Network-based Landscape of Research Strengths of Universities in Mainland China

Zihua Liu¹, Qin Xiao², Qian Zhan³, Cahngui Gu¹, and Huijie Yang^{1*}

¹ *Business School, University of Shanghai for Science and Technology, Shanghai 200093, China*

² *College of Sciences, Shanghai Institute of Technology, Shanghai 201418, China*

³ *Ordinance Engineering College, Shijiazhuang, Hebei Province 050003, China*

A landscape of a complex system presents a quantitative measure of its global state. The profile of research strength in mainland China is investigated in detail, by which we illustrate a complex network based framework to extract a landscape from detailed records. First, a measure analogous to the Jaccard similarity is proposed to calculate from the presided funds similarities between the top-ranked universities. The neighbor threshold method is employed to reconstruct the similarity network of the universities. Second, the network is divided into communities. In each community the node with the largest degree and the smallest average shortest path length is taken as the representative of the community, called central node. The node bridging each pair of communities is defined to be a boundary. The central nodes and boundaries cooperatively give us a picture of the research strength landscape. Third, the evolutionary behavior is monitored by the fission and fusion probability matrices, elements of which are the percentage of a community at present time that joins into every community at the next time, and the percentage of a community at next time that comes from every present community, respectively. The landscapes in three successive 4-year durations are identified. It was found that some types of universities, such as the medicine&pharmacy and the finance&economy, conserve in single communities in the more than ten years, respectively. The agriculture&forest universities tend to cluster into one community. Meanwhile the engineering type distributes in different communities and tends to mix with the comprehension type. This framework can be used straightforwardly to analyze temporal networks. It provides also a new network-based method for multivariate time series analysis.

keywords: Complex network; Landscape of complex system; Research strength in Mainland China

PACS numbers: 89.75.Hc; 89.75.Fb; 87.23.Kg

I. INTRODUCTION

A landscape of research strength is beneficial for us in multi-ways [1, 2]. It can provide an evaluation of the academic level for the government to monitor the scientific development, a guidance for fund providers and high-educational administrators to adjust supporting policies, and a direction for researchers to find collaborators.

In Mainland China, most of the researchers working in basic disciplinary fields distribute in the top universities. Research activities of the universities can be regarded as representative of the national academic level. In the present paper by using the database of National Science Foundation of China (NSFC) [3] we reconstruct a similarity network of top universities [4]. The structural behaviors of the network and their changes in successive durations are used to illustrate the landscape of the research strength and its evolution.

Technically, we select the top 116 universities to represent the institutions working on basic researches. All the NSFC-supported projects presided by these universities in a duration of twelve years are collected, which are catalogued according to the applying system into 101 sub-disciplinary fields [3]. The overlapping of distribu-

tions of projects is used to measure similarities between the universities. We construct subsequently the similarity network by linking every university with the two universities most similar with it.

It was found that the universities cluster into several communities. In each community the university with the largest number of neighbors and the smallest average shortest path length is taken as the representative of the community, called central node. The universities connecting different communities are defined to be boundaries. The central nodes and boundaries are identified, which cooperatively give us a profile of the research strength. The profiles in successive durations show us a picture of global evolution. Some types of universities, such as the medicine&pharmacy and the finance&economy, stay in single communities in the more than ten years, respectively. The agriculture&forest universities tend to cluster into one community. The engineering type distributes in different communities and tends to mix with the comprehension type. Though great progress has been achieved, the main infrastructure of the educational system remains unchanged. The proposed framework can be used to display a picture of a temporal network. It provides also a new method to analyze multivariate time series from the perspective of complex network theory.

*Corresponding author. E-mail: hjyang@ustc.edu.cn

Institution	Sub-catalogue identification number	Project title (Project No.)	Amount (k RMB)	Duration	Researcher
Tsinghua University	G01	China's financial derivatives pricing and foam under Short-sale constraints(71101080)	200	2012/01/01-2014/12/31	Y. Wang
Tsinghua University	F03	Traffic sign detection, recognition and understanding in complex conditions(91120301)	2800	2012/01/01-2015/12/31	C. Zhang
Tsinghua University	E08	Algorithms for the calculation of carbon emissions in city residential areas(51178238)	600	2012/01/01-2015/12/31	N. Liu

TABLE I: Several examples of project records from NSFC.

II. MATERIALS AND METHODS

Data

In 1986, China's State Council approved the establishment of the National Science Foundation of China (NSFC) [3]. The NSFC Committee is in charge of its running. The aim is to support in the nationwide free explorations of basic scientific problems, and play consequently a guiding role in the scientific development. Researchers in universities and research institutions can freely apply the support by providing their proposals. The support will be given to the applications with competitive strengths and innovative ideas. Researchers working in basic scientific fields in Mainland China receive fund supports mainly from NSFC, though it is not the sole way.

A typical record of a NSFC project contains the name of institution, sub-catalogue identification number, title, amount, duration, and researcher's name, as shown in Table I several examples. Herein, we are interested in the sub-catalogue identification number. All the research fields are divided into 101 sub-catalogues (see Table A1 in Appendix).

Each year several organizations will independently issue ranks for all the universities in Mainland China. Though the details of the rank lists are different, the listed top universities are almost the same. In the present work we select the top 124 universities issued by Prof. Shulian Wu's group [4]. Because the records of eight universities are bad in quality, these universities are deleted in our consideration, which results into a set of 116 universities (see Table A2 in Appendix). Here the ranks are used as ID numbers of the universities. In the duration from 2004 to 2015, the selected universities applied successfully a total of 134,064 projects (51.07 billion RMB) from the NSFC, while the total number of projects supported by the NSFC is 252,168 (108.58 billion RMB). The 116 universities have covered a ratio of $\frac{134064}{252168} \sim 53.16\%$ of the supported researches. Hence, the selected universities can be used as representative of the communities working in basic disciplinary fields. In calculations, some records are cleared out due to low qual-

ties and the final number of projects is 131298, covering $\frac{131298}{134064} \sim 97.94\%$ of the total records.

