spillover effect [2]. China now receives over 40 percent of global VC investments [3], and more and more Chinese startups and urban economies are benefiting from VC investments [4,5]. A variety of VC-backed companies, such as Tencent, Alibaba, and Baidu, have had a huge influence on China and the worldwide economy.

Making joint-investments with other VC institutions, which forms an ever growing syndication network with complex structure, is a good way to share information [6], resources [7] and benefit from mutual complementarity [8], diversity [9,10] to against uncertainties [11–14].

VC institutions are bounded by their current and past syndication, i.e., joint investments with others. Besides identifying leading VC institutions, the syndication network of VC institutions can be used to predict the alliance's exit [15], quantify the social capital, evaluate the reputation of VC institutions [16], and predict the portfolio failure rate [17]. Unveiling the latent structure of VC syndication networks and investment behaviors is

crucial for deepening our understanding of the VC industry and boosting the healthy development of the market and economy. However, VC syndication network encodes nodal interactions but neglects other investment behavior information, such as the number of IPO (Initial Public Offering), M&A (Merge and Acquisition), and the number of investments, and hidden richer higher-order structure should be better revealed, all of which are important for VC analysis.

Previous studies indicate that there is a leader-follower phenomenon in the Chinese VC market, for example, a few leading VC institutions have better access to high-quality startups and they usually set up investment plans, while other VC institutions have a higher tendency to follow [18,19]. When young inexperienced VC institutions make joint investments with leading VC institutions of great reputation, they would have a smaller chance to renege or shirk [20], and usually can gain a better position for competition by building a word-of-mouth reputation connections [20,21]. And leading VC institutions also can act as bridges for resources and investment skills exchange, as they are usually hub nodes in the syndication network [19]. The ecosystem of VC market is way more complicated than a leader-follower structure, but most previous studies only focus on identifying leading VC institutions [22–24], and they usually do not have an objective criterion for identifying other groups of VC institutions. For example, Batjargal et al. [22] analyzed empirical data to reveal investment patterns associated with leading VC institutions. Huang et al. [23] found the triadic closure probability is higher if a leading VC institution is in it. Yet, the above works only used the terminology of “leading VC institutions”, but the classification is determined according to subjective criteria or practical experience from VC practitioners. Recently, a few works are dedicated to identifying leading VC institutions from joint-investment networks. The task is then equivalent to determine the most influential nodes via various centrality measurements in complex network analysis [25–29]. Yang et al. [24] employed a weighted k-means algorithm to identify leading VC institutions based on various centrality measurements. With the aid of graph convolutional deep learning model, Gu et al. [30] discovered that the attention matrix can be applied to rank VC institutions. However, most related works are based merely on network centrality measurements or partial nodal interactions, and these works pay more attention to leading VC institutions but do not make a classification of remaining institutions. Besides, they also neglect the richer higher-order structure and other investment behavior information such as IPO, M&A, and investment frequency, which hinders gaining a more comprehensive understanding of the VC market. To the best of our knowledge, few works are focusing on the objective classification of VC institutions and further ranking of obtained groups.

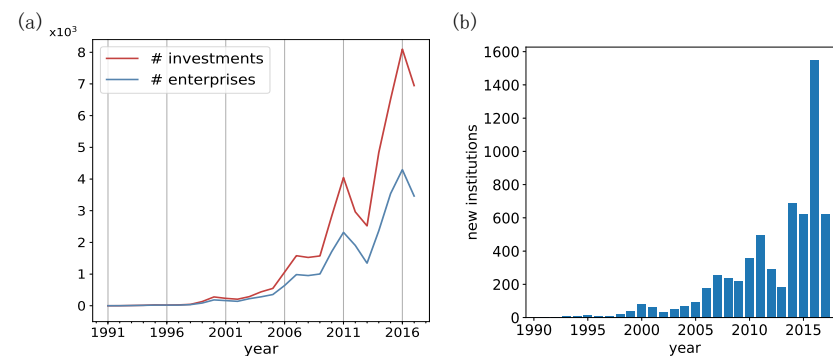
Starting a new VC institution is an activity that entails a high level of risk, both success, and failure cases are valuable. Failure is also an important characteristic of the entrepreneurial reality though it is also been viewed from a negative perspective [31]. In this paper, we propose an iterative Loubar method [32–34] based on Lorenz curve [35] to make objective classification of successful as well as unsuccessful VC institutions based on their principled features, which do not require setting arbitrary thresholds to separate different categories and the number of categories can be automatically determined. The features fed to our method include the number of IPO (Initial Public Offering)&MA (Merge and Acquisition) and the number of investments of VC institutions, which are important indicators of investment performance [12].

The systematical differences between positions in the VC syndication network and investment behaviors among VC institutions in each category prove the validity and practicability of our method. VC institutions in top-ranked categories generally enter more industries and get involved in more investment stages, have a higher network centrality, and have better investment performance. In addition, through the network embedding techniques [36], we unveil a latent structure of different categories of VC institutions, where top-ranked VC institutions have their territories. Our study can benefit on developing better investment strategies and portfolio optimization.

## 2. Data Preprocessing

In this paper, we conduct the study based on the SiMuTong database that records detailed information about each investment, including the VC institution that made the investment, the startup company that got invested, investment date, investment amount, currency, share, investment stage, and some basic information about the startup, for example, its industry and the location of its headquarter. We remove investment records without information about the investor or startup. And, for better consistency, we map the industry type of startup companies provided by Simutong to the ones in the Industrial Classification for National Economic Activities of China (ICNEAC). In addition, we convert the investment amount that was in foreign currency into RMB.

In the western world, the VC industry emerged in 1946 and experienced its first boom and bust cycle from 1982 to 1993 [12]. While in China, VC is still a relatively newly emergent industry (see Figure 1a). Before the 1990s, China received few VC investments. After 2000, The size of the annual investment amount and the number of investment activities underwent a rapid increase, and they both manifest a roughly five-year cycle (see Figure 1a). From 2014 to 2017, there was a spur of VC institutions that around four thousand new VC institutions entered the market (see Figure 1b), which further indicates the urgency of developing a quantitative science of VC institutions.



**Figure 1.** Basic statistics of VC investment activities in China: (a) The number of investments and the number of startups got invested in each year; (b) The number of new VC institutions that enter the Chinese market each year.

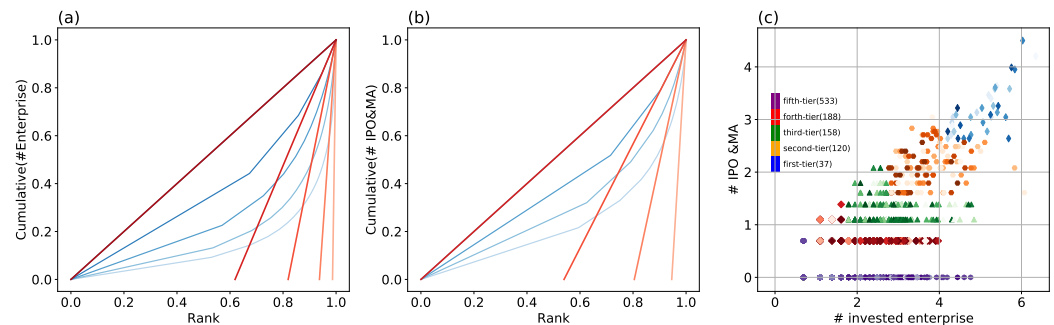
## 3. Iterative Loubar Algorithm for Objective Classification

In the VC literature, traditional ranking algorithms use either the Delphi method based on interviews with practitioners, which are usually limited on the sampling size and influenced by subjective experience, or the network-based centrality measures, which neglect higher-order structure. For example, for the Delphi method, the number of experts got interviewed is usually quite small, and it lacks an objective evaluation criterion and a clear conflict reconciling mechanism when two or more interviewed experts do not come to a consensus on the same question. Besides, both the Delphi and the centrality-based methods pay disproportionately more attention to leading VC institutions rather than underperforming ones, and they are unable to give an objective classification of VC institutions. In this paper, we devise an iterative Loubar method without setting arbitrary thresholds and the number of categories, both of which can be automatically determined.

