Physica A 557 (2020) 124959

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

# Scholar's career switch adhesive with research topics: An evidence from China[☆]

Yinghong Ma [a,*], Le Song [a], Zhaoxun Ji [a], Qian Wang [b], Qinglin Yu [c]

[a] *School of Business, Shandong Normal University, Jinan, 250014, China*
[b] *Jinan Engineering Polytechnic, Jinan, 250200, China*
[c] *Department of Mathematics and Statistics, Thompson Rivers University, BC, Canada*

ARTICLE INFO

*Article history:*
Received 20 October 2019
Received in revised form 18 June 2020
Available online 28 July 2020

*Keywords:*
Two-layer coupling network model
Markov chain
Career stage
Topic type

ABSTRACT

Despite persistent efforts in untangling the mechanism of scientists switching between research topics, little is investigated for the relationship of scholars' career stage switch leading to dynamics of research topics. In this paper, aiming to reveal career stage and its influence on research topics, we construct a two-layer network model, coauthors collaboration network ($\alpha$-layer) for scholars research career stages and papers similarity network ($\beta$-layer) for research topic types, and analyze the relationship between the career stage switch and the topic type change. Applying the data set SMSEC from China to the model, the different statistics of the two layers show different forming mechanisms, the preference attachment and the rule of similarity inherited in the two layers, respectively. The coupling mechanism of the two layers is displayed by correlation of career stages and topic types, and presented by a framework with contributions of new added papers and associated scholars. The results show that the longer of research career is, the bigger contribution on the type of divided topics is; a scholar with large topic scopes would more likely insist in his/her research.



## 1. Introduction

The increasing availability of data on publications, research funding information, collaborations, citations and affiliations provides unprecedented opportunities to explore insights of structures and evolutions of scholarly output in scientific activities [1]. In the meantime, rapid development of state-of-the-art computing softwares enable us to analyze deluge of information by mathematical models to predict evolution of research activities more accurately [2]. Studies have shown that research collaboration can support coauthors in delivering a high research productivity [3,4] or obtaining a far-reaching research impact [5,6], because it increases the number of available communication channels [7,8]. With the increasing number of cooperation among researchers in different disciplines, new research results are better transferred and combined through collaboration. Therefore, new research topics are inevitable emergencies, and create opportunities for scholars to change their research themes or study multiple topics [9], and in turn research themes affect scholars' career trajectory [10]. The research topics affects the scholar career trajectory [10]. It was noted that the number of collaboration works drops along increasing of geographical distance among collaborators [11]. And conversely, research themes tend to fit into authors' environments and interests [12].

☆ This work is supported partly by National Natural Science Foundation of China (No.71471106) and Teaching Reform Project of Shandong Province (Z2016Z025).

* Corresponding author.
*E-mail address:* yinghongma71@163.com (Y. Ma).

https://doi.org/10.1016/j.physa.2020.124959
0378-4371/

As a result, exploring choice and trade-off of continuation of current research topics or switch to other research topics that a scholar faces during his/her career has become a very challenging task. It is particularly hard choice when a scholar's research career maintains high productivity. Hence, a critical decision that a scholar frequently faced is whether or when to continue current topics or terminate it and start new one. Change of topics might be influenced by adjustment of research projects [13], excavation of new research topics [7,14], etc. However, a great deal of research devoted to explore reasons for choices of topics in scholar career trajectory, and found that research interest inspires choices of topics by different expertise and experiences [15–17]. Some other scholars investigated trends of research topics evolving with authors' careers by quantitative methods and models [9,13,18,19].

Process of scholars' career development and span of research topics might be highly correlated. Because career developments are unpredictable, some scholars tend to change their research topics over times. A special angle to detect topic change were focusing on an author's self-citation network, co-authorships and keywords in self-citing articles [20,21]. While different topics had been obtained by different topic detect methods with the same data, document clustering is obtained by standard $k$-Means and the Louvain community detection algorithm using the semantic representation of documents, published as chapter "Clustering articles based on semantic similarity" in article [22] by Wang and Koopman. The process of exploring scholars' career development and relationship between topics is not clearly apart from mass data calculation.

Recent models of processes using multilayer networks showed that ignoring layer interdependencies can lead one to miss why a layer formed the way it did, or draw erroneous conclusions [23]. The main motivation of this work is to detect relationships between scholar collaborations and similarity of their publications, to conform relationships between transformation of research stages and research topics. This work will help researchers to identify topics that an individual scientist is involved, and arrange their career stages to acquire and maintain high academic positions.

The contributions in this work can be summarized in four aspects:

• A two-layer network model is presented, coauthors collaboration network and papers similarity network, coupled by evolution of career stages and topic types of the two layers.

• By computing the collected data set SMSEC, statistics of the two layers are found different which indicate that forming mechanisms of layers are different, in particular, the preference attachment and the rule of similarity inherited in $\alpha$-layer and $\beta$-layer are different.

• A coupling mechanism of the two layers is the contribution of career stages to topic types. The emergence of new connections between scholars and papers empowers model evolution.

• The size of topic is positive proportion to length career; conversely, a scholar with large topic scopes would more likely insist in his/her research.

The rest of this paper is organized as follows: in Section 2, the description of two layers (referred as $\alpha$- and $\beta$-layer) and the construction process of the two-layer network model are presented. Briefly, scholar stages and topics are defined on the two layers, respectively. In Section 3, the method for collecting the data set SMSEC is given. The statistical properties of the model on the data are calculated. Impact of increasing new nodes and new edges are analyzed, the contributions of nodes to new edges are computed in $\alpha$- and $\beta$-layers, respectively. We also use Markov chain to analyze movements of career stages switching and research topic type transition. And in Section 4, the coupling mechanism of scholar career stages and topic types is investigated. Finally, in Section 5, discussion and conclusion on this work is presented, and limitation and future works are also discussed.

## 2. Modeling two-layer network

Scientific collaborators are usually modeled as collaboration networks. Author's career stages and roles of collaboration have been studied. However, relationships between authors' choices of research topics and career stage changes have not been well-studied. Hence, a two-layer network model combining co-authorship and paper similarity is introduced in this section.

### 2.1. Construction of two-layer network

In this subsection, a two-layer network model is presented to reveal evolving relationship between research fields and collaborations. At timestamp $t$, a two-layer network model is denoted by

$$G(t) = (V_\alpha \cup V_\beta, E_\alpha \cup E_\beta \cup E_{\alpha\beta}),$$

where $V_\alpha$ and $V_\beta$ represent the cumulated authors' set and the cumulated papers' set, $E_\alpha$, $E_\beta$ and $E_{\alpha\beta}$ are three cumulated edge sets from initial time to $t$. Note that $\alpha$-layer $= (V_\alpha, E_\alpha)$ and $\beta$-layer $= (V_\beta, E_\beta, W_\beta)$ are two subnetworks, where $W_\beta$ is the edge weight set, representing coauthor collaboration network and paper similarity network, respectively, $E_\alpha = \{(u, v) \in E_\alpha | u, v \in V_\alpha$, and they cooperated in a paper$\}$, $E_\beta = \{(a, b) \in E_\beta | a, b \in V_\beta, sim(a, b) \geq \varepsilon\}$, and $sim(a, b)$ is the similarity of papers $a$ and $b$ in $\beta$-layer, $\varepsilon$ is the threshold which is a given positive real number. Usually, $\alpha$-layer is a binary network in which a link indicates if two authors have cooperated or not, while $\beta$-layer is a weighted network with link weight being similarity, i.e., $w_{ab} = sim(a, b)$. The edge set $E_{\alpha\beta}$ indicates the inter-relations of $\alpha$- and $\beta$-layers, $E_{\alpha\beta} = \{(x, y) | x \in V_\alpha, y \subseteq V_\beta, x$ is one of the authors of $y\}$. A concept example of the two-layer network model is visualized in Fig. 1.

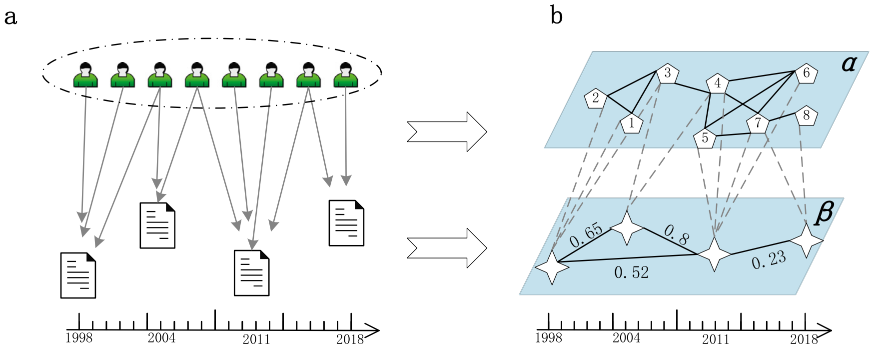



**Fig. 1.** A concept example of the two-layer network model, where there are 8 researchers and 4 publications in different timestamps. (a) is the cumulated relationship between researchers and their publications, and a links represents that a researcher is an author of a paper. (b) indicates the cumulated process of construction of the model corresponding panel (a): pentagons and stars represent researchers and papers respectively. Two pentagons are connected by a link if two researchers are authors of the same paper, which form the $\alpha$-layer; two stars are linked if the similarity of two papers is bigger than the given threshold $\varepsilon$ in $\beta$-layer.

### 2.2. Scholar's career stages in $\alpha$-layer

The $\alpha$-layer in $G(t)$ is a time cumulated collaboration network. Therefore, the node set and edge set increase over times. At timestamp $t$, an author $v$ is called an *old node* if $v$ had published a paper or collaborated with any other author before the time $t$. Correspondingly, an author $v$ is called a *new node* if $v$ never published a paper or cooperated with others, and publishes a paper or cooperates with any other author at the time $t$. In $\alpha$-layer, each new coauthor is added when he/she had new collaborations with the old ones. The collaboration networks had been discussed in many works [24–26], such as small world phenomena, communities and preferential attachment and so on. In this work, the main attention focuses on research career stages and its dynamics.

We define three stages of a scholar in his/her research career from the time of the first publication to the time of the last publication. It was found that an author's first publication is the beginning of his/her scientific career [19,27,28]. The three stages are defined as *novice stage*, *potential stage* and *reputation stage* based on the publication in the first year, the following seven years after the first publication, and more than eight years, respectively. The practical basis for careers length division is from postgraduate education system of China, in which the longest education time for doctor graduate students is seven years. Usually, the first publication takes place in the master graduate period, and more publications would come out during doctor graduate study period, and more papers would be published if graduated doctors work on research institutions or universities. An example to illustrate the three stages of a researcher, shown in Fig. 2. *A*, *B* and *C* are three different researchers, panels (a), (b) and (c) in Fig. 2 show their three stages, respectively.