About sixty years ago, the present government conducted a plan of reuniting universities, namely, the existing universities were broken into parts according to specialties and the parts belonging to identical or similar specialties were subsequently re-integrated into new universities [5]. From then on the universities are catalogued into different types [5], including comprehension (COM), science&technology (ENG), agriculture&forest (AGR), medicine&pharmacy (MED), normal (NOR), and finance&economy (FIN), etc.. About thirty years ago, the opening Mainland China realized the limitations of the catalogues and all the universities began to develop independently. But due to the unbalance of social developments in different regions and the persistence of the government's policies, not every university can develop as its expect. The government is still issuing for each university a catalogue type by using the traditional evaluation system, as shown in Table A2 in Appendix.

Similarity networks

Let us consider a specific duration of time T . From the records one can obtain a matrix describing the distribution of supports from the NSFC, $\{S(m, n), m = 1, 2, \dots, 116; n = 1, 2, \dots, 101\}$, where $S(m, n)$ is the number of projects in the n 'th sub-catalogue that researchers in the m 'th university are supported.

A measure analogous to the Jaccard similarity [6] is used,

$$J(m_1, m_2) = \frac{\sum_{n=1}^{101} \frac{\min[S(m_1, n), S(m_2, n)]}{\max[S(m_1, n), S(m_2, n)]}}{\sum_{n=1}^{101} H(\max[S(m_1, n), S(m_2, n)])}, \quad (1)$$

where $H(x)$ is the Heaviside function, namely, $H(x > 0) = 1, H(x \leq 0) = 0$, and if $S(m_1, n) = S(m_2, n) = 0$, we define $\frac{\min[S(m_1, n), S(m_2, n)]}{\max[S(m_1, n), S(m_2, n)]} \equiv 0$. That is, we do not consider the sub-catalogues being empty both for the two universities. One can find that only when the project numbers for every sub-catalogue, $S(m_1, n)$ and $S(m_2, n)$,

are very close the similarity will tends to 1. A high similarity between a pair of universities has bi-fold implication, namely, the total number of projects they obtained are comparable, and the distributions of the projects are also similar. Hence, the definition can be used to describe research strength similarity.

To extract useful information from the similarity matrix, let us construct from the similarities a network [7–15]. A simple fact is that each university is interested in only several universities being most similar with it. Accordingly, a strategy called neighbor threshold method is used in this work, namely, we scan all the universities and link every university with a specified number of universities most similar with it [9, 12]. If a pair of universities are linked twice by this procedure, we just record the link state and ignore the times. The resulting network is called a similarity network.

Our purpose is to extract a profile of the network, which is displayed by the central nodes and boundaries as mentioned below. Accordingly, an indirection network is constructed. In the above neighbor threshold method, if every university is linked to its most similar neighbors we will have a directional network. A link from A to B implies that A is interested in B . The directions affect significantly microscopic structures of and physical processes on the network, which are not the focus of the present work.

Validity of the neighbor threshold method

To make the estimations of similarity informative, the university-catalogue matrix S should be sparse enough. Herein we use the inverse participation ratio and the structural entropy [16–21] to display the distribution behaviors of the funds.

From the distribution matrix S , one can calculate the ratio of the funds covered by every specific element, $p(m, n) \equiv \frac{S(m, n)}{\sum_{i=1}^M \sum_{j=1}^N S(i, j)}$, where M and N denote the total numbers of universities and sub-catalogues respectively. Regarding this ratio as a probability, the inverse participation ratio reads [16, 17],

$$R_{\text{part}} = \frac{1}{M \cdot N \cdot \sum_{m=1}^M \sum_{n=1}^N p^2(m, n)}. \quad (2)$$

If the funds occur homogenously only in a total of K elements of S , i.e., the distribution obeys a step-like function, we have $R_p = \frac{K}{M \cdot N}$. Hence, it gives us an equivalent ratio of the number of elements occupied by the funds over the total number of elements. A small value of R_{part} implies that the funds tend to aggregate on a small number of elements, i.e, a high degree of distribution sparsity.

Generally, the distribution has detailed patterns rather than the simple step-like shape, which can be described

by the structural entropy, which reads [18],

$$E_{\text{stru}} = \sum_{m=1}^M \sum_{n=1}^N -p(m, n) \ln p(m, n) - \ln(M \cdot N \cdot R_{\text{part}}). \quad (3)$$

The first term is the Shannon entropy of the distribution. The second term is the Shannon entropy of the equivalent step-like distribution. If the distribution has a step-like shape, the structural entropy will vanish. A large value of E_{stru} implies that there exist significant and rich patterns in the distribution, i.e, a high degree of heterogeneity. The structural entropy displays the deviation of the distribution from a step-like shape.

One can introduce analogous measures of distribution sparsity and heterogeneity for every university,

$$\begin{aligned} R_{\text{part}}(m) &= \frac{1}{N \cdot \sum_{n=1}^N \left[\frac{S(m, n)}{\sum_{j=1}^N S(m, j)} \right]^2}, \\ E_{\text{stru}}(m) &= - \sum_{n=1}^N \frac{S(m, n)}{\sum_{j=1}^N S(m, j)} \ln \left[\frac{S(m, n)}{\sum_{j=1}^N S(m, j)} \right] \\ &\quad - \ln [N \cdot R_{\text{part}}(m)], \\ m &= 1, 2, \dots, M, \end{aligned} \quad (4)$$

and for every sub-catalogue,

$$\begin{aligned} R_{\text{part}}(n) &= \frac{1}{M \cdot \sum_{m=1}^M \left[\frac{S(m, n)}{\sum_{j=1}^M S(j, n)} \right]^2}, \\ E_{\text{stru}}(n) &= - \sum_{m=1}^M \frac{S(m, n)}{\sum_{j=1}^M S(j, n)} \ln \left[\frac{S(m, n)}{\sum_{j=1}^M S(j, n)} \right] \\ &\quad - \ln [M \cdot R_{\text{part}}(n)], \\ n &= 1, 2, \dots, N, \end{aligned} \quad (5)$$

called inverse participation ratio and structural entropy for the m 'th university and the n 'th sub-catalogue, respectively.