The input evaluation indicators of VC institutions fed to the algorithm are the number of investments and the number of IPO&MA. VC institutions that can make a large number of investments usually have substantial funds, which is a manifestation of good financial condition and great social capital; while the number of investments exited through IPO and M&A is the most successful way of exit for VC institutions [12].

Our iterative Loubar method works as follows: First, ascendingly ordering all VC institutions according to their number of investment deals, accumulate their values on the vertical axis, and normalize both axes by the corresponding maximum (i.e., normalize the data by the cumulative number of investments for the vertical axis, and the number of

VC institutions for the horizontal axis). This curve is also named the Lorenz curve [32] (see the lightest blue curve in Figure 2a). Then take the derivative of the Lorenz curve at (1, 1) and extrapolate it to the point at which it intersects the horizontal axis, which gives us an objective threshold to separate the first level of VC institutions from others on this indicator. All VC institutions above this threshold will be classified as temporary level 1 on this indicator. The same process is repeated on the other indicator, here, the indicator is the number of IPO&MA (see Figure 2b), which will also classify some VC institutions into level 1 on this indicator. Eventually, only the VC institutions in both level 1 on the two indicators can be classified as the first category (named “first-tier”, see Figure 2c).



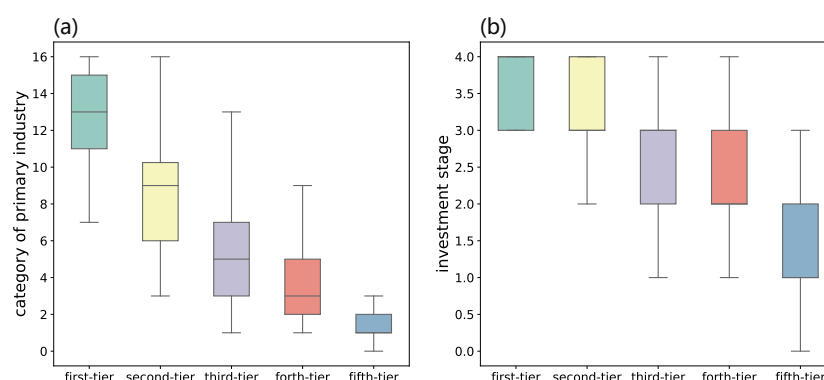
**Figure 2.** The illustration of our iterative Loubar method on classifying VC institutions: The partition based on (a) the number of investments, (b) the number of IPO and M&A. (c) the classification results are based on the number of investment deals and the number of IPO and M&A. The digit in the parentheses indicates the size of the category. The color of the symbols indicates the density of data points, a darker color corresponds to a higher density.

We then exclude VC institutions in first-tier from the Lorenz curve and re-normalize the remaining curve to repeat the previous process for both indicators to obtain VC institutions in level 2 (we name the second category as “second-tier”, see orange hexagon in Figure 2c), which would have the ones in both level 2 on the two indicators and the ones in level 1 on one indicator and in level 2 on the other indicator. If necessary, we can make three subdivisions for the second category. The algorithm goes on until no VC institution is left (see Figure 2a,b), and here we obtain five categories (see Figure 2c), which are also in meaningful ranking from first-tier to fifth-tier as we will illustrate later. By contrast, ordinary clustering methods cannot make ranking for obtained clusters. The Lorenz curves are depicted in progressively transparent shades of blue (from light to dark ones, corresponding to level 1 to level 5 in Figure 2a,b), and the corresponding derivatives at (1, 1) from light red to dark red. Our method can be easily extended to situations with more input indicators. More details and discussions of the iterative Loubar algorithm can be found in Appendix A.

We eventually classified firms into five categories. Those leading VC institutions in the first category are labeled as “first-tier”, the ones in the following are categorized as “second-tier”, “third-tier”, “forth-tier”, and “fifth-tier”, respectively (see Figure 2c). The VC institutions classified into the first category (e.g., Starter Story, Statista, Sequoia Capital China, and China Growth Capital) by our method are in accordance with authority ranking reports, which rank VC institutions based on their capability in earning money, reputation, investment frequency and so on [37]. Compared to the first category, the ones in the second category (orange hexagon in Figure 2c) have fewer investments or IPO and M&A. This group also contains VC institutions with a large number of investments (in level 1 on this indicator) but fewer IPO and M&A (level 2 on the other indicator), and vice versa. The majority of VC institutions in the fifth category (“fifth-tier”, see purple hexagons in Figure 2c) have only a few investments and IPO and M&A exits, but there are some VC institutions that make a large number of investments but end up with quite a few successful exits, which generally have a worse investment performance and much less successful.

#### 4. Investment Behaviors and Performance Analysis

We make further investigations on VC institutions of each category and discover that they have quite different investment behaviors. We first analyze the number of investment deals, IPO and M&A, investment stage, and investment industries for VC institutions in each category. During different investment stages, which consist of seed, initial, expansion, and maturity, startup companies need different resources or guidance from VC institutions. Thus the number of stages that a VC institution got involved in can reflect their ability on the whole investment cycle. While the number of industries that a VC enters is also a manifestation of their knowledge in various fields and the size of funding. As shown in Figure 3, VC institutions in the first category not only enter a wider range of industries but are also involved in more investment stages. There is a clear decrease in such indicators for VC institutions from the fi.

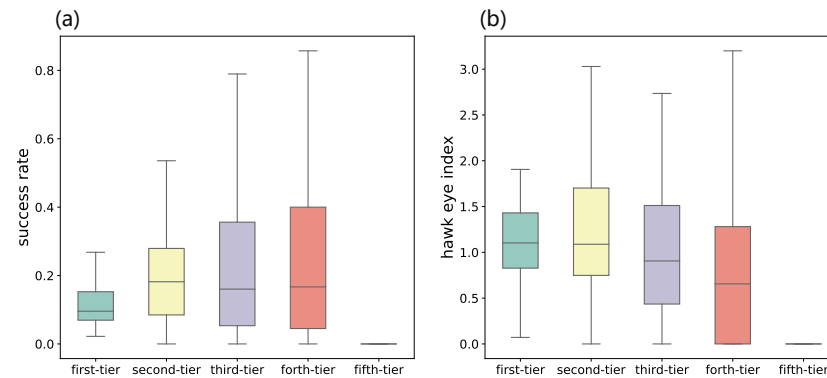


**Figure 3.** Comparisons between VC institutions from different categories on the number of (a) industry, and (b) investment stages that they involved.

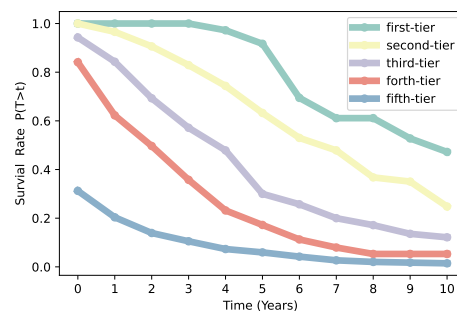
We then evaluate the success rate of VC institutions from different categories by dividing the number of IPO&MA by the total number of invested startups. As shown in Figure 4a, the IPO and M&A rate of VC institutions from the first category is even lower than the ones from second- to forth-tier. This is counter-intuitive since many VC ranking systems give higher ranks to VC institutions based on their high IPO rates [37,38]. Some evidence also suggests that leading VC institutions have a larger fraction of investments that exit through an IPO or M&A [39]. To better investigate the investment performance, we propose a new indicator named hawk eye index (HEI), which is defined as the amount of money invested into the startups that exit through IPO divided by the total amount of money of all investments. It is formulated as follows:  $HEI = (Amt_{IPO} / Amt_{Total}) / (\#investments_{IPO} / \#investments_{Total})$ . As shown in Figure 4b, the median HEI of VC institutions in the first category ranks first and decreases over the remaining four tiers in order. A higher HEI indicates that the VC institution might try out for a larger number of investments, but have a better ability to allocate more resources on deals that are of greater potential and eventually successful.