The three stages of scholars, the novice, the potential and the reputation stages, are defined by the research career's lengths. In order to calculate Markov process, we add a *zero state* to describe the research career state of scientists who quit but may be back to scientific research in the future. In fact, a scholar may retire or quit from any one of the three stages, and a novice scholar can grow into a potential scholar and go on growing into a reputation scholar. Because these states are determined by the length of research career, except the novice state as a transient state, other states can be transformed into itself. Fig. 3(a) shows the complete process of Markov state transition. To calculate state transition probability, we need to combine the real data and use the statistical method to obtain, and calculate the state change frequency of scholars with two adjacent timestamps. For example, if we calculate the probability of scholars in novice state turning into potential state from time $t$ to $t+1$, we need to count the number of scholars in novice state at time $t$, which is recorded as $N^{t}_{novice}$, and count the number of scholars in novice state turning into potential state at time $t+1$, which is recorded as $N^{t\to t+1}_{novice\to potential}$; so from time $t$ to $t+1$, the probability of scholars' transition from novice state to potential state is $\frac{N^{t\to t+1}_{novice\to potential}}{N^{t}_{novice}}$. For transition probability of scholars' state in $\alpha$-layer, the Section 3.4 provides a more detailed description and operation in combination with real data.

### 2.3. Topics in $\beta$-layer

$\beta$-layer$=(V_\beta, E_\beta, W_\beta)$ is defined by paper's similarity. $W_\beta$ is a weight set of time dependent edge, where $w_\beta(a,b) = sim(a,b)$. In this paper, the similarity between two papers $a$ and $b$ in $\beta$-layer is measured by the cosine similarity score [29], which is used to find the most similar of given objects in the fields of information retrieval and text mining [30]:

$$sim(a,b) = \frac{\sum_{i=1}^{n} a_i b_i}{\sqrt{\sum_{i=1}^{n}(a_i)^2 \sum_{i=1}^{n}(b_i)^2}}.$$

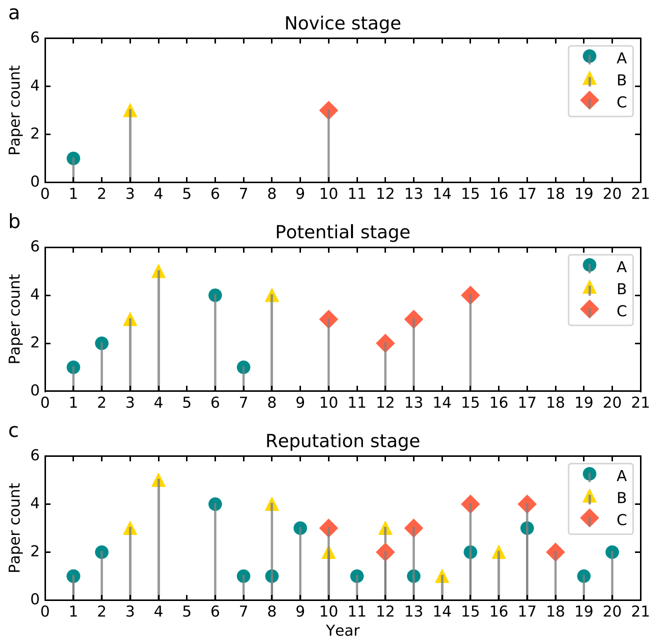
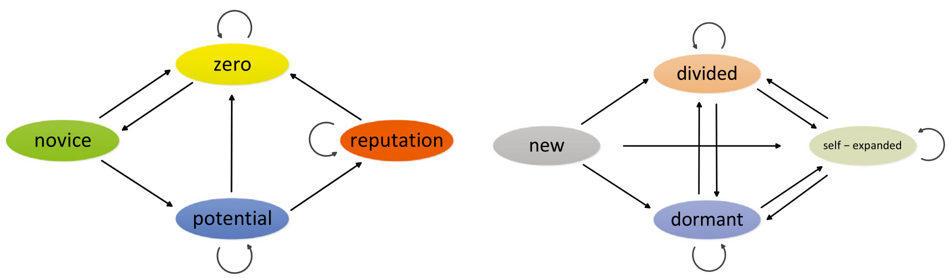



**Fig. 2.** The three stages of researches' careers. *A*, *B* and *C* represent three researches shown in green, yellow and red, respectively. Panels (a), (b) and (c) represent the three stages of three researchers. The vertical axis is the number of authors' publications, and the horizontal axis shows the career length of researches.




(a) transition between the career states. (b) transition between the topic states.

**Fig. 3.** Markov state transition: (a) shows the process of transition between the research career states: the zero, the novice, the potential, and the reputation states; (b) shows the process of state transition between the four types of topics: the new, the self-expanded, the dormant, and the divided topics.

Published papers $a$ and $b$ are represented by vectors combining titles, keywords and abstracts. We use the term frequency–inverse document frequency weighting measures [31], to convert words vectors to numerical vectors. Hence, papers $a$ and $b$ are denoted by $a = (a_1, a_2, \ldots, a_n)$ and $b = (b_1, b_2, \ldots, b_n)$, where $a_i$ and $b_i$ are the $i$th valued feature of papers $a$ and $b$, respectively.

In fact, the similarity of any two papers is greater than 0, so the threshold $\varepsilon$ is a key factor of $\beta$-layer. Too big or too small of $\varepsilon$ can lead to ignored or excessive links and thus drawing erroneous conclusions. In this paper, we set $\varepsilon = avg(sim) + \delta \times \sigma(sim)$, where $\delta$ is a parameter, usually $\delta$ is chosen from 0.7 to 0.8, and $avg(sim)$ and $\sigma(sim)$ are the average and standard deviation of cosine similarities.

We assume that the most similar papers have large similarity score between them. Intuitively, similar papers would have many common terminologies, methods or contents. The relative close contents in weighted network are performed by community [32]. According this assumption, a community is defined as a topic in $\beta$-layer. Thus the number of topics that an author engaged in is the number of communities of his/her published papers belong to, where community structures and their size are detected with the fast unfolding algorithm which is a heuristic method based on modularity optimization [33].

Recall that the size of a community is the number of elements in it. It is naturally to define the size of a topic. In $\beta$-layer, the size of a topic is defined as the size of the community in which the number of most similarity papers are counted. At time $t$, denote the size of a topic by $topic_i$ by $|topic_i|$.

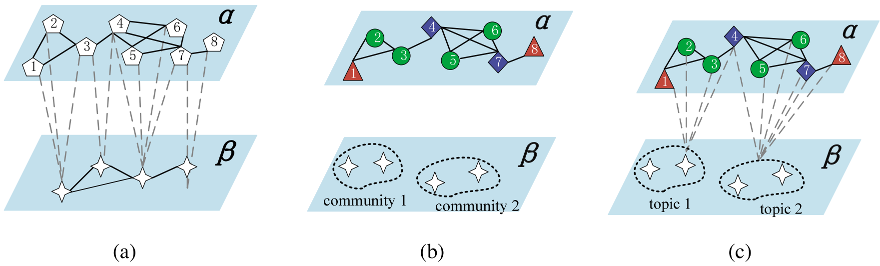



**Fig. 4.** A concept example of the coupling two-layer network model. Red triangles are the novice scholars, green cycles are the potential scholars, and blue diamonds are the reputation scholars. The dotted circles are communities in $\beta$-layer. The dotted lines in panel (a) show authors and papers, while the dotted lines in panel (c) display the authors and their topics.

Here, the definition of the two-layer network is a time evolving model, so the size of community or the topic is time dependent too, and it also shows cumulative effect. In order to analyze which topics cause the evolution of the network, we partition the topics into four types: the new topic, the divided topic, the self-expanded topic, and the dormant topic.

The topic $topic_i$ is called a *new topic* at time $t$, if all the papers that make up the $topic_i$ are newly published at time $t$ in accumulating case. The topic $topic_i$ is called a *divided topic* at time $t$, if the $topic_i$ is obtained at time $t$ by dividing any other topics that has appeared at time $t-1$. In other words, with the papers adding to $\beta$-layer, the average similarities between papers in a topic might decrease with times, since this topic splits into several divided topics, and the original topic disappears. The topic $topic_i$ is called a *self-expanded topic* at time $t$, if the scale of papers belonging to the $topic_i$ is increasing at time $t$ comparing with the scale at time $t-1$, but this topic is not divided at time $t$. The topic $topic_i$ is called a *dormant topic* at time $t$, if the $topic_i$ is fixed from time $t-1$ to $t$. That is, from time $t-1$ to time $t$, there is no new papers in the $topic_i$.

These four topics states are defined by the state of communities which the topic belongs. It is easy to understand that the community evolves with times. As a result, the state change of members in different communities is a dynamic process, that is, the paper states of different topics dynamically changes in the above four states. For the new topic, it can only be new and cannot be transformed from other states. These new topics are bound to change to one of the other three states in the next timestamp. And the three states besides the new topic can be transformed to each other. Fig. 3(b) shows the state transformation of topics.

The probability of state transition about topics in $\beta$-layer is similar with the career stage transition in $\alpha$-layer. In the same way, statistical method is used to calculate the state transition frequency of topics with two adjacent timestamps. The difference is that the way to determine the topic is different from that of scholars. It is certain that scholars are who they are. But we use the method, community division, to determine the topic. Therefore, when we determine the transformation of topics state, it is based on the change of papers in these communities. According to the definitions of the four topic types, it is easy to see that, at two adjacent timestamps $t-1$ and $t$, if the papers in a community are completely composed by newly published papers at time $t$, then the topic represented by the community is the new topic at time $t$; if the papers in a community come from the split of other existing communities, whether there are new papers or not, the topic represented by this community is the divided topic at the time $t$; if the number of papers in a community increases at time $t$, and these increased papers are all newly published papers at time $t$, then the topic represented by the community is self-expanded topic at time $t$; if the papers in a community have not changed, then the topic represented by the community is a dormant topic at the time $t$. In the same way as the calculation of transition probability of scholars' career, we need to count the number of topics in different types at time $t-1$ and $t$ respectively, so as to calculate the transition probability between the topic types. For the transition probability of the topic types in $\beta$-layer, the Section 3.5 provides a more detailed description and calculation applying to the real data.

### *2.4. Coupling of $\alpha$-layer and $\beta$-layer*

Stages of scholar careers and their research topic types might have inherent relationship of $\alpha$ and $\beta$-layers, which worth to be discussed in the following.

According to the definition of two-layer network model, $\alpha$-layer and $\beta$-layer are coupled by the affiliations of authors and papers. As a result, the relationships between scholars and research topics are closely correlated.

Fig. 4 is a concept example to illustrate $\alpha$-layer and $\beta$-layer coupled with their affiliations. There are 8 authors in $\alpha$-layer, 4 published papers in $\beta$-layer. The number of authors are 3, 2, 4 and 2 in the four published papers, shown in dotted lines in Fig. 4(a). In order to show the scholar career stages more clearly, we use the different colors to mark three stages and the number of authors in the novice, the potential and the reputation. In this case, there are 2, 4 and 2 different authors in $\alpha$-layer, respectively. The four papers in $\beta$-layer are divided into 2 communities based on the similarity, and marked with two dotted cycles, shown in Fig. 4(b). And the affiliation relationship between scholars and communities are shown in Fig. 4(c).