Besides that the similarity should be informative, a mainly centralized distribution of similarity may leads to the invalidity of the neighbor threshold method. Selection of the number of the nearest neighbors may affect heavily the results. In the section of results we will show the distribution behaviors of the similarities.

Properties of the similarity network

To display the landscape of research strengths of the universities and its evolution, our attentions will focus on the structural patterns at different scales, including communities, central node within and boundaries of a community, and fission and fusion probability matrices.

Community [22]. In a complex network the nodes cluster generally into some densely connected groups, called communities. The communities are connected loosely with a few edges. The Newman-Girvan method [22] is used to find the communities, in which a modularity measure Q is defined to evaluate the performance of the separation. A larger value of Q means that the network has a much distinguishable modularity-structure.

Representative of a community. In the constructed university network, there are generally different paths to get from one node to another node, among which the path with the smallest number of hops is called the shortest path between the two nodes. Let us specify a node and find the shortest paths from it to all the other nodes. Average length of all the shortest paths is defined to be the average shortest path length of the specified node [23], a small (large) value of which implies that the node is very close to (far away from) the center of the network.

Degree of a node refers to the number of nodes directly linked with it [23]. The nodes whose degrees are significantly large compared with that of the others are called hubs. If a university has a significant large number of neighbors, it is used as the representative of those universities.

Herein, the average shortest path length and the degree of a university are jointly used to find the representatives of the communities. We isolate each community from the whole network and find within this community the university that has the largest degree and the smallest average shortest path length. If there are several universities turn out to be candidates, the university with the smallest ID number (the strongest one among them) is selected. The university singled out by this procedure is called central node, which is taken as the representative of the community.

Boundary of a community. We are also interested in the universities through which different communities are bridged together. These universities should have characteristics shared by the connected communities. Some universities are trying to enhance their levels, i.e., from the present community to a new one with higher strength of research. The boundary universities can act as an intermediate goal and provide a good reference and an optimal route. If there are several universities turn out to be candidates, the university with the smallest ID number (the strongest one among them) is selected.

Evolution of a network. Let us denote the communities at present duration with $C1_i, i = 1, 2, \dots, P$, and that at the successive next duration with $C2_j, j = 1, 2, \dots, F$. The total number of communities at the two durations are denoted with P and F , respectively. In the evolutionary process, every community at present time will split into pieces, which subsequently take part in the formations of the communities in the next time [24–31]. This evolution can be described quantitatively by the fission and fusion matrices, p_{fis} and p_{fus} [30, 31]. The element $p_{\text{fis}}(i, j)$ is the percentage of the nodes in $C1_i$ that joins into $C2_j$, the element $p_{\text{fus}}(i, j)$ the percentage of nodes in $C2_j$ that comes from $C1_i$. If the values of $p_{\text{fis}}(i, j)$ and $p_{\text{fus}}(i, j)$ are both large, most part of $C1_i$ forms the most part of $C2_j$, i.e., the community is conserved. If $p_{\text{fis}}(i, j) \gg p_{\text{fus}}(i, j)$, most part of $C1_i$ takes part in the formation of $C2_j$ as only a small part of the members. As for the case of $p_{\text{fis}}(i, j) \ll p_{\text{fus}}(i, j)$, a small part of $C1_i$ forms the main part of $C2_j$.

	C1_1	C1_2	C1_3	C1_4	C1_5	C1_6	C1_7	C1_8	C1_9
C1_1	10	41	—	—	—	—	—	—	—
C1_2	22	71	99	—	—	28	—	—	—
C1_3	—	55	45	11	30	—	85	108	—
C1_4	—	—	32	32	31	—	—	—	—
C1_5	—	—	90	75	37	—	101	39	90
C1_6	—	50	—	—	—	76	—	49	—
C1_7	—	—	96	—	106	—	38	106	—
C1_8	—	—	105	—	94	93	94	105	—
C1_9	—	—	—	—	79	—	—	—	74

TABLE II: Representatives and boundaries of the communities in the duration of 2004 – 2007. The element corresponding to the row $C1_x$ and the column $C1_y$ is the university that belongs to $C1_x$ and acts as the boundary between the communities $C1_x$ and $C1_y$.

The diagonal element at the row $C1_x$ is the representative of the community $C1_x$.

	C2_1	C2_2	C2_3	C2_4	C2_5	C2_6	C2_7	C2_8	C2_9	C2_10
C2_1	20	25	69	—	—	—	—	—	—	—
C2_2	41	8	3	—	—	—	—	—	27	—
C2_3	73	28	49	—	—	—	16	—	88	94
C2_4	—	—	—	9	—	—	—	—	—	—
C2_5	—	—	—	—	107	11	11	—	—	—
C2_6	—	—	—	—	85	82	—	—	114	110
C2_7	—	—	37	—	37	—	32	—	56	32
C2_8	—	—	—	—	—	—	—	72	—	—
C2_9	—	45	43	—	—	84	97	—	43	113
C2_10	—	—	80	—	—	38	65	—	58	65

TABLE III: Representatives and boundaries of the communities in the duration of 2008 – 2011. The element corresponding to the row $C2_x$ and the column $C2_y$ is the university that belongs to $C2_x$ and acts as the boundary between the communities $C2_x$ and $C2_y$.