In addition, VC institutions play an important role in the development of the economy, their survival rate and sustainable development are closely related to economic growth. Here, the survival rate is defined as the proportion of VC institutions that have yet to continue to invest after  $t$  years after their first investment. We calculate the survival rate of VC institutions from different categories obtained by our iterative Loubar classification method. We find that the longevity curves rank in decreasing order in accordance with ordered categories by our iterative Loubar classification algorithm that VC institutions from the first-tier have the highest survival rate and the fifth-tiers have the lowest (see Figure 5). Based on statistical analysis, Govindarajan et al. discovered that nearly 40% of companies cannot survive within the first 5 years [40], which is just an average for the whole market, but the whole system can be quite heterogeneous. From Figure 5, we can observe that over 95% of first-tier VC institutions can survive over 5 years, while more than 90% of fifth-tier VC institutions died in 5 years after their first investment. The variety

of survival rates might be caused by the heterogeneity of VC institutions. The longevity survival curves, which rank in decreasing order for first-tier to fifth-tier VC institutions, also confirm the effectiveness of our iterative Loubar classification algorithm.



**Figure 4.** Comparisons between VC institutions from different categories on (a) the success rate, which is measured by the fraction of invested startups that exit through IPO or M&A, and (b) the hawk eye index (HEI). Note that the first-tier category has a low median success rate but the highest median HEI, which indicates that they tend to invest more money on eventually successful startups out of a large number of alternative choices.

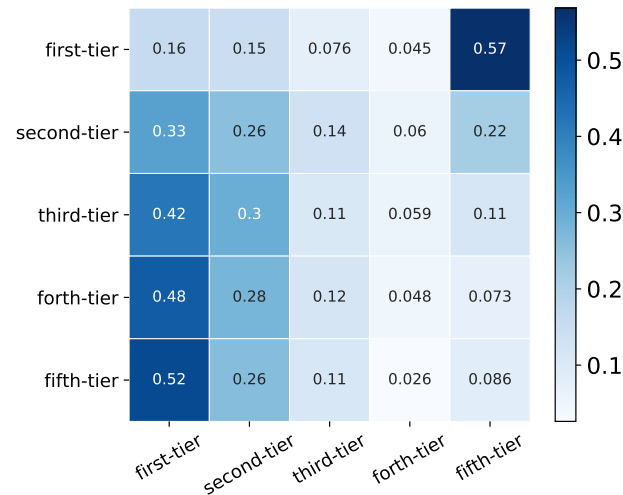


**Figure 5.** Comparisons between VC institutions from five categories on the survival rate.

Syndication between VC institutions is a common investment strategy to share resources, access to better deal flows, reduce risks as well as complement skills [12]. In this paper, we analyze the difference in syndication behaviours of VC institutions from different categories. We first compute the number of joint investments  $W_{ab}$  between VC institutions from categories  $a$  and  $b$ , and we then calculate the total number of joint-investments of each category  $W_a = \sum_b W_{ab}$ . Then the syndication tendency from category  $a$  to  $b$  can be represented as  $W_{ab}/W_a$ , which indicates that each row is subject to normalization, and it is worth noting that the matrix of  $W_{ab}/W_a$  is asymmetric (see Figure 6).

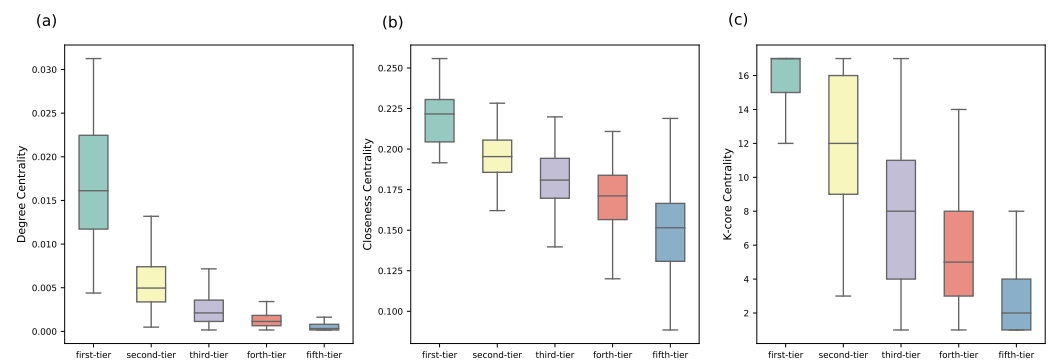
From Figure 6, we can clearly observe that the VC institutions from the first-tier category syndicate more frequently with VC institutions from the same group than with VC institutions from other groups, which indicates a rich-club phenomenon. VC institutions in other categories except for the ones from the fifth-tier category also have the highest tendency to make joint investments with VC institutions from the first-tier category, which are indicated by darker color in the first column of Figure 6. VC institutions in the fifth-tier category, most of which are new entrants, have the strongest tendency to syndicate within its category. This might suggest that resources, better deal flows, and even information is more concentrated in those VC institutions from the first-tier category, and most VC institutions would like to mingle, but certain barriers exist. It is worth noting that VC institutions in every category have the lowest syndication tendency with VC institutions from the forth-tier category (see the fourth column in Figure 6), in comparison, the syndication tendency with VC institutions from the fifth-tier category is even much larger (see the fifth

column in Figure 6), which is roughly at least twice higher than with the forth-tier category. These new VC institutions from the fifth-tier category might be more attractive to other VC institutions to collaborate with, but the forth-tier category might already pass such a new entrance phase and has less potential or resources to revive.

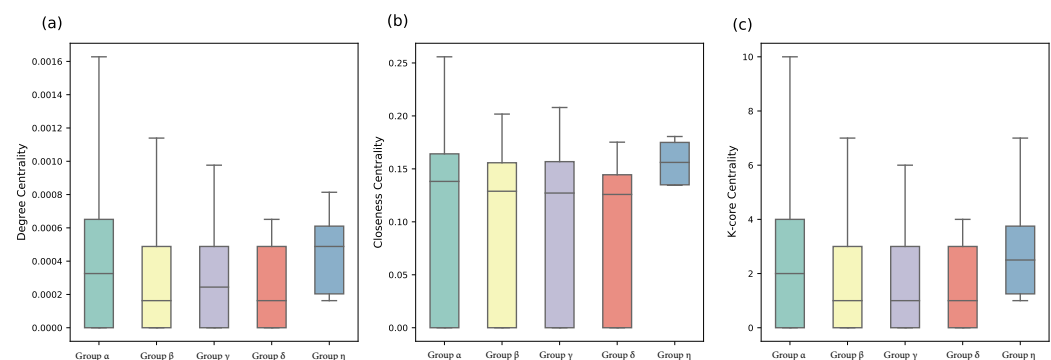


**Figure 6.** Syndication probability  $W_{ab}/W_a$  of VC institutions from different categories. The first-tier group and the fifth-tier group have the highest tendency to make joint-investment with VC institutions within the same category. Each row is subject to normalization, but each column is not subject to normalization.