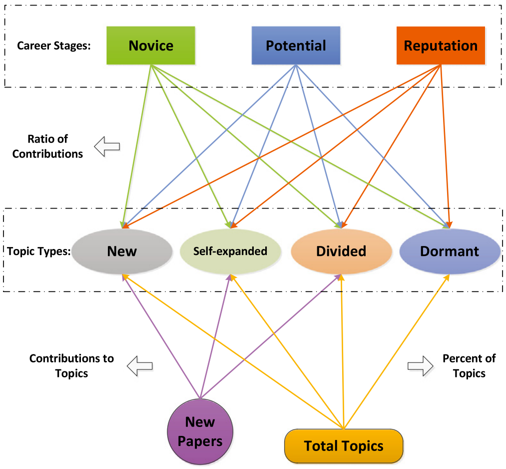



**Fig. 5.** The framework for coupling diagram of scholar career stages and topic types between $\alpha$-layer and $\beta$-layer.

The transitions of stages of careers and the types of topics are displayed in Fig. 3(a) and Fig. 3(b), respectively. While coupling between the two layers and the pushing power for the network evolution depend on new published papers, the new papers add scholars and collaborations between scholars. As a result, we present a framework for coupling diagram of scholar career stages and topic types between $\alpha$-layer and $\beta$-layer, shown in Fig. 5. The coupling mechanism of the two layers is determined by contribution ratios of scholars in different stages to different topic types, and contributions of the new papers to different topic types. In the following, we calculate the coupling between different career stages and different type topics, and contributions of new papers based on the case study in China for this framework.

## 3. Data set and statistical analysis

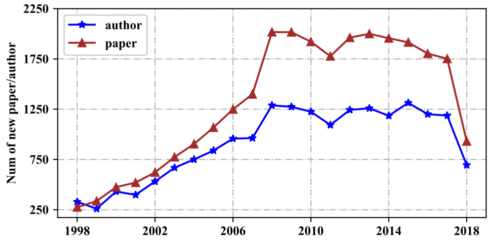
In this section, a collected data set is presented, the two-layer network model is realized from the data. The topological analysis and the evolving structures are also calculated. The coupling mechanism of the two layers are simulated from the data too.

### 3.1. Data set

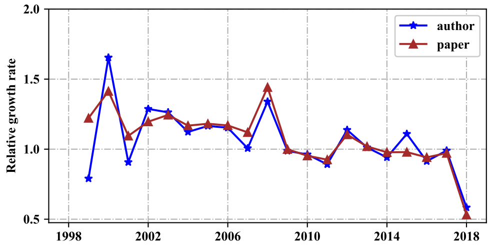
The data set includes the published papers and their authors in the disciplines of management science and engineering in China, denoted by SMSEC. The authors include the directors of the third councils of Society Management Science and Engineering of China obtained through the website of http://www.glkxygc.cn/ and their collaborators from 1998 to 2018. The published papers of those authors are downloaded from CNKI platform of China (https://www.cnki.net/) which is one of the most authority retrieval platforms for Chinese scientific papers. From those downloaded papers, the directors and their collaborators from 1998 to 2018 were chosen. And then we take five attributes of the papers: titles, years, authors, keywords and abstracts.

The SMSEC data set is a representative for the field of Management Science and Engineering. The management science and engineering of China is one of the first-tier disciplines in the management field according to the discipline system of China, which was established by the Chinese Ministry of Education in 1997. It is the milestone for the discipline of management science and engineering, and it is a multi-disciplinary field involving a wide range of knowledge areas including theories and methods from industrial engineering, modern management, economics, mathematics, system science and information science and so on. In other words, this discipline combines scientific research and its applications in various directions, and outputs of the core researchers of this disciplines represent the mainstream research, trends and popular topics in China. We use the data of SMSEC to analyze topic evolution and career stage switch in the two-layer network model.

The data cleaning includes author's name disambiguation and deletion of publications of lacking keywords, published years or abstracts. The disambiguation algorithm for SMSEC data takes authors' research fields, institutions and work experience together to figure out which papers belongs to the same author and eliminate bad data. Publications that contain no keywords, published years, or abstracts were deleted. We end up with 27,672 papers and 18,969 authors from 1998 to 2018 in the SMSEC data set.




(a) The increments.




(b) The relative incremental ratios.

**Fig. 6.** Increments of new authors and new papers each year in the SMSEC data set, where the relative increment ratios of authors and papers are defined by the ratio of increments at $t$ time to that at $t-1$ time.

### *3.2. Statistical analysis of $\alpha$- and $\beta$-layers*

From the SMSEC data set, the real world two-layer network consisting of author collaboration network ($\alpha$-layer) and paper-similarity network ($\beta$-layer) is constructed. Taking a year as a timestamp, the topological properties of the two-layers network, such as numbers of authors and papers, and increments of them, number of edges between authors, number of similarity edges between papers, average degrees and clusters, and so on, are calculated, which are shown in Table 1 of Appendix, where the similarity threshold is set to $\varepsilon = 0.3$.

The blue and the red line charts in Fig. 6(a) show the incremental trends of new authors and new published papers each year, respectively. And the two relative incremental ratio of author (paper) is defined by the ratio of increments of authors (papers) at time $t$ to increments of authors (papers) at time $t-1$ for $t = 1, 2, \ldots$. Fig. 6(b) shows the relative incremental ratios of authors and papers. The year of 2009 is a milestone, as shown in Fig. 6(a), because that the numbers of new authors and papers increase dramatically before 2009, while increments are relatively steady after 2009. And this also shows that the annual increase of authors and papers is positively correlated. By the way, the data of 2018 is not the whole year, so we will not discuss it in the following text. The average cluster coefficient and the average distance in $\alpha$-layer and $\beta$-layer show much different behaves. The $\alpha$-layer shows higher cluster and longer distance than $\beta$-layer, which might be caused by different forming mechanisms, the former is the "richer get richer" preference attachment inherited in collaboration network [3,7], and the latter is resulted by the rule of similarity [31] of "Birds of a feather flock together".

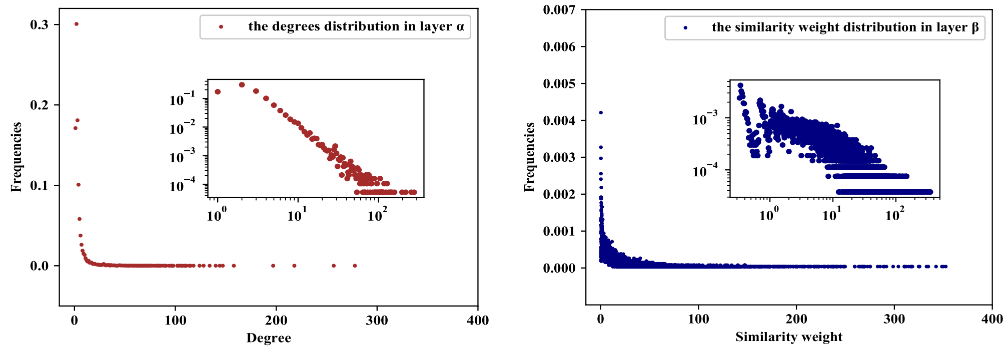
Since $G(t)$ is a time dependent network, node degree of $\alpha$-layer shows cumulative properties, as a collaboration network often looks like a binary networks in which the cumulated degree of a node is the number of coauthors of him/her. While the similarity weight of $\beta$-layer does not display the cumulation because the similarity between two nodes is a fixed value, the node set of $\beta$-layer is the cumulated papers. The node degree distribution and the similarity weights distribution are displayed in Fig. 7. The fitting slops of the two distributions with the power-law function are 2.0042 and 0.7658, respectively, which means the structure of $\alpha$-layer displays the scale-free property [34], while the similarity weight of nodes in $\beta$-layer does not show the power-law distribution. What is more, it can be found from Table 1 that with the increase of years and the expansion of network scale, the average cluster coefficient in $\beta$-layer $\langle C_i \rangle_\beta$ can be considered approximately as a constant value, floating in a small range around 0.54, while the average similarity weight of nodes in $\beta$-layer $\langle W_{node} \rangle_\beta$ shows the trend of growth, so there is no direct relationship between $\langle C_i \rangle_\beta$ and $\langle W_{node} \rangle_\beta$ in terms of quantity.

### *3.3. New edges of the two layers*

$G(t)$ is defined as a time evolving model, and the evolution of it is caused by the new nodes and new edges added. A new edge links two nodes: two authors, two papers, or one paper and one author. In this subsection, new edges and the associated pairs of nodes will be discussed to explore growth mechanism of new edges in $\alpha$-layer and $\beta$-layer. We find that, with development and maturity of a discipline, cooperation between new and old authors tends to be stable and more inclined to the cooperation with old authors in $\alpha$-layer; in $\beta$-layer, the old nodes are highly active and play an important role in topic cohesion. The contributions of the nodes for the evolution of $G(t)$ are discussed below.

#### *3.3.1. New edges in $\alpha$-layer*

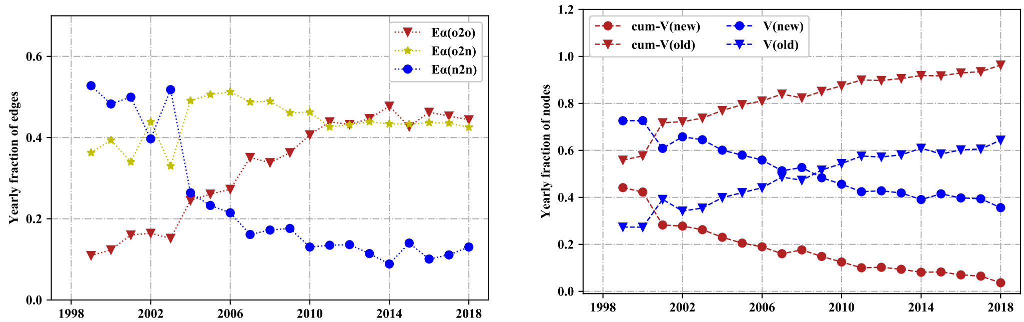
In $\alpha$-layer, we defined the new author or the new node by the year of him/her first published paper, and then call an old author or the old node from then on. Over times, the number of authors grows, and the collaborations between them increase too, where the collaborations are joined among the new authors, the old authors or between them. As a result, three kinds of new edges would emerge from the new collaborations between authors and coauthors. At time $t$, an old node is called as an *active old one* if he/she publishes a paper at time $t$. We assume that the three types of new edges came out between two new authors, between an old author and a new, and between two old authors, denoted by




(a) The node degree distribution of $\alpha$-layer.

(b) The similarity weight distribution of $\beta$-layer.

**Fig. 7.** The inner panels of the two graphs are the log–log plots of the two distributions respectively. The degree distribution of $\alpha$-layer is the node degree cumulated over times. The similarity weights of $\beta$-layer is the similarity of papers at the end of running time.