The diagonal element at the row $C2_x$ is the representative of the community $C2_x$.

	C3_1	C3_2	C3_3	C3_4	C3_5	C3_6	C3_7	C3_8	C3_9
C3_1	1	4	—	—	—	—	—	—	—
C3_2	3	20	—	—	28	29	—	—	—
C3_3	—	—	57	114	—	—	—	—	—
C3_4	—	—	82	97	55	—	—	—	—
C3_5	—	50	—	70	71	—	30	—	—
C3_6	—	25	—	—	—	39	65	—	—
C3_7	—	—	—	—	37	116	48	40	—
C3_8	—	—	—	—	—	—	103	76	—
C3_9	—	—	—	—	—	—	—	—	62

TABLE IV: Representatives and boundaries of the communities in the duration of 2012 – 2015. The element corresponding to the row $C3_x$ and the column $C3_y$ is the university that belongs to $C3_x$ and acts as the boundary between the communities $C3_x$ and $C3_y$.

The diagonal element at the row $C3_x$ is the representative of the community $C3_x$.

FIG. 1: (Color online) Distribution behaviors of the funds. The number of funds decays exponentially, as shown in (a) The number of funds presided by every university. The universities are ranked according to a descending order of the number of funds; and in (b) The number of funds in every sub-catalogue. The sub-catalogues are ranked according to a descending order of the number of funds. The distributions are sparse enough and have rich patterns rather than a step-like shape, as shown in (c) Scatter plot of structural entropy versus inverse participation ratio for the universities; and in (d) Scatter plot of structural entropy versus inverse participation ratio for the sub-catalogues. The open stars refer to the results for the distribution matrix S in Eq.(1).

FIG. 2: (Color online) Similarity distribution. (a) In the duration of 2004 – 2007. (b) In the duration of 2008 – 2011. (c) In the duration of 2012 – 2015. The similarities between all the universities distribute normally (see the black solid circles and the black solid line), respectively. The similarities for two nearest neighbors and three nearest neighbors distribute mainly in the region of 5% largest similarities (indicated by the dotted black lines), and between the distributions of which there exists only a slight difference.

III. RESULTS

From the distribution matrix S we calculate the number of funds presided by every university and the number

of funds in every sub-catalogue, as shown in Fig.1(a) and (b), where the universities and the sub-catalogues are ranked with a descending order of the numbers of funds, respectively. The number of funds decays exponentially.

FIG. 3: (Color online) Similarity networks constructed with records in the durations of 2004 – 2007, 2008 – 2011, and 2012 – 2015, as shown in (a),(b), and (c), respectively. The universities clustered into several communities, marked with different colors. The representatives of the communities are denoted with squares.

Fig.1(c) and (d) present the inverse participation ratios and structural entropies for the universities and the sub-catalogues. One can find that for the most universities or sub-catalogues the funds cover an equivalent ratio of 5% to 25% of the total elements. And the structural

entropies are significantly larger than zero, i.e, the distributions have rich and significant patterns rather than a simple step-like shape. The distribution S is sparse enough, which makes the similarities informative.

We calculate the similarities between all the universi-

FIG. 4: (Color online) Distributions of universities in the communities according to their types. The distributions in the durations 2004 – 2007, 2008 – 2011, and 2012 – 2015 are shown in (a), (b), and (c), respectively. In each duration most of the medicine&pharmacy and finance&economy type universities cluster in single communities, respectively. The agriculture&forest type evolves from a distributed state to a sole community. The comprehension type distributes in almost all the communities.

ties, which distribute normally, as shown in Fig.(2) with the black solid circles. The dotted lines indicate the specific values of similarity, the fractions of the similarities larger than which are 5% of the total number of similarities. The largest 5% similarities scatter in a large range of ~ 0.3 . We collect also the largest two and the largest three similarities for every university, whose distributions are depicted with red solid circles and blue solid circles, respectively. Most of them occur in the 5% largest similarities. And there is only a slight difference between them.

It should be noted that in the neighbor threshold method, from Fig.2 one can find that we will construct some links between some pairs of universities with small similarities. If the similarity corresponding to a link from a university is small, the university has small similarities with all the other universities. In the constructed network, the university will not be a hub. In the present paper we are interested in the communities and the universities affecting macroscopic structures of the communities (central nodes). Hence, the results are not sensitive to the specific number of neighbors in the neighbor threshold method.

In the present paper the two-neighbor threshold is used, namely, we link every university with the two universities having largest similarity values with it. The resulting networks corresponding to the durations of 2004 – 2007, 2008 – 2011, and 2012 – 2015 are shown in Fig.3. The universities cluster into 9, 10 and 9 commu-

nities, numbered with $C1_x$ ($x = 1, 2, \dots, 9$), $C2_y$ ($y = 1, 2, \dots, 10$) and $C3_z$ ($z = 1, 2, \dots, 9$), respectively. The modularity measures for the three networks are $Q = 0.719, 0.722$ and 0.749 , respectively. Few universities (e.g., the universities No. 81, 87, and 95 in the period 2004 – 2007) are isolated nodes due to the non-validation of records, namely, the universities have no funds from the NSFC.

Development of a university is constrained by several essential factors, such as the positive-feedback mechanism and the competition between universities for limited resources. The cooperative effect of the factors is enforced further by the strategy so-called "adding flowers to the brocade", employed usually by the government to promote the high-educational development [32]. The universities are ranked into several classifications. The government will accordingly provide favorable policies and more financial supports to the top universities [33]. Hence, the universities will cluster into communities.