The syndication patterns can be better revealed through a complex network perspective. If two VC institutions make a joint investment, then we connect these two nodes in the network. We further compare the structural differences of VC institutions from different groups via computing the degree centrality, closeness centrality [41], and  $k$ -core centrality [42]. Degree centrality measures the number of edges that a VC institution has in the joint-investment network. With more ties, a VC institution usually has more investment opportunities and better deal flows [12]. Closeness measures its average distance to all other nodes that VC institutions with a high closeness centrality are averagely closer to other nodes [43]. The coreness (also referred to  $k$ -core) is a more sophisticated centrality measure, which can be obtained by iteratively removing vertices with a degree less than  $k$  [42] or by iterating the  $H$ -index operator [44]. The core-periphery structure indicates a higher-order organization in the VC syndication network. VC with a high  $k$ -core value is generally in a more central position with a stronger influence in the network [42]. We find that the VC institutions from different categories are of typically different values, and the first-tier category are more central than VC institutions from other categories (see Figure 7). From first-tier to fifth-tier VC institutions, there is a clear decreasing trend on all these network indicators, which also confirms the effectiveness of ranking for groups obtained by our iterative Loubar method. We further validate the quality of classification by our algorithm by comparing the structural differences with the algorithm proposed by Yang et al. [24], which uses a weighted  $k$ -means method to identify leading VC institutions and classify VC institutions into different groups. Recent evidence indicates that weighted  $k$ -means is a pretty good clustering algorithm for grouping VC institutions when compared with other clustering methods [24], and thus in this paper, we compare our classification results with the weighted  $k$ -means clustering algorithm. We discover that groups identified by the weighted  $k$ -means algorithm have no significant difference on all three network centrality indicators (see Figure 8). Since network centrality measurements are classical ways of quantifying how important a VC institution is, our algorithm has a better classification performance.



**Figure 7.** The difference between VC institutions of five categories identified by our iterative Loubar algorithm on three network centrality indicators: (a) degree; (b) closeness; and (c) k-core.



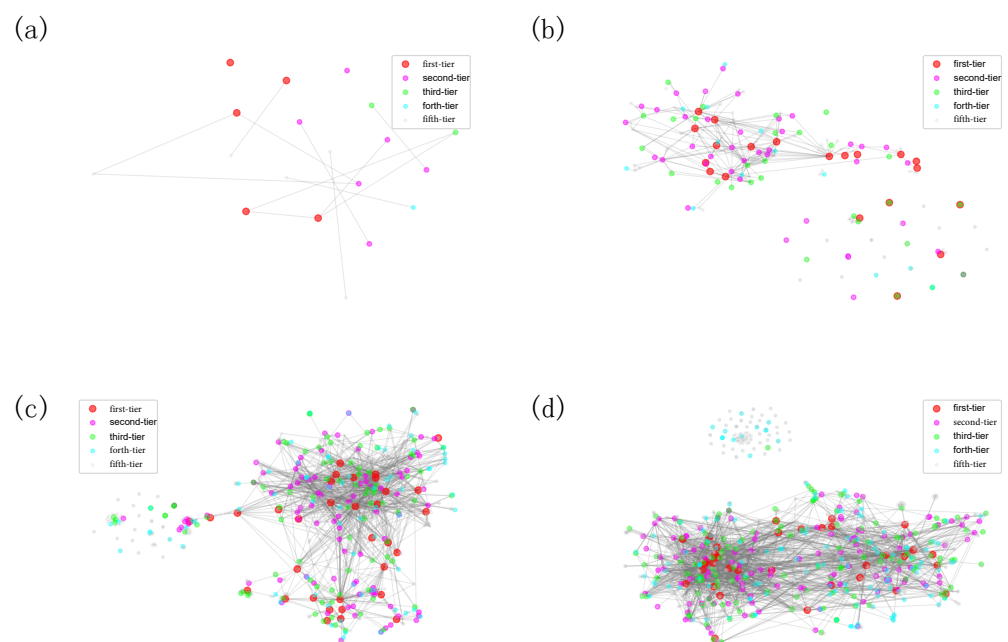
**Figure 8.** The difference between VC institutions of five categories identified by the weighted k-means algorithm [24], which cannot rank obtained clusters, on three network centrality indicators: (a) degree; (b) closeness; and (c) k-core. Note the  $\alpha, \beta, \gamma, \dots$  only denote different groups but not the rank.

## 5. Latent Structure of Joint-Investment Behaviors

To further investigate the phenomena observed above, we exploit network embedding algorithms to better unveil the latent structure of VC institutions in the Chinese VC market. Our embedding method builds upon the DeepWalk architecture [45], which is a direct adaptation of the word2vec [46] model in the context of graphs. Random walk sequences on the syndication network, which are considered as “sentences” in the natural language processing, are fed to the SkipGram model [46] to obtain real-value vector representations for each VC institution. If two VC institutions are of a closer relationship in the syndication network, then they have a close relationship in the embedded metric space. The vector representation can be easily integrated into other computational models to make accurate predictions on joint investment behaviors and to cluster similar VC institutions. Since the embedding vectors of syndication also encode the complex network structure in the investment space, they can also help us have a better view of the evolution of the Chinese VC industry.

In this work, to get their vector representations, based on the accumulative joint-investment networks, we embed VC institutions into a 100-dimensional metric space, which is suggested by the recent advance on determining an appropriate embedding dimension [47]. To visualize the embedding vector, we use t-SNE [48] to project the high-dimensional vector representation into two dimensions. We visualize the syndication evolution process in Figure 9 and we can tell that there are only a few VC institutions that have joint investments before 2000. Compared with the VC institutions from other categories, first-tier and second-tier VC institutions entered the Chinese market at a relatively earlier time (see Figure 9a). But some first-movers might be less successful and fall into lower-ranked categories (see Figure 9a). In 2003, many first-tier VC institutions already take

the central position (see Figure 9b), which indicates that their investments play a key role in the Chinese VC industry. The syndication around first-tier and second-tier VC institutions began forming densely connected communities. VC institutions with fewer syndication events are located far from the majority (See Figure 9b–d), among which we notice several first-tier VC institutions that entered the investment market for the first time. These VC institutions are sort of rising stars that started from the fringe of the network but ended up as a superstar with a central position. They gain good opportunities to syndicate with second-tier and first-tier VC institutions and gradually move to the densely connected elite community (see Figure 9c,d). It is worth noting that there are two separate groups in the Chinese market, one of which is mainly consisted of VC institutions from the forth-tier and fifth-tier categories, which are in accordance with the results presented in Figure 6. The fifth-tier group has a densely connected structure among them, so they might have limited opportunities to source high-quality deal flow (i.e., select promising startup companies) and to nurture investments (i.e., add value to portfolio companies) [12].

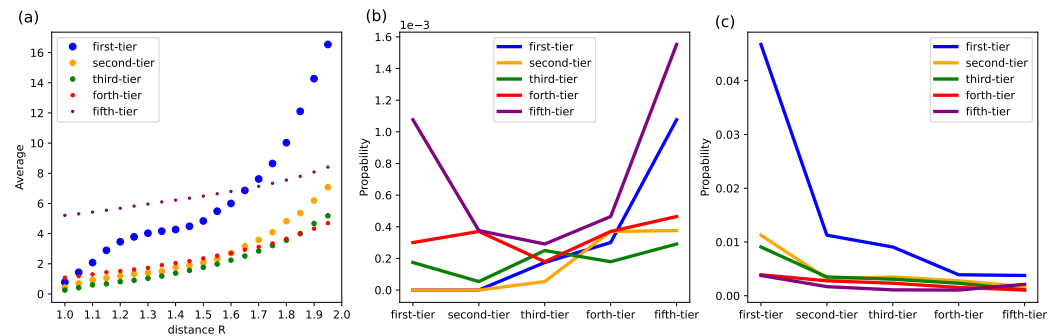


**Figure 9.** The two-dimensional projection of the cumulative syndication networks obtained by t-SNE algorithm in the year (a) 1998, (b) 2003, (c) 2008, and (d) 2011. Each node represents a VC institution with node color indicating the classification category obtained by our iterative Loubar method, and the gray line signifies the syndication relation between VC institutions.

Network embedding not only encodes the syndication relationship between VC institutions but also provides a natural explanation for the architecture of real complex networks with a latent metric space. We measure the geometric properties of VC institutions by quantifying the number of nodes within a certain distance  $R$ .