(a) The ratios of the new edges.

(b) The ratios of nodes.

**Fig. 8.** The growth trends of the nodes and the new edges in $\alpha$-layer. (a) shows the proportions of the three types of new edge collaborations, $E_\alpha(n2n)$, $E_\alpha(o2n)$ and $E_\alpha(o2o)$ shown in blue dots, yellow stars and red triangles, respectively. (b) shows the ratios of new added coauthors and the active old authors at each time in blue curve lines. And the red curves are ratios of the cumulated number of the old authors and the new author to the total number of authors at each timestamp, respectively. The data in this figure is based on Table 1 of Appendix.

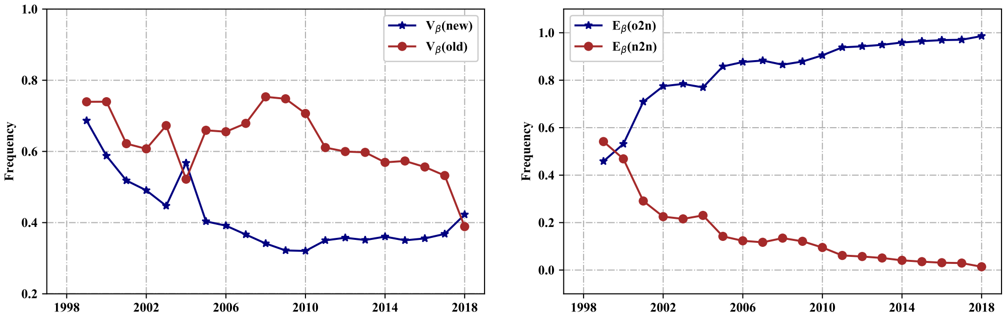
$E_\alpha(n2n)$, $E_\alpha(o2n)$, and $E_\alpha(o2o)$, respectively. The growth rates of the three types of new edges are shown in Fig. 8(a) with blue dots, yellow stars, and red triangles, respectively. In the three types of edges, the size of the set $E_\alpha(n2n) \cup E_\alpha(o2n)$ is greater than $E_\alpha(o2o)$ before the year 2010, and from then on, $E_\alpha(n2n)$ is lower than 0.2, $E_\alpha(o2n)$ and $E_\alpha(o2o)$ tend to 0.4. We find that, with development of the discipline, the proportion of three types of new edges tends to be stable. That is, with development and maturity of the discipline, cooperation between the old authors and the new authors will become stable, and more focus on the new authors with the old authors as well as the old authors with the old authors, which can be found in Fig. 8(a).

In fact, the absolute increments of the new edges are not proportional to the increments of the new nodes, shown in Table 1 of Appendix. Clearly, the new edges are contributed by both the new nodes and the active old nodes. Analyzing the number of active old nodes and the new nodes further, the ratio of the new nodes decreases while the active old ones increases. The trends of the two curves go crossover around the year 2009, and then go to opposite directions, shown in Fig. 8(b). This phenomenon also shows that, with development of the discipline, active old nodes play an more important role in author cooperations.

In Table 1 of Appendix, the size of new nodes is not proportional to the cumulated old nodes in the total node set. However, with growth of the old authors set, the impact of the old authors set became more apparent, that the contribution of the old nodes to the total incremental edges is greater than the new ones. It is not hard to comprehend the above phenomena since the new nodes became an old one after he/she first involvement in the collaboration. The node set become bigger and bigger, so does the base of the new nodes.

*3.3.2. New edges in $\beta$-layer*

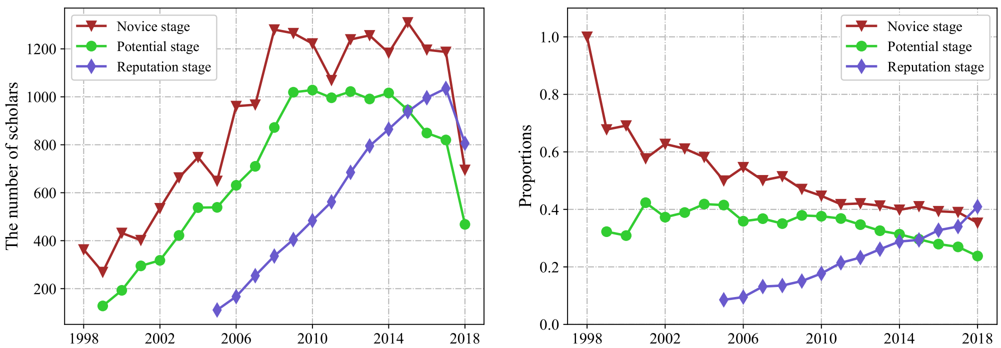
Fig. 6(a) shows that the relative ratios of paper increments near to 1, which implies that the paper publications of each year is relative stable from 2009 to 2017. At time $t$, the emergence of new edges in $\beta$-layer includes the new edges between two new papers, an old paper and a new paper. Denote the two types of new edge sets by $E_\beta(n2n)$ and $E_\beta(n2o)$, respectively. We still call an old node as an active old nodes who takes part in the new edges forming in $E_\beta(n2o)$. Then




(a) The ratios of nodes.

(b) The ratios of the two new edges.

**Fig. 9.** The growth trends of the nodes and the new edges in $\beta$-layer. (a) the red dotted line shows the ratio of the active old nodes to the total old nodes, the blue star line shows the ratio of the newly added nodes (i.e., papers) to the sum of the active old nodes and the newly added nodes. (b) shows the proportions of the two new edge sets $E_\alpha(n2n)$ and $E_\alpha(o2n)$ in $\triangle E_\beta$ of Table 1 of Appendix.




(a) The number of scholars in three stages.

(b) The ratio of scholars in three stages.

**Fig. 10.** (a) shows the number of papers in three stages of scholars. (b) shows the proportions of scholars in their three stages at each year. The ratio of the reputation stage from 1998 to 2004 is zero because of its definition.

the ratios of the active old node to the total old nodes and the ratios of the new nodes to the total two nodes of the new edges including the active old nodes and the new added nodes are shown in Fig. 9(a). In addition to 2018, among all the old nodes, the proportion of active old nodes is always higher than 50%, which indicates that the activities of old nodes are very high. With development of the discipline, the proportion of active old nodes will be higher than 50% of the nodes connected by new edge at each timestamp. It might be caused by that the increasing of new edges is more dependent on the old nodes, so the old nodes are crucial for cohesion and development of the discipline.

The contribution of $E_\beta(n2o)$ outperforms the set of $E_\beta(n2n)$ for the growth of new edges, shown in Fig. 9(b), that indicates the agglomeration of the active old papers is bigger than the new ones. The active old nodes contribute almost all the new edges at the end of evolving time. So, we believe that the big topics in $\beta$-layer are the accumulated result of the active old papers. We will talk about it in the Section 3.5.

### 3.4. Career stages in $\alpha$-layer

By the example in Fig. 2, the panel (a) shows the time of their first publications. Scholars A, B and C are at their novice stage at the first year, the third year and the tenth year. Then the potential stage of the scholar A is from the 2nd year to the eighth year, and the reputation stage began from ninth year. From then on, the scholar A has a big probability to stay steady in his reputation stage until he retired or quit the stage.

The incremental productions of a scholar had required steady supports of publications [13]. Therefore, the scholar's career length and the scholar's collaboration relationships are uncertain. We need to explore probability of scholars to stay their research careers. Computing the three stages of a scholar's career based the cumulated data set SMSEC, the average proportions of scholars in three stages are 60.87%, 26.05% and 13.07%, respectively. And for each timestamp, the number and proportion of scholars in the three stages are shown in Figs. 10(a) and 10(b), respectively. At the early 10 years, a great deal of scholars is in the novice stage, and then being relative steady from the year 2007. While the ratios of the novice in the population decrease with times. So does the potential scholars. However, the reputation scholars increase each year until to be the biggest in the three stages.

**Table 1**
The structure information of two-layer network on data set SMSEC from 1998 to 2018.

| Time | $\lvert V_\alpha \rvert$ | $\lvert V_\beta \rvert$ | $\lvert E_\alpha \rvert$ | $\lvert E_\beta \rvert$ | $\lvert E_{\alpha\beta} \rvert$ | $\langle Deg \rangle_\alpha$ | $\langle W_{node} \rangle_\beta$ | $\langle C_i \rangle_\alpha$ | $\langle C_i \rangle_\beta$ | $\langle Dist \rangle_\alpha$ | $\langle Dist \rangle_\beta$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1998 | 329 | 142 | 511 | 151 | 190 | 3.106 | 2.1276 | 0.879 | 0.509 | 1.617 | 3.331 |
| 1998–1999 | 589 | 388 | 902 | 585 | 508 | 3.063 | 3.015 | 0.837 | 0.597 | 2.565 | 5.125 |
| 1998–2000 | 1019 | 804 | 1647 | 1750 | 1144 | 3.233 | 4.353 | 0.834 | 0.536 | 2.897 | 6.996 |
| 1998–2001 | 1409 | 1258 | 2707 | 3526 | 2418 | 3.842 | 5.606 | 0.826 | 0.554 | 6.334 | 6.327 |
| 1998–2002 | 1911 | 1823 | 3615 | 6158 | 2644 | 3.783 | 6.719 | 0.819 | 0.55 | 8.571 | 6.026 |
| 1998–2003 | 2545 | 2540 | 6255 | 10 089 | 4693 | 3.943 | 7.944 | 0.81 | 0.563 | 9.747 | 5.935 |
| 1998–2004 | 3257 | 3422 | 7674 | 17 139 | 5351 | 3.843 | 10.017 | 0.804 | 0.56 | 9.616 | 5.597 |
| 1998–2005 | 4087 | 4433 | 9484 | 27 524 | 7348 | 3.758 | 12.418 | 0.801 | 0.558 | 8.847 | 5.391 |
| 1998–2006 | 5044 | 5648 | 11 373 | 41 186 | 9614 | 3.262 | 14.584 | 0.8 | 0.554 | 8.494 | 5.204 |
| 1998–2007 | 6007 | 6998 | 14 003 | 61 327 | 12 392 | 3.788 | 17.527 | 0.797 | 0.552 | 7.639 | 5.003 |
| 1998–2008 | 7297 | 8948 | 14 003 | 96 993 | 16 081 | 3.84 | 21.769 | 0.789 | 0.546 | 7.106 | 4.789 |
| 1998–2009 | 8571 | 10 920 | 16 924 | 139 393 | 19 888 | 3.951 | 25.53 | 0.784 | 0.546 | 6.491 | 4.65 |
| 1998–2010 | 9797 | 12 801 | 19 791 | 181 427 | 23 279 | 4.041 | 28.346 | 0.784 | 0.545 | 6.491 | 4.596 |
| 1998–2011 | 10 890 | 14 536 | 22 615 | 223 275 | 26 611 | 4.154 | 30.72 | 0.782 | 0.544 | 6.26 | 4.543 |
| 1998–2012 | 12 133 | 16 454 | 25 738 | 270 743 | 29 837 | 4.244 | 32.909 | 0.781 | 0.539 | 6.035 | 4.502 |
| 1998–2013 | 13 391 | 18 419 | 29 057 | 323 312 | 33 332 | 4.341 | 35.553 | 0.779 | 0.527 | 5.868 | 4.59 |
| 1998–2014 | 14 575 | 20 289 | 32 276 | 361 991 | 37 357 | 4.429 | 37.842 | 0.779 | 0.529 | 5.729 | 4.411 |
| 1998–2015 | 15 888 | 22 196 | 36 019 | 435 127 | 41 313 | 4.535 | 39.604 | 0.78 | 0.526 | 5.609 | 4.421 |
| 1998–2016 | 17 088 | 23 965 | 39 602 | 493 462 | 45 076 | 4.636 | 41.182 | 0.781 | 0.532 | 5.493 | 4.379 |
| 1998–2017 | 18 276 | 25 688 | 43 249 | 550 884 | 47 734 | 4.734 | 42.89 | 0.782 | 0.531 | 5.418 | 4.351 |
| 1998–2018 | 18 969 | 26 606 | 45 509 | 582 017 | 50 197 | 4.8 | 43.751 | 0.783 | 0.531 | 5.363 | 4.336 |