Figure 4 shows the distributions of universities in the communities according to their types. One can find that in every duration most of the medicine&pharmacy and finance&economy type universities cluster in single communities, respectively. These two types are conserved in the more than ten years' period. The agriculture&forest type, which scattered initially in different communities, tends to cluster into a sole community. Meanwhile, the comprehension type universities distribute in almost all the communities. The agricul-

FIG. 5: (Color online) Evolutionary behavior of the system. (a) Fission matrix. (b) Fusion matrix. The widths of the arrows are proportional to the value of the corresponding elements.

ture and medicine&pharmacy types cluster into single communities ($C3_3$ and $C3_8$), respectively. The engineering type universities tend to mix with the comprehension type universities. A detailed type-based distribution of universities in the communities is provided in Table A3 in appendix.

The findings can be explained by using simple facts occurring in Mainland China. For instances, the medicine&pharmacy and finance&economy types are closely related with comparatively specific and conserved techniques and servant fields, respectively. With the changes of economical and social structures the role of agriculture & forest becomes less important and less competitive. The strategy of this type universities is to develop uniquely in the agriculture and/or forest related fields. Meanwhile the rapid development of economy in Mainland China provides unusual opportunities to the engineering type universities. And the academic evaluation system encourages them to develop toward the comprehension type.

Figure 5(a) and (b) depicts the fission and fusion matrices, respectively. The widths of the arrows are proportional to the values of the corresponding elements.

From Fig.5(a) one can find that, for instance, 63% of $C1_9$ participates in the formation of $C2_5$, whose 75% will take part in further the formation of $C3_9$. Meanwhile, Fig.5(b) shows that 67% of $C3_9$ comes from $C2_5$, whose 63% comes from $C1_9$. Hence, the evolution trajectory $C1_9 \rightarrow C2_5 \rightarrow C3_9$ is a conserved path. As shown in Fig.4, this conserved evolution corresponds to the conservation of the finance&economy type universities in the community $C1_9$. At the same time, there are also some communities that break into small parts and participate subsequently in different communities. For an example, the community $C1_3$ is broken into seven parts that join into all the communities in the duration-2 except the three communities $C2_1$, $C2_4$ and $C2_8$; the community $C3_7$ comes from many small parts of different communities in the duration-2, and each part cover only a small percentage of the total universities, which correspond to the active states of the engineering and comprehension type universities in community $C3_7$.

From Fig.4 one can find that the universities belong to the same type tend to cluster together, which can be regarded as traces of main infrastructure of the educational system with a life of more than sixty years. Meanwhile,

Fig.5 shows us great changes in the educational system in the 12-year duration. It is a cooperative result of many factors at different levels such as the change of administration system of the government, the change of supporting policies of the research funders, and the merge of universities. To determine contributions of the factors requires detailed and professional works.

In Table I, II, and III we summarize the representative for every community, namely, the university having the largest degree and the smallest average shortest path length. The boundaries of all the communities are also presented. If several universities are candidates of representative, the representative is chosen to be the one with the smallest ID number (the strongest one among them).

It should be noted that the representative and boundaries in each community are weakly relative to the large scale (strongest) universities in Mainland China. It can be understood from the definition of similarity in this paper. A larger value of similarity implies two facts, namely, the fund distributions are similar and the total numbers of funds are comparable. A large scale university can obtain a significant large number of funds. The funds distribute widely in different sub-catalogues. However, as shown in Fig.1(a) the number of funds decreases exponentially, namely, in Mainland China there are only a few top universities with significantly strong research strengths and the differences between them are also significant. Hence, the similarities of a large scale university with other universities should be small. The few strongest universities will not be the central nodes in the constructed network, and cluster mainly in one community because the similarities between them are comparatively large with high possibilities. From Fig.2 one can find that if we use alternatively a specific similarity as a threshold to filter out the low similarities, the most of the resulting links have been covered in the similarity networks shown in Fig.3. The few strongest universities will not be the central nodes also. Hence, it's the specific definition of similarity that leads to the result.

Just as mentioned above, the strategy of "adding flowers to the brocade" in the development has enlarged the differences between universities, even between the few strongest universities (see the exponential decay of fund number in Fig.1(a)). The few strongest universities are high-peaks of research strength, however they can not be taken as the representatives of the global level of the research strength over the whole Mainland China. Hence, the result is reasonable and has potential use in evaluation of the policies.

IV. CONCLUSION AND FUTURE WORK

A complex system can be described by a network, in which the nodes are the elements and the edges the relationships between the elements. A basic problem is how to extract profile of the topological structure. In the present paper we develop a framework to extract and dis-

play landscape of a complex system. First, from detailed records one can reconstruct the network of the system. Second, dividing the system into different communities, the nodes having the largest degrees and the smallest average shortest path lengths within the communities are used as representatives of the communities, called central nodes, respectively. Meanwhile, one can also find the boundaries that bridge different communities together. The central nodes and the boundaries jointly form a profile of the system. Third, to monitor the evolutionary process of the complex system, one can draw the fission and fusion matrices from present communities to that at the successive next time.

By using this scheme we investigate, as a typical example, the basic research strength in Mainland China. The funds supplied to the top universities in a duration of 12 years are collected, which are separated into three durations. By calculating the similarities between the universities we reconstruct the similarity network of the universities, which evolves in the considered durations. One can find that great changes have occurred, while there exist also significant traces of the impact of the educational administration system constructed more than sixty years ago.

An evolutionary process of the system can be described with a temporal network [34–36]. The proposed framework can be used straightforwardly to map the temporal network to a series of landscapes for the instant networks, which displays a clear picture of the evolution.

From a multivariate time series one can reconstruct the relationship network between the variables [37–40], whose topological structure evolves with time. Hence, the framework provides us a new network-based method for multivariate time series analysis, which has been attracting special attentions in recent years [7–11, 13, 14, 41–45]. We hope our work can stimulate a series of works on evolutionary behaviors of complex systems.