As shown in Figure 10a, the average number of VC institutions that are in the vicinity of a first-tier VC institution varies significantly across categories. VC institutions in the first-tier category are more central as the average number in their vicinity grows faster when  $R$  is larger. We discover that there is a plateau for a first-tier VC institution to encounter other first-tier ones when  $R$  is around 1.2 to 1.5, that there is no clear significant increase. This suggests the existence of possible territories of first-tier VC institutions that generally a few of them would dominate a certain range, where no more other first-tier institutions can be encountered. We then set  $R = 1.35$ , which resides in the plateau phase, and we find that first-tier VC institutions are surrounded by many fifth-tier VC institutions (indicated by a large fraction of entities in the fifth category, see the first value

of the purple line in Figure 10b), while first-tier VC institutions themselves seem to avoid each other, which is indicated by a near-zero fraction. While fifth-tiers are either closer to first-tier VC institutions or closer to each other. When we look at a wider range, e.g.,  $R = 2$  (see Figure 10c), first-tier VC institutions are still much closer to each other than the rest. This indicates that the first-tier VC institutions cooperate frequently, but they all have some smaller VC institutions that are only closer to themselves, which might form their own territories.



**Figure 10.** (a) The average number of VC institutions in the vicinity of a first-tier VC institution. The horizontal axis is the distance  $R$  in the embedded space and the vertical axis is the average number of VC institutions from different categories in the vicinity of a VC institution. The results are averaged for all VC institutions from the first-tier category. (b) The average fraction of VC institutions from different categories encountered in the vicinity with  $R = 1.35$  and (c)  $R = 2$  for VC institutions from different categories.

## 6. Conclusions

In this paper, we devise an iterative Loubar algorithm to make an objective classification of VC institutions and ranking obtained categories, which does not require setting arbitrary thresholds and the number of categories and can be easily extended to cases with more indicators. We discover that the average performances of VC institutions in identified categories, from the first-tier category to the fifth-tier, on various investment behaviors indicators, largely follow a descending order, which confirms the validity of our algorithm. VC institutions from top-ranked first-tier categories generally enter a wider range of investment industries, involve in more investment stages, as well as have a larger number of syndication with VC institutions from other categories. The only exception lies in the success rate, where those VC institutions in the first-tier category are not high, but we find that they have the highest hawk eye index, which indicates that they might make many tryouts but are able to allocate more resources into startups with great potential. This also suggests that the ordinary success rate, which equals the fractions of exits through IPO and M&A [12], is not a proper indicator to identify leading VC institutions. In addition, we find that most VC institutions have a stronger syndication tendency than VC institutions from the first-tier category. The exception is the fifth-tier category, which has a much stronger tendency to make joint-investment within its category. It is worth noting that VC institutions from every category have the lowest syndication tendency towards the VC institutions from the forth-tier category. The latent structure of the syndication network unveiled by network embedding further confirms our previous findings and suggests the existence of possible territories of first-tier VC institutions that generally a few of them would dominate a certain range, which is indicated by the plateau in Figure 10a, where no more other first-tier institutions can be encountered.

Our algorithm can be easily extended to make finer classifications. For example, in the second category, there are some VC institutions that have a larger number of investments that are identified as level 1 on this indicator and level 2 on the IPO&MA indicator (see those relatively scattered orange hexagons on the bottom-right in Figure 2c), such institutions might be less successful than those with the number of investments at level 2 but IPO&MA at level 1 (see those a few concentrated orange hexagons on the top-left in Figure 2c).

In the future, a closer investigation of the investment strategies of VC institutions in each category would be critical for developing a quantitative theory on the success of VC institutions. In this study, we focus our analysis on the Chinese market due to the limitation of data accessibility, which poses great challenges for making comparative analyses across countries.

**Author Contributions:** Conceptualization, R.L. and W.G.; Methodology, R.L. and W.G.; Software, L.L. and A.Y.; Validation, R.L., W.G., L.L. and A.Y.; Formal analysis, R.L., W.G., L.L. and A.Y.; Investigation, R.L., W.G., L.L. and A.Y.; Resources, R.L., W.G.; Data curation, R.L.; Writing—original draft preparation, R.L., W.G., L.L. and A.Y.; writing—review and editing, R.L., W.G., L.L. and A.Y.; Visualization, W.G., L.L. and A.Y.; Supervision, R.L. and W.G.; Project administration, R.L.; Funding acquisition, R.L. and W.G. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by National Natural Science Foundation of China grant number 61903020, Fundamental Research Funds for the Central Universities grant number buctrc201825, ZY2110, and the China Postdoctoral Science Foundation grant number 2021M690316.

**Data Availability Statement:** The data on venture capital investments is purchased from SiMuTong dataset of Zero2IPO Group ([www.pedata.cn](http://www.pedata.cn)) (accessed on 13 August 2018).

**Acknowledgments:** We acknowledge helpful discussions with Ran Liu from School of Mathematics and Statistics, Beijing Jiaotong University.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

|        |                                                                     |
|--------|---------------------------------------------------------------------|
| VC     | Venture Capital                                                     |
| IPO    | Initial Public Offering                                             |
| M&A    | Merge and Acquisition                                               |
| ICNEAC | Industrial Classification for National Economic Activities of China |
| HEI    | Hawk Eye Index                                                      |
| t-SNE  | t-distributed Stochastic Neighbor Embedding                         |

## Appendix A

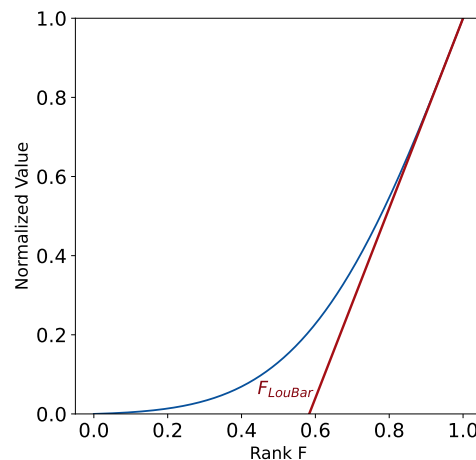
### Appendix A.1. Lorenz Curve

Lorenz curve was first proposed by Max O. Lorenz in 1905 to depict inequality in wealth distribution in economics. It was then developed by Corrado Gini to quantify income inequality or wealth inequality within a nation or a social group. The Lorenz curve is a cumulative distribution showing the proportion of overall income (or wealth) possessed by the bottom  $x\%$  of the people [35]. The percentage of households is plotted on the  $x$ -axis, and the percentage of income is plotted on the  $y$ -axis. For the Lorenz curve, the steeper the curvature, the stronger the inequality. Besides measuring inequality in wealth distribution, Lorenz curve is widely applied to quantify inequality among the size of individuals in ecology [49] and identifying the hotspots in urban studies [32].

### Appendix A.2. Loubar Algorithm

The  $x$ -axis of Lorenz curve is the normalized cumulative fraction of entities, and the  $y$ -axis is the normalized cumulative fraction of the sum value of all entities. The area between the diagonal and the curve corresponds to the standard Gini coefficient. The ordinary Loubar algorithm classifies objects into different classes based on the tangent line at  $(1, 1)$  of the Lorenz curve. The steeper the slope, the smaller the number of dominant individuals, which have a large value (e.g., a high income). In order to find the classification threshold  $\delta$ , we take the intersection point  $F^*$  (see Figure A1) between the tangent line of normalized  $L(F)$  at point  $(1, 1)$  and the horizontal  $x$ -axis  $L = 0$  (see Figure A1) [32]. In this analysis, we extend the method in an iterative way to obtain consecutive VC levels. Once

VC institutions at level  $l$  have been extracted, they are excluded from the distribution, and the threshold is recalculated to identify VC institutions at level  $l + 1$ .