**Table 2**
The number of types of topics in $\beta$-layer.

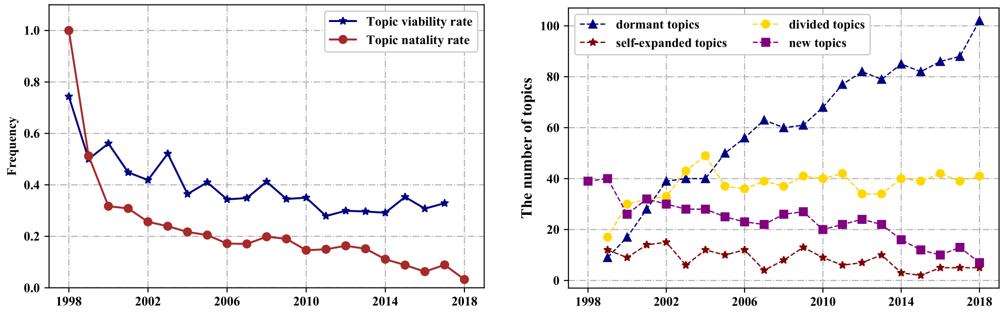
| Year | $\lvert The\ new \rvert$ | $\lvert The\ dormant \rvert$ | $\lvert The\ self\ expanded \rvert$ | $\lvert The\ divided \rvert$ |
|---|---|---|---|---|
| 1998 | 39 | 0 | 0 | 0 |
| 1999 | 40 | 9 | 12 | 17 |
| 2000 | 26 | 17 | 9 | 30 |
| 2001 | 33 | 28 | 14 | 32 |
| 2002 | 30 | 39 | 15 | 33 |
| 2003 | 28 | 40 | 6 | 43 |
| 2004 | 28 | 40 | 12 | 49 |
| 2005 | 25 | 50 | 10 | 37 |
| 2006 | 22 | 56 | 12 | 38 |
| 2007 | 22 | 63 | 4 | 40 |
| 2008 | 26 | 60 | 8 | 37 |
| 2009 | 27 | 61 | 13 | 41 |
| 2010 | 20 | 68 | 9 | 40 |
| 2011 | 22 | 77 | 6 | 42 |
| 2012 | 24 | 82 | 7 | 34 |
| 2013 | 22 | 79 | 10 | 34 |
| 2014 | 16 | 85 | 3 | 40 |
| 2015 | 12 | 82 | 2 | 40 |
| 2016 | 9 | 86 | 5 | 43 |
| 2017 | 13 | 89 | 5 | 39 |
| 2018 | 5 | 102 | 5 | 43 |

The stages of a scholar transform from the novice to the potential, and then to the reputation, shown in Fig. 2. In this work, that a scholar quits or retires from his/her research career is considered to enter the zero state. The quitting, the novice, the potential and the reputation are denoted by the four states $s_0, s_1, s_2, s_3$, respectively. For convenience, we denote the four states by a set $S = \{s_0, s_1, s_2, s_3\}$ and the transform probability between the four states by a matrix $P = (p_{ij})_{4\times4}$. The transform probability $p_{ij} \geq 0$, and $\sum_{j\in S} p_{ij} = 1$ for any $i \in S$, so $P$ is a Markov chain [35].

Since the state transition of researches is a Markov process, the values in $P$ is obtained by calculating SMSEC data set from the year 1998 to 2017, $t = 1, 2, \ldots, 20$. In Statistics, the frequency is an approximating distribution by computing the real data. Therefore, we use the frequencies of states transition to estimate the transition probability $P$.

$$P = (p_{ij})_{4\times4} = \begin{pmatrix} 0.92 & 0.08 & 0 & 0 \\ 0.61 & 0 & 0.39 & 0 \\ 0.2 & 0 & 0.7 & 0.1 \\ 0.1 & 0 & 0 & 0.9 \end{pmatrix}, \tag{1}$$

where $i, j \in \{s_0, s_1, s_2, s_3\}$. The values in the matrix $P$ display the possibilities of the three stages transforms. $p_{10} = 0.61 > p_{20} = 0.2 > p_{30} = 0.1$ show that the probability of a scholar steady in research career correlates positively to his research length. The transition probabilities $p_{12} = 0.39$ and $p_{23} = 0.1$ show that 39% and 10% of scholars transform from the novice to potential, and the potential to the reputation, respectively. However, 90% of the reputations stay at their stage, and




(a) The natality and viability of topics. (b) The proportions of the four types topics.

**Fig. 11.** Panel (a) shows the distributions of the natality and viability of topics defined in Section 3.5.1. The trend of the viability is steady, while the curve of the natality decreases gradually to 0. Panel (b) is the proportion of the four types of topics in each year. The divided and the self-expanded are steady; the number of dormant topics increases, while the number of new topic decreases.

with a tiny chance to quit from research career. Totally, only 8% ($p_{01} = 0.08$) of a scholar re-entering scientific research career, while 91% $= (p_{10} + p_{20} + p_{30})$ of the scholars eventually quit from scientific research careers.

The equilibrium distribution $\pi_0$ of the transform matrix $P$ is calculated by the equation $P\pi_t = \pi_t \to \pi_0$ for $t = 1, 2, \ldots$, and an equilibrium state is $\pi_0 = (0.78, 0.06, 0.08, 0.08)$. It shows 78% of the observed scholars in quit state, 22% = (6% +8%+8%) of them insist on research.

### 3.5. Topics and their types in $\beta$-layer

In this section, statistical properties of topics and their classification in $\beta$-layer are discussed.

#### 3.5.1. Size of topics, natality and viability

A topic is defined by a community in $\beta$-layer, denote the set of topics and its size at time $t$ by $topics(t)$ and $|topics(t)|$, respectively. The number of topics and the sizes of them in SMSEC data set from 1998 to 2018 are calculated and shown in Table 2 in Appendix. Adding all the 21 timestamp $\beta$-layers together, there are 2611 topics from 1998 to 2018. Among them, there are more than 52% of the topics with size 2, and 1234 topics with size more than 3. Computing the size distribution of topics among 1234 topics, 25.04% of them have 3 papers, 11.43% of them have 4 papers, the numbers of topics with the size from tens to hundreds to thousands are less than ten, shown in Fig. 14 and Table 3 of Appendix. There are 1827 topics with scale 2, 3 and 4, accounting for 69.97% of all the topics. However, they only account 4245 papers, 15.34% of the total number of papers. There are 548 topics with size larger than 50, 20.99% of the total topics, but include 19806 papers, 71.57% of the total papers. Therefore, the largest portion of papers belongs to the large scale topics. A Markov process was defined in occupation number space to simulate the spontaneous generation [36], to predict the nonlinear growth properties associated with the emergence of new topics. While, development of a discipline must be accompanied by the change of topic size, which naturally correlates with the birth and the survival of topics. The*topic natality* defined as the ratio of new topics to all the topics in the current timestamp, and the *topic viability* is the proportion of active topics (i.e., both of the self-expanded and the divided topics) to all the topics. Denoted the natality at time $t$ by $natality_t$, where $natality_t$ is the ratio of the number of new topics $N^t_{new}$ at time $t$ to the total number of topics at time t $N^t_{total}$. Denoted the viability at time $t$ by $viability_t$, the ratio of the number of divided topics $N^t_{divided}$ and self-expanded topics $N^t_{self-expanded}$ at time $t$ to the total number of topics $N^{t-1}_{total}$ at time $t-1$. That is,

$$natality_t = \frac{N^t_{new}}{N^t_{total}}, \text{ and } viability_t = \frac{N^t_{divided} + N^t_{self-expanded}}{N^{t-1}_{total}}.$$

The natality and the viability of topics in SMSEC data set are shown in Fig. 11(a). With development of the discipline, the viability of the topic first drops slowly before at the first 10 years, and then relative stable. But the natality of topics declines, and the rate of decline slows down gradually. This means that the proportion of the new topics is reduced, and more and more topics would become dormant topics. This is closely related to the type change of the topic. We will discuss it in Section 3.5.2.

#### 3.5.2. Topic types

Four types of topics, the new, the divided, the self-expanded, and the dormant, were defined in Section 2.3. With the new papers added each year, the types of these topics would changes. As a result, how to distinguish which topics pushing development of the network is an attractive idea.