V. ACKNOWLEDGEMENTS

The work is supported by the National Science Foundation of China under Grant Nos 10975099 (H.-J. Yang) and 11505114 (C.-G. Gu), the Shanghai Natural Science Foundation under Grant No. 16ZR1447200 (Q. Xiao), the Program for Professor of Special Appointment (Oriental Scholar) at Shanghai Institutions of Higher Learning under Grant Nos D-USST02 (H.-J. Yang) and QD2015016 (C.-G. Gu), and the Shanghai project for construction of discipline peaks under Grant No.USST-02-SAI (H.-J. Yang and C.-G. Gu), and the Innovation Program of Shanghai Municipal Education Commission under Grant No. 13YZ072 (H.-J. Yang).

VI. APPENDIX

ID	Sub-catalogue	ID	Sub-catalogue	ID	Sub-catalogue
A01	Math	C20	Food science	H09	Nervous system and mental illness
A02	Mechanics	D01	Geography	H10	Medical immunology
A03	Astronomy	D02	Geology	H11	Skin and its appendages
A04	Physics I	D03	Geochemistry	H12	Ophthalmology
A05	Physics II	D04	Geophysics and space physics	H13	Otolaryngology head and neck science
A06	NSFC-Zhong WuYuan mutual funds	D05	Atmospheric science	H14	Oral cranial and maxillofacial science
A08	Big science mutual funds	D06	Marine science	H15	severe medical/trauma, burn/plastic
B01	Inorganic chemistry	D07	Petrochemical joint fund	H16	Oncology
B02	Organic chemistry	E01	Metal materials	H17	Rehabilitation medicine
B03	Physical chemistry	E02	Inorganic non-metallic materials	H18	Imaging and biomedical engineering
B04	Polymer science	E03	Organic polymer materials	H19	Medical pathogenic microorganisms and infection
B05	Analytical chemistry	E04	Metallurgical and mining	H20	Laboratory medicine
B06	Chemical engineering and industrial chemistry	E05	Mechanical engineering	H21	Special medical
B07	Environmental chemistry	E06	Engineering thermal physics and energy utilization	H22	Radiation medicine
C01	Microbiology	E07	Electrical science and engineering	H23	Forensic science
C02	Botany	E08	Building environment and structural engineering	H24	Local epidemiology
C03	Ecology	E09	Water science and Marine engineering	H25	Geriatric medicine
C04	Zoology	F01	Electronics and information system	H26	Preventive medicine
C05	Biophysics, biochemistry and molecular biology	F02	Computer science	H27	Traditional Chinese medicine
C06	Genetics and bio-informatics	F03	Automation	H28	Chinese materia medica
C07	Cellular biology	F04	Semiconductor and information device	H29	Combine traditional Chinese and western medicine
C08	Immunology	F05	Optics and optical electronics	H30	Pharmacology
C09	Neuroscience, cognitive science and psychology	G01	Management science and engineering	H31	Pharmacology
C10	Biomechanics and tissue engineering	G02	business administration	J02	Teenagers and soft subjects
C11	Physiology and integrative biology	G03	Macro management and policy	J03	Program director of the office funds
C12	Developmental biology and reproductive biology	H01	Respiratory system	L01	Agriculture
C13	Agriculture base and crops	H02	circulatory system	L02	Population and health
C14	Plant protection science	H03	digestive system	L03	Resources and environment
C15	Horticulture and plant nutrition	H04	reproductive system	L04	Materials and engineering fields
C16	Forestry	H05	Urinary system	L05	Electronic information field
C17	Livestock and grassland science	H06	Motor system	L06	Bio-diversity conservation areas
C18	Veterinary medicine	H07	Endocrine system nutrition support	L07	Comprehension utilization of mineral resources and new materials
C19	Fisheries science	H08	Blood system	M01	Appoint committee director of office of funds

TABLE A1: Sub-catalogues and their identification numbers.