**Figure A1.** Illustration of the criteria selection on the Lorenz curve of the Loubar method.

#### Appendix A.3. Iterative Loubar Classification Algorithm

As for the VC partition task, in order to objectively partition them based on their investment performance indicators, we devise an iterative nonparametric Loubar method [32–34,50] based on the derivative of Lorenz curves [35].

We first sort all VC firms ascendingly based on the number of invested start-ups, we then normalize the x-axis by the total number of VC institutions (see Figure 2a). In other words, the value  $n$  on the x-axis represents the first  $n$  percent of VC firms with the smallest investment numbers. The cumulative value of the investment numbers is the y-axis (see Figure 2a), which is also in the range of  $[0, 1]$  after normalization. We then draw the tangential line of the cumulative curve at point  $(1, 1)$  and divide all VC firms by the intercept point of the tangential line on the x-axis. VC firms with a larger investment number than the intercept point are assigned to the first level and then eliminated from the current set. Then we reapply the procedure on the set of remaining VC firms to obtain the second level, and so on (see illustration of the algorithm in Figure 2a). We end up with five levels in total.

Then, we apply the same procedures to the other indicator, the number of IPO and M&A, similarly, we also obtain four levels. Then VC firms that are in both first levels are classified as first-tier, which are VC firms with a larger number of invested start-ups and a larger number of IPO&MA; then, for remaining VC firms that are in the second level or above for both indicators are classified as second-tier. Second-tier VC firms comprise three sub-types: the ones in both the second level on the investment numbers and IPO&MA, ones in the second level on the investment numbers and in the first level on the IPO&MA, and ones in the first level on the investment number and second level on the IPO&MA. If necessary, second-tier VC firms can be further divided into three sub-categories. Similarly, we can obtain fourth-tier and fifth-tier VC institutions, where we can have more sub-categories.

Different from other studies that use empirical experience or arbitrary thresholds to make classification, the applied iterative Loubar method is more objective, with thresholds derived directly from the data per se.

#### Appendix A.4. Iterative Loubar Classification Performance for Different Types of Distributions

Lorenz curve was first proposed to measure inequality, and many real-world distributions (e.g., income) are of high heterogeneity, which is also related to other concepts including dispersion or concentration. Here, we also make some tests on distributions with different heterogeneity.

We find that the iterative Loubar algorithm generally works well on heterogeneous distribution (e.g., scale-free distribution with  $P(x) \propto x^{-\alpha}$ , see Figure A2a), for which the number of categories and their number of entities are reasonable (see Figure A2c). This is also the case in this work, as the distribution of the number of invested start-ups and IPO&MA are all close to scale-free distributions.

We then test the iterative Loubar algorithm on Gaussian distribution (see Figure A2d), and find that its first category already takes more than half of the population (see Figure A2f), which mixes individuals with largest values and near the average. And then the number of entities if the remaining categories decrease fast. Such classification may not be satisfactory and also indicate the limitation of the iterative Loubar method in dealing with normal distributions.

We further consider two extremes. One is that every entity is of the same value (see Figure A2g), then the distribution is highly concentrated (see Figure A2i). In this case, the Loubar algorithm can correctly divide all the data into one group.

The other one is that each entity has a different value, for example, taking values from 1 to  $N$ , respectively, where the distribution is uniform and quite dispersed (see Figure A2j). From the perspective of clustering, we may cluster each entity into a different category or classify the whole population as one group as they form a manifold [51]. However, the iterative Loubar method does not return any of such partitions, due to the fractal nature of such a distribution, the partition goes in a dichotomy way in that the first category contains half of the population, and the second category contains one-quarter, and so on (Figure A2l). This is also a situation in the Loubar method may not be appropriate.

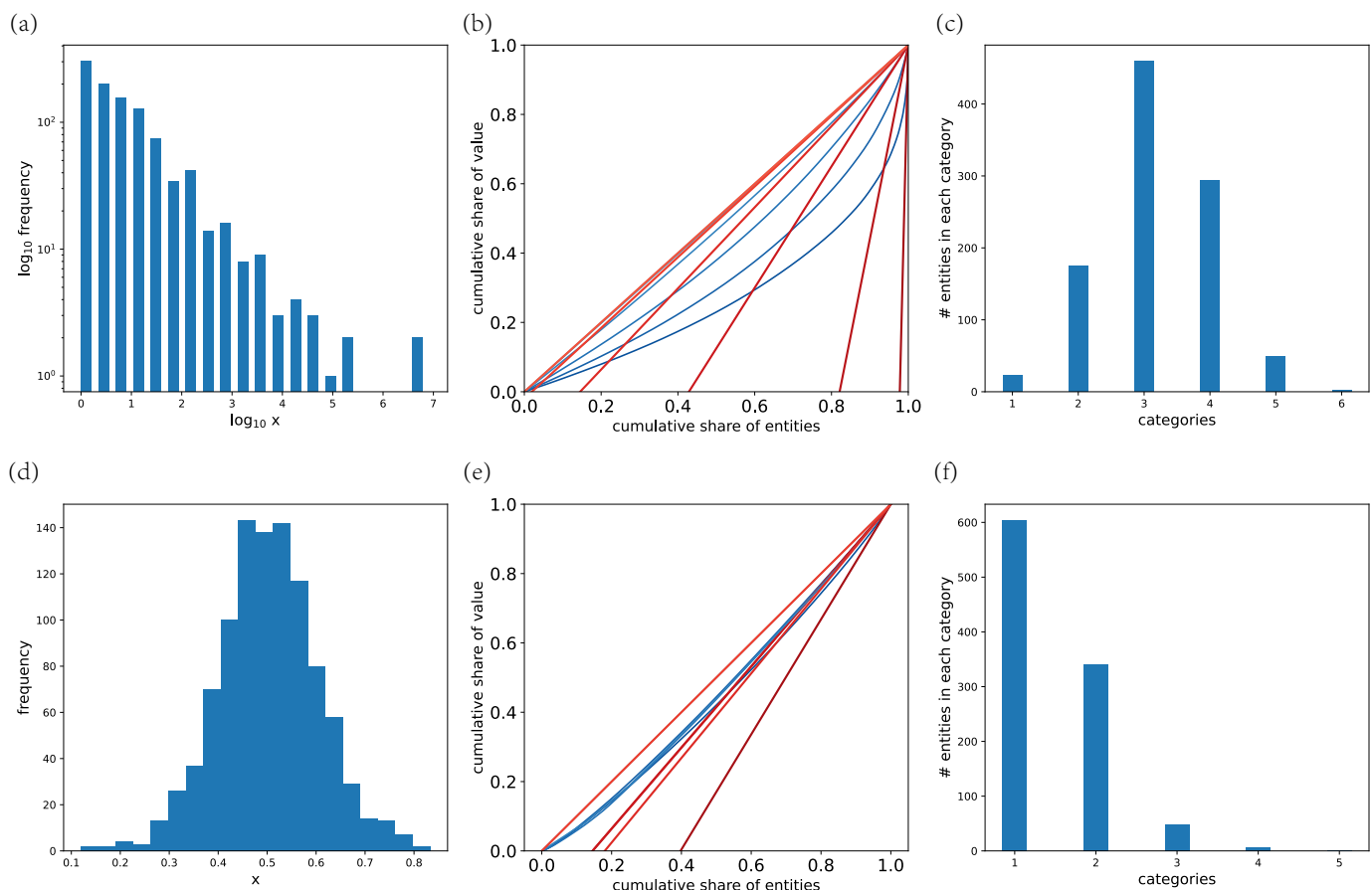
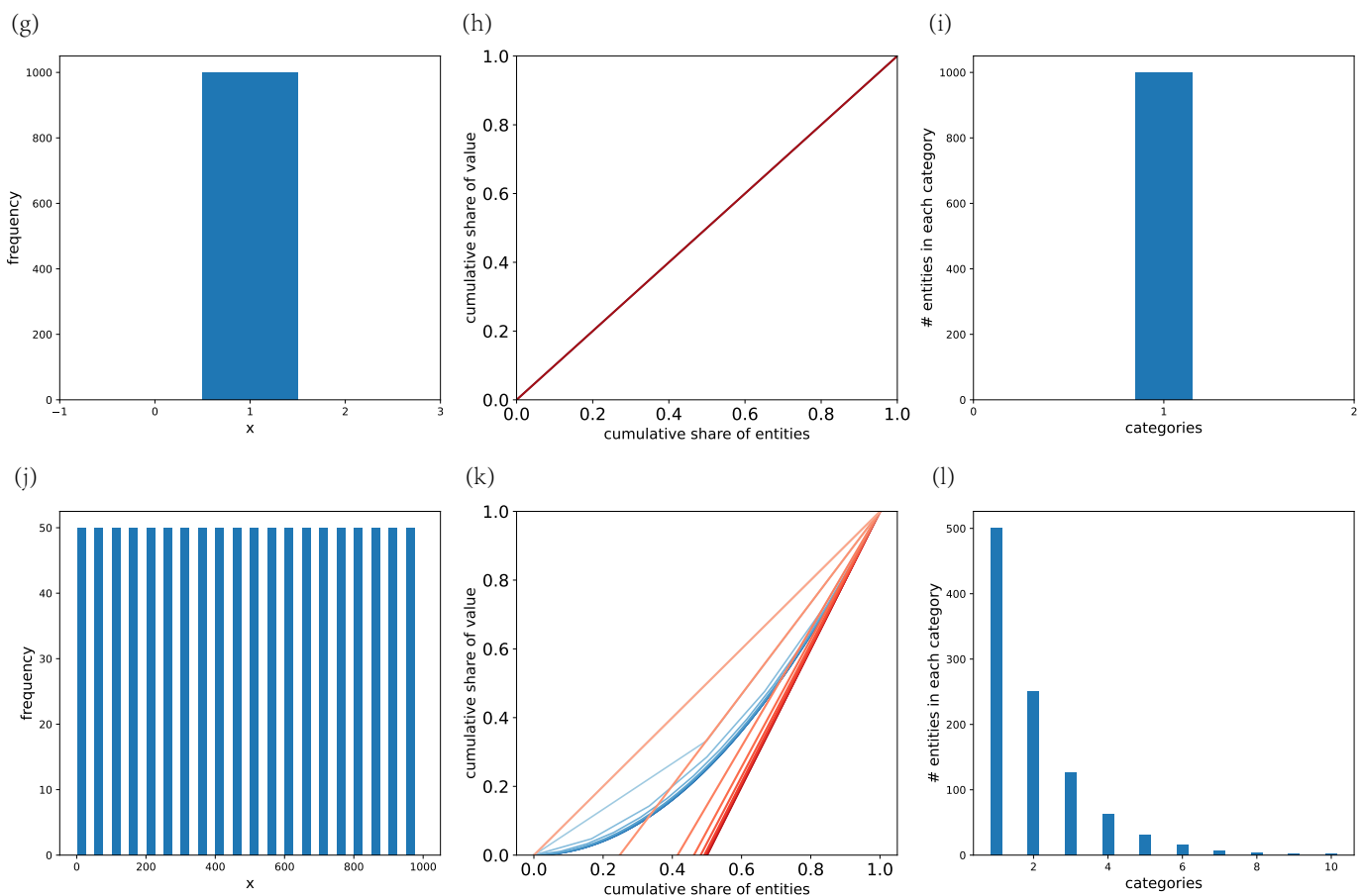