Fig. 11 shows the number of topics in the four types: the number of the dormant topics is the largest, consistent with the decline of topic viability shown in Fig. 11(a), followed by the number of the divided topics that represent the separation

**Table 3**
The distribution of sizes of topics in $\beta$-layer, where the columns *size* and ♯ indicate the sizes of topics and corresponding number of topics with this size.

| Size | ♯ | Size | ♯ | Size | ♯ | Size | ♯ | Size | ♯ | Size | ♯ | Size | ♯ | Size | ♯ | Size | ♯ | Size | ♯ | Size | ♯ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 1377 | 47 | 2 | 89 | 1 | 148 | 3 | 208 | 1 | 281 | 1 | 350 | 1 | 463 | 1 | 648 | 1 | 825 | 1 | 1102 | 1 |
| 3 | 309 | 48 | 3 | 90 | 1 | 150 | 1 | 209 | 1 | 283 | 2 | 352 | 1 | 464 | 1 | 649 | 1 | 826 | 1 | 1128 | 1 |
| 4 | 141 | 50 | 2 | 91 | 2 | 151 | 1 | 211 | 2 | 285 | 1 | 354 | 1 | 468 | 1 | 650 | 1 | 828 | 1 | 1132 | 1 |
| 5 | 54 | 51 | 2 | 92 | 2 | 152 | 1 | 216 | 2 | 286 | 1 | 358 | 2 | 473 | 1 | 653 | 1 | 830 | 1 | 1135 | 1 |
| 6 | 21 | 52 | 4 | 93 | 2 | 153 | 1 | 217 | 1 | 289 | 1 | 361 | 1 | 475 | 3 | 654 | 1 | 832 | 1 | 1154 | 2 |
| 7 | 23 | 53 | 5 | 94 | 2 | 154 | 1 | 218 | 1 | 291 | 1 | 364 | 1 | 478 | 1 | 660 | 1 | 854 | 1 | 1157 | 1 |
| 8 | 19 | 54 | 3 | 95 | 2 | 156 | 2 | 219 | 2 | 292 | 3 | 369 | 1 | 480 | 1 | 661 | 1 | 868 | 1 | 1158 | 1 |
| 9 | 13 | 55 | 2 | 96 | 1 | 159 | 2 | 221 | 2 | 293 | 2 | 372 | 1 | 482 | 1 | 662 | 1 | 871 | 1 | 1202 | 1 |
| 10 | 5 | 56 | 2 | 98 | 1 | 161 | 1 | 223 | 1 | 294 | 1 | 374 | 1 | 484 | 4 | 664 | 1 | 874 | 1 | 1211 | 1 |
| 11 | 1 | 57 | 3 | 100 | 1 | 162 | 3 | 224 | 1 | 296 | 1 | 375 | 1 | 486 | 1 | 673 | 1 | 889 | 1 | 1214 | 1 |
| 12 | 5 | 58 | 3 | 101 | 1 | 163 | 1 | 226 | 3 | 298 | 1 | 377 | 1 | 495 | 1 | 676 | 1 | 899 | 1 | 1221 | 1 |
| 13 | 4 | 59 | 1 | 102 | 1 | 164 | 6 | 227 | 1 | 299 | 1 | 384 | 1 | 503 | 1 | 677 | 1 | 903 | 1 | 1254 | 1 |
| 14 | 2 | 60 | 1 | 105 | 3 | 165 | 5 | 229 | 2 | 300 | 1 | 387 | 2 | 511 | 1 | 680 | 1 | 905 | 1 | 1261 | 1 |
| 15 | 2 | 61 | 1 | 107 | 1 | 166 | 1 | 231 | 1 | 302 | 2 | 390 | 1 | 512 | 1 | 687 | 1 | 909 | 1 | 1291 | 1 |
| 16 | 5 | 62 | 2 | 108 | 3 | 167 | 1 | 233 | 1 | 304 | 1 | 391 | 1 | 514 | 1 | 704 | 1 | 918 | 1 | 1302 | 1 |
| 17 | 5 | 63 | 1 | 109 | 1 | 168 | 1 | 235 | 1 | 307 | 1 | 393 | 2 | 529 | 1 | 714 | 1 | 923 | 1 | 1341 | 1 |
| 18 | 4 | 64 | 2 | 111 | 3 | 169 | 2 | 238 | 1 | 308 | 1 | 395 | 1 | 533 | 1 | 717 | 1 | 925 | 2 | 1366 | 1 |
| 19 | 3 | 65 | 2 | 112 | 2 | 170 | 2 | 239 | 3 | 310 | 1 | 399 | 1 | 535 | 2 | 718 | 1 | 928 | 1 | 1388 | 1 |
| 20 | 4 | 66 | 1 | 115 | 4 | 173 | 1 | 240 | 1 | 311 | 4 | 402 | 1 | 543 | 2 | 728 | 1 | 938 | 1 | 1414 | 1 |
| 21 | 6 | 67 | 2 | 116 | 2 | 175 | 1 | 242 | 1 | 312 | 1 | 404 | 1 | 545 | 2 | 729 | 2 | 940 | 1 | 1428 | 1 |
| 22 | 3 | 68 | 1 | 118 | 1 | 176 | 2 | 243 | 1 | 314 | 1 | 410 | 1 | 550 | 1 | 735 | 1 | 945 | 1 | 1449 | 1 |
| 23 | 3 | 69 | 1 | 119 | 3 | 179 | 1 | 246 | 1 | 315 | 1 | 411 | 2 | 551 | 2 | 737 | 1 | 949 | 2 | 1459 | 1 |
| 24 | 2 | 70 | 1 | 120 | 1 | 180 | 3 | 247 | 1 | 316 | 2 | 413 | 1 | 555 | 1 | 742 | 1 | 957 | 1 | 1479 | 1 |
| 26 | 1 | 71 | 2 | 121 | 1 | 181 | 1 | 248 | 2 | 317 | 1 | 414 | 2 | 557 | 1 | 749 | 1 | 969 | 1 | 1484 | 1 |
| 27 | 4 | 72 | 1 | 122 | 1 | 182 | 1 | 252 | 1 | 318 | 2 | 416 | 1 | 558 | 1 | 755 | 1 | 971 | 1 | 1539 | 1 |
| 28 | 3 | 73 | 1 | 124 | 3 | 183 | 1 | 253 | 1 | 320 | 2 | 422 | 1 | 568 | 1 | 758 | 1 | 979 | 1 | 1576 | 1 |
| 29 | 2 | 74 | 6 | 125 | 1 | 184 | 2 | 254 | 3 | 322 | 2 | 423 | 1 | 570 | 1 | 760 | 2 | 981 | 1 | 1713 | 1 |
| 30 | 3 | 75 | 2 | 126 | 1 | 187 | 1 | 255 | 1 | 323 | 1 | 430 | 2 | 578 | 1 | 762 | 1 | 992 | 1 | 1944 | 1 |
| 31 | 1 | 76 | 1 | 128 | 1 | 189 | 1 | 256 | 1 | 324 | 3 | 431 | 1 | 589 | 1 | 764 | 1 | 1000 | 2 | 2018 | 1 |
| 32 | 3 | 77 | 1 | 129 | 1 | 191 | 1 | 257 | 1 | 332 | 1 | 434 | 1 | 591 | 1 | 771 | 1 | 1023 | 1 | | |
| 34 | 3 | 79 | 4 | 131 | 1 | 192 | 2 | 262 | 1 | 333 | 2 | 437 | 1 | 594 | 1 | 778 | 1 | 1025 | 1 | | |
| 35 | 5 | 80 | 3 | 133 | 1 | 193 | 2 | 263 | 1 | 335 | 2 | 438 | 1 | 598 | 1 | 779 | 1 | 1040 | 1 | | |
| 36 | 3 | 81 | 4 | 136 | 3 | 194 | 2 | 264 | 1 | 336 | 4 | 440 | 1 | 605 | 2 | 780 | 2 | 1048 | 1 | | |
| 37 | 1 | 82 | 2 | 138 | 2 | 195 | 1 | 269 | 1 | 337 | 1 | 442 | 1 | 612 | 1 | 783 | 1 | 1049 | 1 | | |
| 38 | 4 | 83 | 1 | 139 | 1 | 196 | 3 | 271 | 3 | 339 | 1 | 443 | 2 | 615 | 1 | 789 | 2 | 1054 | 1 | | |
| 39 | 4 | 84 | 1 | 140 | 3 | 199 | 1 | 273 | 1 | 340 | 2 | 444 | 3 | 617 | 1 | 795 | 1 | 1070 | 1 | | |
| 40 | 4 | 85 | 1 | 142 | 2 | 201 | 1 | 276 | 2 | 342 | 1 | 450 | 1 | 618 | 1 | 801 | 1 | 1079 | 1 | | |
| 42 | 3 | 86 | 2 | 144 | 2 | 202 | 1 | 277 | 1 | 343 | 1 | 452 | 1 | 620 | 1 | 808 | 1 | 1084 | 1 | | |
| 44 | 1 | 87 | 1 | 145 | 1 | 203 | 1 | 278 | 1 | 345 | 2 | 458 | 1 | 637 | 4 | 809 | 1 | 1097 | 1 | | |
| 45 | 2 | 88 | 2 | 146 | 2 | 204 | 1 | 279 | 2 | 347 | 2 | 459 | 1 | 640 | 1 | 811 | 1 | 1100 | 1 | | |

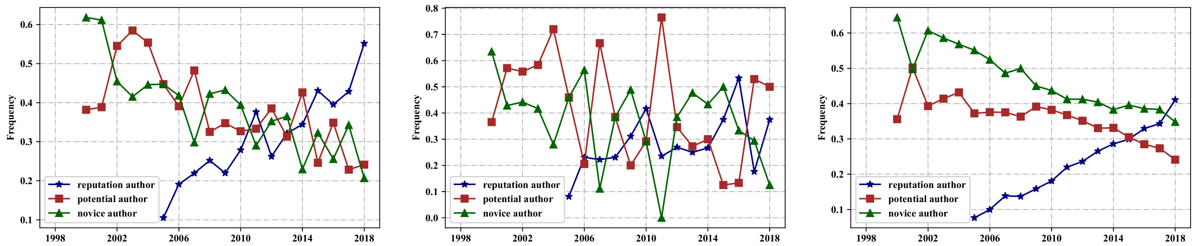
and fusion of topics. Note that the separation and fusion of topics are an important 'behaviors' of topic evolution. Analyzing further, we find that 99% of the divided topics come from the cross integration of existing topics, rather than directly split from a single topic. A possible reason is that management science and engineering itself is an interdisciplinary discipline, scholars with different disciplines work together that push the knowledge fusion of the disciplines, resulting in a large number of mainstream topics split and merged.

In order to further explore the evolution of different types, we use the Markov process to simulate the transit probabilities of the topic types, the transit directions and their relations (Fig. 3(b)). Denote the four types of topics as a set $S' = \{s'_0, s'_1, s'_2, s'_3\}$ and the simulate transition matrix $P' = (p'_{i'j'})_{4\times4}$, where $s'_0, s'_1, s'_2, s'_3$ are the types of the new, the dormant, the self-expanded, and the divided topic, respectively.

From Fig. 3(b), we see that none of four types of topics change to the new topic, and the new topic have probability to change to other type of topics, so all entries of the first column of the transition matrix $P'$ are 0. Hence, $4 \times 4$ matrix $P'$ can be reduced to $P' = (p'_{i'j'})_{3\times4}$. So, we calculate the probability of the topic type transition every year (taking the probability by frequency), and take the average value of each type as the value in the transition matrix of $P'$.

$$P' = (p'_{i'j'})_{4\times3} = \begin{pmatrix} 0.6 & 0.09 & 0.32 \\ 0.73 & 0.07 & 0.20 \\ 0.42 & 0.21 & 0.37 \\ 0.06 & 0.02 & 0.91 \end{pmatrix}. \tag{2}$$

For the values in $P'$, $p'_{33} = 0.91$ means that the most of divided topics have 91% probability to be the divided topic; and $p'_{11} = 0.73$ means that the dormant topic has 73% probability to be the dormant. $p'_{01} = 0.6 > p'_{03} = 0.32 > p'_{02} = 0.09$ means that the new topic has 60% likelihood to change to the dormant topic, and 32% probability to change to the divided topics.




(a) The new topic. (b) The self-expanded topic. (c) The divided topic.