-
- [1] C. W. Matthiessen, and A. W. Schwarz, *Urban Stud.* **36(3)**,453(1999).
- [2] C. W. Matthiessen, A. W. Schwarz, and S. Find, *J. Knowl. Manag.* **10(5)**,14(2006).
- [3] <http://www.nsf.gov.cn/>
- [4] <http://edu.sina.com.cn/gaokao/2014-10-09/1746437114.shtml>
- [5] T. Li, *J. Beijing Univ. Aero. Astro. (Soc. Sci. Ed.)* **17(4)**,74(2004). (in Chinese)
- [6] P. Jaccard, *New Phytol.* **11**,37(1912).
- [7] J. Zhang, and M. Small, *Phys. Rev. Lett.* **96**,238701(2006)
- [8] Y. Yang, and H.-J. Yang, *Physica A* **387**,1381(2008).
- [9] X. Xu, J. Zhang, and M. Small, *Proc. Natl. Acad. Sci. (USA)* **105(50)**,19601(2008).
- [10] L. Lacasa, B. Luque, F. Ballesteros, J. Luque, and J. C. Nuno, *Proc. Natl. Acad. Sci. USA* **105(13)**,4972(2008).
- [11] R. V. Donner, Y. Zou, J. F. Donges, N. Marwan, and J. Kurths, *New J. Phys.* **12(3)**,033025(2010).
- [12] Y. Wang, J. Qi, J. Yan, S. Deng, G. Wu, et al., *Chin. Sci. Bull.* **57(10)**, 875(2012). (in Chinese)
- [13] Z.-K. Gao, X.-W. Zhang, N.-D. Jin, N. Marwan, and J. Kurths, *Phys. Rev. E* **88**,032910(2013).
- [14] Q. Xiao, X. Pan, X. Li, M. Stephen, H.-J. Yang, et al., *Chin. Phys. B* **23(7)**,078904(2014).
- [15] S. M. Cai, W. Chen, D. B. Liu, M. Tang, and X. Chen, *Physica A* **466**,633(2017).
- [16] M. Janssen, *Phys. Rept.* **295**,1(1998).
- [17] A. D. Mirlin, *Phys. Rept.* **326**,259(2000).
- [18] J. Pipek, and I. Varga, *Phys. Rev. A* **46**,3148(1992).
- [19] I. Varga, and J. Pipek, *Phys. Rev. E* **68**,026202(2003). And the references there-in.
- [20] I. Varga, *Phys. Rev. B* **66**,094201(2002).
- [21] G. Zhu, H. Yang, C. Yin, and B. Li, *Phys. Rev. E* **77**,066113(2008).
- [22] M. Girvan, and M. E. J. Newman, *Proc. Natl. Acad. Sci. (USA)* **99(12)**,7821(2002).
- [23] A. Reka, and A.-L. Barabasi, *Rev. Mod. Phys.* **74(1)**,47(2002).
- [24] D. M. Song, M. Tumminello, W. X. Zhou, and R. N. Mantegna, *Phys. Rev. E* **84**,026108(2011).
- [25] M. C. Munnix, T. Shimada, R. Schafer, F. Leyvraz, T. H. Seligman, et al., *Sci. Rept.* **2**,644(2012).
- [26] S. Kumar, and N. Deo, *Phys. Rev. E* **86**,1679(2012).
- [27] A. Bashan A, R. P. Bartsch, J. W. Kantelhardt, S. Havlin, and P. C. Ivanov, *Nat. Commun.* **3**,702(2012).
- [28] Y.-C. Gao, Z.-W. Wei, and B.-H. Wang, *Inter. J. Mod. Phys. C* **24(2)**,135005(2012).
- [29] X. F. Jiang, T. T. Chen, and B. Zheng, *Sci. Rept.* **4**,5321(2014).
- [30] S. Deng, L. Qiu, Y. Yang, and H. Yang, *Physica A* **441**,62(2016).
- [31] L. Qiu, T. Jia, and H. Yang, *Acta Physica Sinica* **65**,198901(2016).
- [32] C. Liu, and H. Li, *High. Edu. China* **1**,23(2012). (in Chinese)
- [33] Y. Li, and D. Zhang, *Stud. Fin. Account. Edu.* **21(3)**,9(2010). (in Chinese)
- [34] P. Holme, and J. Saramaki, *Phys. Rept.* **519(3)**,97(2012).
- [35] W. Wang, M. Tang, H. F. Zhang, H. Gao, Y. Do, et al., *Phys. Rev. E* **90(4)**,042803(2014).
- [36] S. Boccaletti, J. A. Almendral, S. Guan, I. Leyva, Z. Liu, et al., *Phys. Rept.* **660**,1(2016).
- [37] W.-X. Wang, Y.-C. Lai, C. Grebogi, and J. Ye, *Phys. Rev. X* **1**,021021(2011).
- [38] Z. Zhang, Z. Zheng, H. Niu, Y. Mi, S. Wu, et al., *Phys. Rev. E* **91**,012814(2015).
- [39] X. Han, Z. Shen, W.-X. Wang, and Z. Di, *Phys. Rev. Lett.* **114**,028701(2015).
- [40] W.-X. Wang, Y.-C. Lai, and C. Grebogi, *Phys. Rept.* **644**,1(2016).
- [41] W. J. Xu, L. X. Zhong, P. Huang, T. Qiu, and C. Y. Zhong, *Adv. Compl. Syst.* **18**,1550003(2015).
- [42] Z. K. Gao, Y. X. Yang, P. C. Fang, Y. Zou, C. Y. Xia, et al. *Europhys. Lett.* **109(3)**,30005(2015).
- [43] S. Mutua, C. Gu, and H. Yang, *Plos ONE* **10(11)**:e0143015(2015).
- [44] S. Mutua, C. Gu, and H. Yang, *Chaos* **26(5)**,053107(2016).
- [45] Z.-K. Gao, M. Small, and J. Kurths, *Europhys. Lett.* **116**,50001(2016). And references there-in.