Figure A2. Cont.



**Figure A2.** Performing the iterative Loubar method on different distributions of data. (a) A power-law distribution with its exponent  $\alpha = 2.65$ , which is displayed in a log-log fashion. (b) The illustration of the iterative Loubar method on the power-law distribution shown in (a). (c) The distribution of the classification results, which shows the number of categories and the number of entities in each category, from the power-law distribution by the iterative Loubar method. (d–f) The same as (a–c) but on a normal distribution with mean equals 0.5 and standard deviation equals 0.1. (g–i) The same as (a–c) but on a delta distribution, where every entity is of the same value. (j–l) The same as (a–c) but on a uniform distribution, where all  $n$  entities have different values and equals 1 to  $n$ , respectively.

## References

- Gompers, P.A.; Lerner, J. *The Venture Capital Cycle*; MIT Press: Cambridge, MA, USA, 2004.
- Wang, C.; Wang, L.; Xue, Y.; Li, R. Revealing spatial spillover effect in high-tech industry agglomeration from a high-skilled labor flow network perspective. *J. Syst. Sci. Complex.* **2022**, *35*, 849–859. [CrossRef]
- Zinser, S. *The VC Landscape in China*; The Diplomat. 2021. Available online: <https://thediplomat.com/2021/01/the-global-reach-of-chinas-venture-capital/> (accessed on 10 October 2022).
- Li, R.; Lu, L.; Cui, T.; Gu, W.; Ma, S.; Xu, G.; Stanley, H.E. Assessing the Attraction of Cities on VentureCapital From a Scaling Law Perspective. *IEEE Access* **2021**, *9*, 48052–48063. [CrossRef]
- Li, R.; Dong, L.; Zhang, J.; Wang, X.; Wang, W.; Di, Z.; Stanley, H.E. Simple spatial scaling rules behind complex cities. *Nat. Commun.* **2017**, *8*, 1841. [CrossRef]
- Burt, R.S. *Structural Holes: The Social Structure of Competition*; Harvard University Press: Cambridge, MA, USA 2009.
- Katila, R.; Rosenberger, J.D.; Eisenhardt, K.M. Swimming with sharks: Technology ventures, defense mechanisms and corporate relationships. *Adm. Sci. Q.* **2008**, *53*, 295–332. [CrossRef]
- Uzzi, B. The sources and consequences of embeddedness for the economic performance of organizations: The network effect. *Am. Sociol. Rev.* **1996**, *61*, 674–698. [CrossRef]
- Podolny, J.M. Networks as the pipes and prisms of the market. *Am. J. Sociol.* **2001**, *107*, 33–60. [CrossRef]
- Manigart, S.; Wright, M.; Lockett, A.; Meuleman, M.; Landström, H.; Bruining, H.; Desbrières, P.; Hommel, U. Why do European venture capital companies syndicate? *Erim Rep. Ser. Ref.* **2002**, ERS-2002-98-ORG.
- Lerner, J. The syndication of venture capital investments. *Financ. Manag.* **1994**, *23*, 16–27. [CrossRef]