**Fig. 12.** The contribution of the new paper to different types of topics. (a) shows contributions of the three stages in the new topic. The novice and the potential decrease, while the reputation increase. (b) displays disorder. (c) has significant trends of the three stages, the reputation increase and the other two stages go to the opposite direction.

Comparing Figs. 3(a) and 3(b), there is a closed transition cycle including all four stages of careers in Fig. 3(a), but the new topic is excluded in the closed cycle formed by the other three types. That is, there is no steady state of $P'$. With this observation, we simulate the trends of the topic type. First, we take the average of all transition probabilities in 1999–2018 as the initial times, $\pi'_0 = (s'^0_1, s'^0_2, s'^0_3) = (0.57, 0.08, 0.36)$. Then, we multiply and iterate 19 times of the transition matrix (2), the result is $\pi'_{19} = (0.20, 0.03, 0.65)$. And if we continue the iterate process, the results go on the trend of decreasing which means that there is no steady state during the topic types evolution.

In Section 3, we focus on the statistical characteristics and states of the career stages in $\alpha$-layer and the types of topics in $\beta$-layer. An in-depth question is about relationship between the types of topics and authors' career stages. So in the following section we discuss the evolving relationships between the topics in $\beta$-layer and the collaborations in $\alpha$-layer.

## 4. Coupling mechanism of $\alpha$- and $\beta$-layers

In this section, we reveal the coupling mechanism between the two layers. It is found that both of the new scholars and the old ones prefer to the divided topics, especially the old scholars. For new topics and self-expanded topics, the proportion of new and old authors is almost the same, even under the continuous accumulation of old authors.

With development of the discipline, from the perspective of cumulative contribution or annual incremental contribution, reputation scholars' contribution to divided topic is particularly obvious.

### *4.1. Contributions of the new paper to the topics*

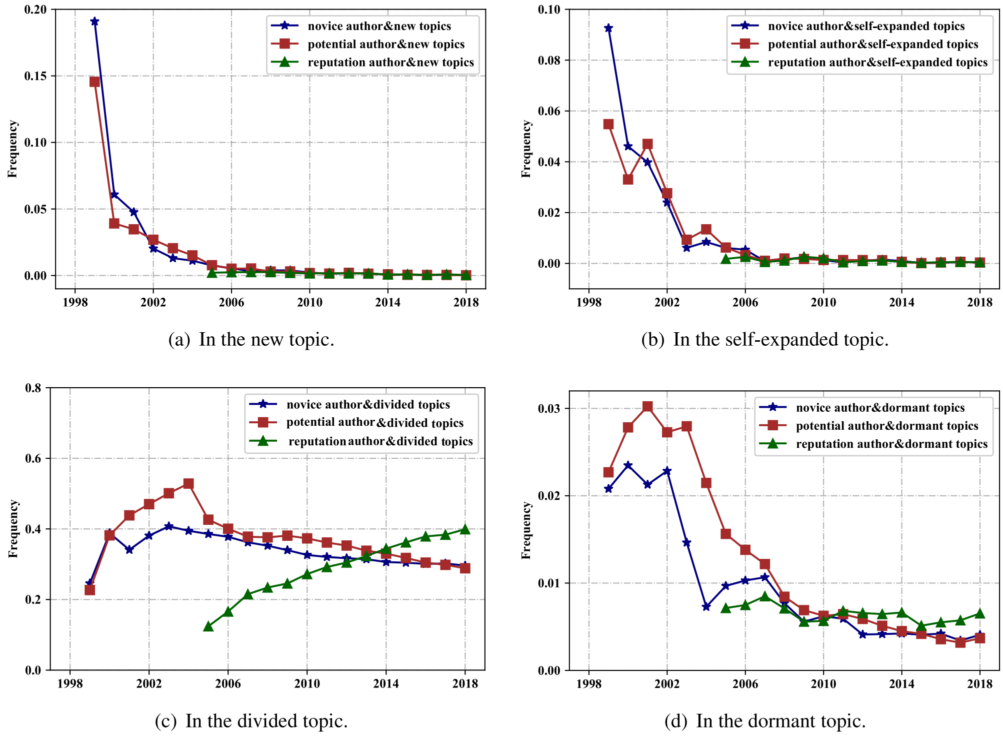
In order to observe the coupling relationship between the career stages and the type of topics more closely, we focus on the forming process of the type of the topics when a new paper added in $\beta$-layer. When a new scholar is added in $\alpha$-layer, this scholar has one stage, the novice. However, a new paper is published and added in $\beta$-layer, it could be one of three possible type or communities: the new, the self-expanded and the divided. Thus we detect which type of topics belongs to when a new paper is added in the $\beta$-layer. That is, we would reveal the coupling relationship between the topic types of new papers and the career stages of their authors. Exploring the contribution of different scholars' career stages to the topic types for new added papers, we could have a deeper understanding of the preference of different topic types matching with different scholars.

Calculating the scholars' stages and the topic' types in SMSEC data set, we get the annual coupling as shown in Fig. 12. It can be seen from Figs. 12(a) and 12(c) that the contributions of the reputation scholars to the new and the divided topics increase gradually, while the novice and the potential scholars to the new and the divided topics serrated decreasing. For the self-expanded topic, although the proportions of scholars with the three topic types have large fluctuation, it is generally equal, which can be seen from Fig. 12(b).

Based on the different proportions of scholars with respect to different types of topics, the proportion of the reputation scholars increases from 0.1 to 0.5 in the new topics, and it diverse from 0.1 to 0.4 in the self-expanded topics. However, Figs. 13(c) and 12(c) are very similar which coincides with the large number of divided topics, and it shows that the most important contribution is from the reputation scholars to the divided topics.

### *4.2. Coupling of the career stages and the topic types*

The change of scholars' cooperation relationship in $\alpha$-layer drives the change of topic types. Therefore, a natural approach is to find the coupling of career stages and topic types. According to the definition of reputation status in Section 2.2, the reputation scholars first appear in 2005 in SMSEC data. Hence, the data analysis of the reputation scholars starts at the year 2005. According to the timestamp coupling of two layers, we calculate the distributions of the career stages in the different topic types, and the results are shown in Fig. 13.




(a) In the new topic. (b) In the self-expanded topic.

(c) In the divided topic. (d) In the dormant topic.

**Fig. 13.** The proportions of career stages in four different types of topics. (a) and (b) show the proportions of three career stages devote to the new and the self-expanded topics. The two types have the same trends approaching to zeros after the year 2006. (c) shows clear trends of three stages in the divided topics. (d) also displays similar trends in the dormant topics, but those curves have relatively steady values.

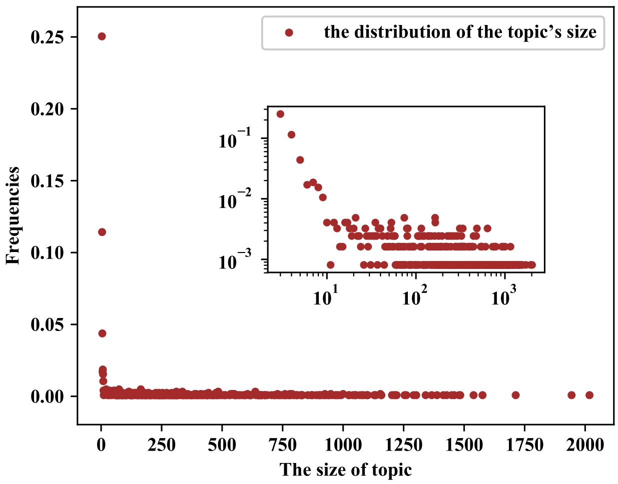



**Fig. 14.** The distribution of sizes of topics.

It can be seen from Figs. 13(a), 13(b) and 13(d), at whatever stages, scholars have almost the same preference for the new topic, the self-expanded topic and the dormant topic. The proportions of the scholars devoting to the new and the self-expanded topics are tending to zeros, while the dormant topics is about 0.5%. But, Fig. 13(c) shows much different picture than the other three panels. In the early development of any scholar's stages, their preference for divided topics increases, that is, the scholar pays more attention to cross topic fusion research in their early career stages. However, with the development of the scholar's career, the proportion of the novice scholars and the potential scholars devoting to the divided topics gradually decreases, and the reputation scholars devoting to the divided topics increases. Since 2014, the proportion of the reputation scholars for the divided topics has exceeded the novice and the potential scholars. In this view, the reputation scholars have the most contribution to the growth of the divided topics.

Based on the above analysis, we find that scholars in different stages have different extent contributions to the different types of topics. The contribution of the reputation scholars to the divided topics is particularly visible. In addition, the numbers of the new topics and the self-expanded topics are less than that of the divided topics, so the two types of topics are diluted in cumulative statistics. However, from the perspective of annual increment, the contribution of the reputation scholars to the new topics is still the most obvious in the long run, and the contributions of three career stages to the self-expanded topics are equal.

## 5. Discussion and conclusions

In this study, we propose a conceptional model of a two-layer network $G(t)$, one layer is the co-authorship network and the other is the paper similarity network, simply referred as $\alpha$-layer and $\beta$-layer, respectively. We define the three stages of scholars' careers by their research length and four type of topics in $\alpha$-layer and $\beta$-layer, respectively. Applying SMSEC data set to $G(t)$, the statistical analysis on data show the different between the two layers: higher cluster and longer distance in $\alpha$-layer than in $\beta$-layer, which suggest different forming mechanism, the "richer get richer" preference attachment and the "birds of a feather flock together" rule of similarity inherited in the two layers, respectively. The coupling mechanism of the two layers is calculated by the career stage changing and the topic type transition, in which the stages transition is a closed Markov process and has steady state, while the topic type changing has no steady state. It is found that both of the new scholars (the novice) and the experienced ones (the potential and the reputation scholars) prefer to the divided topics, especially the later. With the development of research, the contribution from the reputation scholars to the divided topic is extremely obvious whenever from the perspective of accumulation or annual increment. For the new topics and the self-expanded topics, the proportions of the new and experienced scholars have almost the same contribution.

Four statistical results are obtained: (1) the active old nodes play an important role in the development of disciplines and in terms of cooperation in $\alpha$-layer, new authors depend on the old ones. What is more, there are about 91% (i.e., $p_{10} + p_{20} + p_{30}$) scholars quitting from the research career; (2) the new papers are more related to old papers in $\beta$-layer. That is, new papers born from the existing topics. The scales of topic types are different, among which the largest one is the dormant topics, followed by the divided topics, the new and the self-expanded. Furthermore, the number of large-scale topics is relatively small, and the number of small-scale topics is large. However, there are about 20.99% of large-scale topics (i.e., the size is greater than 50) accounting for 71.57% of papers. With the development of the discipline, the topic types have about 65% probability to transit to divided topics; (3) the coupling mechanism of $\alpha$ and $\beta$-layer in the two-layer network $G(t)$ reveals the contributions of scholars in different stages to different type of topics; (4) the scholars contribute a lot to the development of the divided topics, especially the experienced ones. And their contributions to the new topics and the self-expanded topics are almost the same, and the proportion of these two coupling relations is less than that of the divided topics. The contribution of the reputation scholars to the divided topics is significant regardless with respect to cumulative contributions or annual increments.