ID	Institution	Type	ID	Institution	Type
1	Zhejiang University	Comprehension	59	University of Science & Technology Beijing	Engineering
2	Peking University	Comprehension	60	Shantou University	Comprehension
3	Nanjing University	Comprehension	61	Beijing University of Technology	Engineering
4	University of Science and Technology of China	Comprehension	62	Shanghai University of Finance & Economics	Financial
5	Tsinghua University	Comprehension	63	Northeastern University	Engineering
6	Fudan University	Comprehension	64	China University of Geosciences	Engineering
7	Shanghai Jiaotong University	Comprehension	65	Harbin Engineering University	Engineering
8	Nankai University	Comprehension	66	Hunan Normal University	Normal
9	China Agricultural University	Agriculture	67	Fuzhou University	Engineering
10	Sun Yat-sen University	Comprehension	68	Xiangtan University	Comprehension
11	Renmin University of China	Comprehension	69	Jiangsu University	Engineering
12	Beijing Normal University	Normal	70	Southwest University	Comprehension
13	Dalian University of Technology	Engineering	71	Yangzhou University	Comprehension
14	Harbin Institute of Technology	Engineering	72	China Medical University	Medical
15	Tianjin University	Engineering	73	Zhejiang Sci-Tech University	Engineering
16	Eastern China University of Science and Technology	Engineering	74	Central University of Finance and Economics	Financial
17	Beijing University of Chemical Technology	Engineering	75	Yanshan University	Engineering
18	Wuhan University	Comprehension	76	ChongQing Medical University	Medical
19	Sichuan University	Comprehension	77	Zhejiang Gongshang University	Financial
20	Shandong University	Comprehension	78	Capital Normal University	Normal
21	Nanjing Agricultural University	Agriculture	79	Huzhou University	Normal
22	Lanzhou University	Comprehension	80	Beijing University of Posts and Telecommunications	Engineering
23	Beihang University	Engineering	81	Shanghai International Studies University	Language
24	Capital Medical University	Medical	82	Anhui Normal University	Normal
25	Southeast University	Engineering	83	Shanghai Normal University	Normal
26	Huazhong University of Science and Technology	Engineering	84	Beijing Forestry University	Agriculture
27	Northeast Normal University	Normal	85	Northwest Normal University	Normal
28	Amoy University	Comprehension	86	Beijing Jiaotong University	Engineering
29	Xi'an Jiaotong University	Engineering	87	Beijing Foreign Studies University	Language
30	Jiangnan University	Engineering	88	Yunnan University	Comprehension
31	Hunan University	Comprehension	89	University of International Business and Economics	Financial
32	Beijing Institute of Technology	Engineering	90	Wuhan University of Technology	Engineering
33	Huazhong Agricultural University	Agriculture	91	Wenzhou University	Comprehension
34	Nanjing Medical University	Medical	92	Tianjin Medical University	Medical
35	Northwestern Polytechnical University	Engineering	93	Tsingtao University of Science & Technology	Engineering
36	Central South University	Engineering	94	University of Shanghai for Science and Technology	Engineering
37	Donghua University	Engineering	95	China University of Political science and Law	LAW
38	Zhejiang Normal University	Normal	96	Henan Normal University	Normal
39	Nanjing University of Science and Technology	Engineering	97	Jiangsu Normal University	Normal
40	Shenyang Pharmaceutical University	Medical	98	Beijing Language and Culture University	Language
41	Soochow University	Comprehension	99	Shanxi University	Comprehension
42	Jilin University	Comprehension	100	Communication University of China	Comprehension
43	Northwestern University	Comprehension	101	China Jiliang University	Engineering
44	Tongji University	Engineering	102	China University of Mining and Technology, Beijing	Engineering
45	Shaanxi Normal University	Normal	103	Shanghai University of Traditional Chinese Medicine	Medical
46	University of Electronic Science and Technology of China	Engineering	104	Hohai University	Engineering
47	Nanjing Technology University	Engineering	105	Huaqiao University	Engineering
48	China University of Petroleum	Comprehension	106	Nanjing University of Information Science & Technology	Engineering
49	Ocean University of China	Comprehension	107	Zhongnan University of Economics and Law	Financial
50	Jinan University	Comprehension	108	Nanjing University of Posts and Telecommunications	Engineering
51	Xidian University	Engineering	109	Anhui University	Comprehension
52	Nanjing Normal University	Normal	110	Hebei University	Comprehension
53	zhejiang university of techonlogy	Engineering	111	Shandong Agricultural University	Agriculture
54	Chongqing University	Engineering	112	Southern Medical University	Medical
55	Southern China Normal University	Normal	113	Fujian Normal University	Normal
56	Shanghai University	Comprehension	114	Zhejiang A & F University	Agriculture
57	Northwestern A&F University	Agriculture	115	Anhui Medical University	Medical
58	Ningbo University	Comprehension	116	North China Electric Power University	Engineering

TABLE A2: The top 116 universities, their identification numbers (here the ranks are used as ID numbers), and their type assignments.

	AGR	COM	ENG	FIN	LAN	LAW	MED	NOR
C1_2		1,2,3,4,5,6,7,10, 18,19,20,41,42	25,26,29,36,44					
C1_2	9,21,33,57, 84,111	8,22,28,43,49, 71,88,99						
C1_3		25,26,29,36,44, 11,58,70,91,110	30,108					27,45,52,55,66,78, 82,83,85,97,113
C1_4		31,56	13,14,15,17,23,32,35,54, 61					
C1_5		48,68	16,37,39,46,47,51,53,59, 63,65,67,69,73,75,80,86, 90,101,102,104,116					
C1_6		50,60					24,34,40,72, 76,92,103, 112,115	
C1_7	114	109	106					38,96
C1_8			64,93,94,105					
C1_9		100		62,74,77, 89,107	98			79
C2_1		1,2,6,7,10,18,19, 20,42,50,60	25,26,29,36, 44,69					
C2_2		3,4,5,8,41						12,27
C2_3		22,28,49,70,71,88	16,30,47,64,73,94,105, 116					
C2_4	9,21,33,57, 111							
C2_5		11		62,74,77, 89,107	87	95		
C2_6	114	91,100,109,110			98			82,83,85,96
C2_7		31,56,68	13,14,15,17,23,32,35,37, 39,46,53,54,59,61,63,67, 86,90,93,102,104,106					
C2_8							24,34,72,76, 92,112,115	
C2_9	84	43,48,99					40,103	45,52,55,66,78,79, 97,113
C2_10		58	51,65,75,80,101,108					
C3_1		1,2,4,5,6,7,10						
C3_2		3,8,18,19,20,28, 41,42	26,29,36,44					
C3_3	9,21,33,57, 84,111,114							
C3_4		11,109,110						38,52,55,78,82,83, 85,96,97,113
C3_5		22,43,50,58,60,70, 71,88,99	30,69					
C3_6			13,14,15,23,25,32,35,39, 46,51,54,59,61,63,65,75, 80,86,90,108					
C3_7		31,48,49,68,91	16,17,37,47,53,64,67,73, 93,94,101,102,104,105, 106,116	77			40	79
C3_8							24,34,72,76, 92,103,112, 115	
C3_9		100		62,74,89, 107	81,87,98	95		

TABLE A3: Type-based distributions of universities in the communities.