12. Hochberg, Y.V.; Ljungqvist, A.; Lu, Y. Whom you know matters: Venture capital networks and investment performance. *J. Financ.* **2007**, *62*, 251–301. [\[CrossRef\]](#)
13. Sorenson, O.; Stuart, T.E. Syndication networks and the spatial distribution of venture capital investments. *Am. J. Sociol.* **2001**, *106*, 1546–1588. [\[CrossRef\]](#)
14. Luo, J.D.; Li, R.; Fan, F.; Tang, J. Repeated Cooperation Matters: An Analysis of Syndication in the Chinese VC Industry by ERGM. In *Social Network Analysis*; CRC Press: Boca Raton, FL, USA, 2017; pp. 177–194.
15. Wu, J.; Luo, C.; Liu, L. A social network analysis on venture capital alliance's exit from an emerging market. *Complexity* **2020**, *2020*, 4650160. [\[CrossRef\]](#)
16. Alexy, O.T.; Block, J.H.; Sandner, P.; Ter Wal, A.L. Social capital of venture capitalists and start-up funding. *Small Bus. Econ.* **2012**, *39*, 835–851. [\[CrossRef\]](#)
17. Dimov, D.; De Clercq, D. Venture capital investment strategy and portfolio failure rate: A longitudinal study. *Entrep. Theory Pract.* **2006**, *30*, 207–223. [\[CrossRef\]](#)
18. Luo, J.D.; Rong, K.; Yang, K.; Guo, R.; Zou, Y. Syndication through social embeddedness: A comparison of foreign, private and state-owned venture capital (VC) firms. *Asia Pac. J. Manag.* **2019**, *36*, 499–527. [\[CrossRef\]](#)
19. Gu, W.; Jar-Der, L.; Liu, J. Exploring small-world network with an elite-clique: Bringing embeddedness theory into the dynamic evolution of a venture capital network. *Soc. Netw.* **2019**, *57*, 70–81. [\[CrossRef\]](#)
20. Tykvová, T. Who chooses whom? Syndication, skills and reputation. *Rev. Financ. Econ.* **2007**, *16*, 5–28. [\[CrossRef\]](#)
21. Liu, Z.; Chen, Z. Venture Capital Networks and Investment Performance in China. *Aust. Econ. Pap.* **2014**, *53*, 97–111. [\[CrossRef\]](#)
22. Batjargal, B. Comparative social capital: networks of entrepreneurs and venture capitalists in China and Russia. *Manag. Organ. Rev.* **2007**, *3*, 397–419. [\[CrossRef\]](#)
23. Huang, H.; Tang, J.; Liu, L.; Luo, J.D.; Fu, X. Analysis and prediction of triadic closure in online social networks. In *Social Network Analysis*; CRC Press: Boca Raton, FL, USA, 2017; pp. 105–136.
24. Yang, H.; Luo, J.D.; Fan, Y.; Zhu, L. Using weighted k-means to identify Chinese leading venture capital firms incorporating with centrality measures. *Inf. Process. Manag.* **2020**, *57*, 102083. [\[CrossRef\]](#)
25. Aral, S.; Walker, D. Identifying influential and susceptible members of social networks. *Science* **2012**, *337*, 337–341. [\[CrossRef\]](#)
26. Li, Q.; Zhou, T.; Lü, L.; Chen, D. Identifying influential spreaders by weighted LeaderRank. *Phys. A Stat. Mech. Its Appl.* **2014**, *404*, 47–55. [\[CrossRef\]](#)
27. Li, R.; Wang, W.; Di, Z. Effects of human dynamics on epidemic spreading in Côte d'Ivoire. *Phys. A Stat. Mech. Its Appl.* **2017**, *467*, 30–40. [\[CrossRef\]](#)
28. Lü, L.; Chen, D.; Ren, X.L.; Zhang, Q.M.; Zhang, Y.C.; Zhou, T. Vital nodes identification in complex networks. *Phys. Rep.* **2016**, *650*, 1–63. [\[CrossRef\]](#)
29. Li, R.; Wang, W.; Shu, P.; Yang, H.; Pan, L.; Cui, A.; Tang, M. Review of Threshold Theoretical Analysis about Epidemic Spreading Dynamics on Complex Networks. *Complex Syst. Complex. Sci.* **2016**, *13*, 1–39.
30. Gu, W.; Gao, F.; Lou, X.; Zhang, J. Discovering latent node Information by graph attention network. *Sci. Rep.* **2021**, *11*, 6967. [\[CrossRef\]](#) [\[PubMed\]](#)
31. McGrath, R.G. Falling forward: Real options reasoning and entrepreneurial failure. *Acad. Manag. Rev.* **1999**, *24*, 13–30. [\[CrossRef\]](#)
32. Louail, T.; Lenormand, M.; Ros, O.G.C.; Picornell, M.; Herranz, R.; Frias-Martinez, E.; Ramasco, J.J.; Barthelemy, M. From mobile phone data to the spatial structure of cities. *Sci. Rep.* **2014**, *4*, 5276. [\[CrossRef\]](#)
33. Bassolas, A.; Barbosa-Filho, H.; Dickinson, B.; Dotiwalla, X.; Eastham, P.; Gallotti, R.; Ghoshal, G.; Gipson, B.; Hazarie, S.A.; Kautz, H.; et al. Hierarchical organization of urban mobility and its connection with city livability. *Nat. Commun.* **2019**, *10*, 4817. [\[CrossRef\]](#)
34. Gao, T.; Xu, Z.; Liu, C.; Yang, Y.; Shang, F.; Li, R. Quantifying relation between mobility patterns and socioeconomic status of dockless sharing-bike users. *arXiv* **2022**, arXiv:2204.03894.
35. Beach, C.M.; Kaliski, S.F. Lorenz curve inference with sample weights: an application to the distribution of unemployment experience. *J. R. Stat. Soc. Ser. C (Appl. Stat.)* **1986**, *35*, 38–45. [\[CrossRef\]](#)
36. Grover, A.; Leskovec, J. node2vec: Scalable feature learning for networks. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 855–864.
37. These are the Top Venture Capital Companies in China (2021). Available online: <https://welpmagazine.com/these-are-the-top-venture-capital-companies-in-china-2021> (accessed on 23 August 2022).
38. Jain, A. These are the Top Venture Capital Firms of 2020. Available online: <https://www.entrepreneur.com/article/361514> (accessed on 23 August 2022).
39. Lange, J.E.; Bygrave, W.; Nishimoto, S.; Roedel, J.; Stock, W. Smart money? The impact of having top venture capital investors and underwriters backing a venture. *Ventur. Cap. Int. J. Entrep. Financ.* **2001**, *3*, 309–326. [\[CrossRef\]](#)
40. Govindarajan, V.; Srivastava, A. The scary truth about corporate survival. *Harv. Bus. Rev.* **2016**, *12*, 24–25.
41. Bonaventura, M.; Ciotti, V.; Panzarasa, P.; Liverani, S.; Lacasa, L.; Latora, V. Predicting success in the worldwide start-up network. *Sci. Rep.* **2020**, *10*, 345. [\[CrossRef\]](#) [\[PubMed\]](#)
42. Kitsak, M.; Gallos, L.K.; Havlin, S.; Liljeros, F.; Muchnik, L.; Stanley, H.E.; Makse, H.A. Identification of influential spreaders in complex networks. *Nat. Phys.* **2010**, *6*, 888–893. [\[CrossRef\]](#)
43. Evans, T.S.; Chen, B. Linking the network centrality measures closeness and degree. *Commun. Phys.* **2022**, *5*, 1–11. [\[CrossRef\]](#)

- 
44. Lü, L.; Zhou, T.; Zhang, Q.M.; Stanley, H.E. The H-index of a network node and its relation to degree and coreness. *Nat. Commun.* **2016**, *7*, 10168. [[CrossRef](#)]
  45. Perozzi, B.; Al-Rfou, R.; Skiena, S. Deepwalk: Online learning of social representations. In Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 24–27 August 2014; pp. 701–710.
  46. Mikolov, T.; Chen, K.; Corrado, G.; Dean, J. Efficient estimation of word representations in vector space. In Proceedings of the Workshop at ICLR, Scottsdale, AZ, USA, 2–4 May 2013.
  47. Gu, W.; Tandon, A.; Ahn, Y.Y.; Radicchi, F. Principled approach to the selection of the embedding dimension of networks. *Nat. Commun.* **2021**, *12*, 3772. [[CrossRef](#)]
  48. Van der Maaten, L.; Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 11.
  49. Damgaard, C.; Weiner, J. Describing inequality in plant size or fecundity. *Ecology* **2000**, *81*, 1139–1142. [[CrossRef](#)]
  50. Louf, R. Wandering in cities: a statistical physics approach to urban theory. *arXiv* **2015**, arXiv:1511.08236.
  51. Shang, F.; Chen, B.; Expert, P.; Lü, L.; Yang, A.; Stanley, H.E.; Lambiotte, R.; Evans, T.; Li, R. Local dominance unveils clusters in networks. *arXiv* **2022**, arXiv:2209.15497.