In the data set of SMSEC, for the three states of scholars, the average numbers of papers published by these authors in the novice, potential and reputation are 1.21, 1.74 and 2.88 respectively. Since the scholars reach his/her potential stage after the novice state, and the reputation stage after the potential and the novice states, the number of published papers is cumulated and increase with the growth of scholars' research length. In addition, from Fig. 13, we see that the total contribution of reputation scholars accounting for 13.07% of the total number of scholars to various topics can reach 40%. However, the novice and the potential scholars, accounting for 60.87% and 26.05% of the total number of scholars, contribute 30% and 30% to various topics, respectively. It follows that scholars contribution to the number of topics increase with the growth of career length. To sum up, the growth of scholars' career is not only conducive to improve output of papers, but also conducive to the exploration of research topics. Therefore, scholars should spend more time on research topics. Long-term adherence not only advance in level of research, but also expand research fields and scope of cooperators. Reputation scholars should strengthen their research mentor-ship in novice scholars, generally majority of novice scholars are graduate students, and they are also potential scholars. The stages of novices and potentials are also career stages of growth process of reputation scholars. As a result, reputation scholars should help and coach novice and potential scholars going through challenge period of research establishment, enhancing their confidence, and commitment to scientific research. Moreover, the growth process of the number of reputations would also bring along development of the discipline and scientific research with applications.

Although we present a new perspective to understand the research topics evolution in this study, several limitations exist. First of all, SMSEC data set is obtained from the scholars of management science and engineering of China, and based on CNKI database of China that does not contain all the publications of scholars. Therefore, the results might deflect due to the data deficiency. Another defect is that this study merely focuses on the quantity of publications not considering the quality of the papers. The quality of papers and the topics might be more benefit to measure scholars' research stages.

We analyze the coupling relationship between scholars of different career stages and their topic types (states) based on statistical perspective in this paper. However, data analysis and mining of the coupling relationship are only discussed from the quantitative perspective. For further study, we will use the references of the research, such as the multiplex social network structuring processes [23], propagation paths enhance locating the source of diffusion [37], and the independent cascade [38] and so on, to pay more attention to the relationships between the two layers, in which the coupling mechanism between research topics and career stage switch is an attractive issue. We would like to discuss in much details for the behavior of scholars, processes of knowledge diffusion, and possible cascading effects in the two layer networks.

The plan for the future work includes to refine the discussion of scholars' choice behaviors on topics and to introduce knowledge diffusion to explain coupling between two layers. From Fig. 13, we find that scholars prefer to the divided topics. But, we did not carry out semantic analysis on it or specific preferring behaviors on choices of divided topics. For example, if a topic $topic_{parent}$ is divided into two topics $topic^1_{sub}$ and $topic^2_{sub}$ in the next timestamp after semantic analysis, we need to determine whether $topic_{parent}$ and two sub-topics $topic^1_{sub}$ and $topic^2_{sub}$ are similar or not in the future work. And then we would extract rules of topic division. With such an investigation, coupling mechanism behind topics would be revealed. This mechanism can be viewed as a matching rule for scholars of different academic careers and papers of different states. Therefore, coupling rules can be simulated by fitness in agents which embeds in intelligent computing. Using known data as training set can help us to predict contribution of scholars in a certain research field to a topic.

The cooperation among authors who study different topics would provide a chance to exchange or fusion different domain knowledge even the reputation scholars had published many papers and researched in many topics. The knowledge transmission would begin at their cooperation and be contained in published papers. By dividing topics of papers, the authors of the same topic will surely promote diffusion of domain knowledge through cooperation, because they have a greater chance of reading each other's published papers and establishing a partnership. Scholars have a great probability to transfer their domain knowledge to their collaborators when they cooperate with other scholars which may also promote topic change of their collaborators. We suspect that diffusion of domain knowledge in author cooperation layer would promote spread of intellectual knowledge, such as research methods, technologies and prospect of researches and so on. And knowledge diffusion also provides possible support for topics division. Based on the above discussion, we hope to further explore the specific evolution rule of divided topic through text semantic analysis, and then provide a richer explanation for this phenomenon from the perspective of knowledge diffusion.

## CRediT authorship contribution statement

**Yinghong Ma:** Conceptualization, Methodology, Writing - original draft. **Le Song:** Writing - original draft, Visualization. **Zhaoxun Ji:** Data collection and computing, Visualization. **Qian Wang:** Software editing. **Qinglin Yu:** Language polish.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

We show our great appreciation to all the anonymous reviewers who presented nice comments and advices, which greatly benefit to improve the quality and the readability of this paper. We also show our thanks for National Science Foundation of China (No. 71471106) and Teaching Reform Project of Shandong Province, China (Z2016Z025) supporting partly for this research.

## Appendix

Table 1 shows the following rules: (1) the sizes of the nodes $|V|$ and edges $|E|$ changes in $\alpha$-layer and $\beta$-layer from 1998 to 2018: in $\alpha$-layer, the number of authors $|V_\alpha|$ increases from 329 to 18969 people, and the number of edges $|E_\alpha|$ increases from 511 to 45509; in $\beta$-layer, the number of papers $|V_\beta|$ increases from 142 to 26606, and the number of similarity edges $|E_\beta|$ increases from 151 to 582017. Furthermore, The number of coupling edges between two layers $|E_{\alpha\beta}|$ increases with times. (2) the average degrees $\langle Deg\rangle_\alpha$ of $\alpha$-layer are relatively steady. (3) in $\beta$-layer, the average similarity weights of the papers $\langle W_{node}\rangle_\beta$ increases. (4) the size of average clusters of $\alpha$-layer $\langle C_i\rangle_\alpha$ is greater than that of $\beta$-layer $\langle C_i\rangle_\beta$. (5) the trends of the average distances of the two layers, $\langle Dist\rangle_\alpha$ and $\langle Dist\rangle_\beta$, are same.

## References

[1] S. Fortunato, C.T. Bergstrom, K. Borner, J.A. Evans, D. Helbing, S. Milojevio, A.M. Petersen, F. Radicchi, R. Sinatra, B. Uzzi, Science of science, Science 359 (6379) (2018) aao0185.
[2] A. Clauset, D.B. Larremore, R. Sinatra, Data-driven predictions in the science of science, Science 355 (6324) (2017) 477.
[3] J.S. Katz, B.R. Martin, What is research collaboration? Res. Policy 26 (1) (1997) 1–18.
[4] S. Lee, B. Bozeman, The impact of research collaboration on scientific productivity, Soc. Stud. Sci. 35 (5) (2005) 673–702.
[5] A. Gazni, F. Didegah, Investigating different types of research collaboration and citation impact: a case study of harvard university's publications, Scientometrics 87 (2) (2011) 251–265.
[6] R. Sooryamoorthy, Do types of collaboration change citation? Collaboration and citation patterns of South African science publications, Scientometrics 81 (1) (2009) 177.
[7] E.Y. Li, C.H. Liao, H.R. Yen, Co-authorship networks and research impact: A social capital perspective, Res. Policy 42 (9) (2013) 1515–1530.

[8] M. Belkhouja, H. Yoon, How does openness influence the impact of a scholar's research? An analysis of business scholars' citations over their careers, Res. Policy 47 (2018) 2037–2047.
[9] T. Amjad, A. Daud, M. Song, Measuring the impact of topic drift in scholarly networks, in: WWW '18: Companion Proceedings of the The Web Conference 2018, pp. 373–378.
[10] A. Rzhetsky, J.G. Foster, I.T. Foster, J.A. Evans, Choosing experiments to accelerate collective discovery, Proc. Natl. Acad. Sci. USA 112 (47) (2015) 14569–14574.
[11] J. Sienkiewicz, K. Soja1, H. Janusz, M.A.S. Peter, Categorical and geographical separation in science, Sci. Rep. 8 (8253) (2018) 1–12.
[12] H. Ankarali, O. Pasin, S. Ankarali, A new index for evaluating academic performance: Hos-index, Int. J. Comput. Sci. Issues 44 (3) (2017) 78–86.
[13] J. Tao, D. Wang, B.K. Szymanski, Quantifying patterns of research-interest evolution, Nat. Hum. Behav. 1 (4) (2017) 0078.
[14] A. Aleta, S. Meloni, N. Perra, Y. Moreno, Explore with caution: mapping the evolution of scientific interest in physics, EPJ Data Science 8 (2019) 27.
[15] R. Klavans, K.W. Boyack, Research portfolio analysis and topic prominence, J. Informetr. 11 (4) (2017) 1158–1174.
[16] H. Zuckerman, Theory choice and problem choice in science, Sociol. Inquiry 48 (3–4) (2010) 65–95.
[17] T.F. Gieryn, Problem retention and problem change in science, Sociol. Inquiry 48 (3–4) (2010) 96–115.
[18] A. Zeng, Z. Shen, J. Zhou, Y. Fan, Z. Di, Y. Wang, H. Eugene Stanley, S. Havlin, Increasing trend of scientists to switch between topics, Nature Commun. 10 (1) (2019) 34–39.
[19] P. Deville, D. Wang, R. Sinatra, C. Song, V.D. Blondel, A. Barabási, Career on the move: Geography, stratification, and scientific impact, Sci. Rep. 4 (7497) (2014) 4770.
[20] I. Hellsten, R. Lambiotte, A. Scharnhorst, M. Ausloos, Self-citations, co-authorships and keywords: A new approach to scientists' field mobility? Scientometrics 72 (3) (2007) 469–486.
[21] M. Ausloos, R. Lambiotte, A. Scharnhorst, I.R. Hellsten, Andrzej Pekalski networks of scientific interests with integral degrees of freedom through self-citation analysis, Internat. J. Modern Phys. C 19 (3) (2008) 371–384.
[22] J. Gläser, W. Glänzel, A. Scharnhorst, Same data–different results? Towards a comparativeapproach to the identification of thematic structures in science, Scientometrics 111 (3) (2017) 981–998.
[23] A. Curtis, G. Piotr, J. Matthew, H. Janusz, Why understanding multiplex social network structuring processes will help us better understand the evolution of human behavior, 2019, arXiv.org, 1903.11183.
[24] M.E. Newman, The structure of scientific collaboration networks, Proc. Natl. Acad. Sci. 98 (2) (2001) 404–409.
[25] J. Wang, Knowledge creation in collaboration networks: Effects of tie configuration, Res. Policy 45 (1) (2016) 68–80.
[26] Q. Gui, C. Liu, D. Du, Globalization of science and international scientific collaboration: A network perspective, Geoforum 105 (2019) 1–12.
[27] A. Petersen, M. Riccaboni, H. Stanley, F. Pammolli, Persistence and uncertainty in the academic career, Proc. Natl. Acad. Sci. 109 (14) (2012) 5213–5218.
[28] A.M. Petersen, O. Penner, Inequality and cumulative advantage in science careers: a case study of high-impact journals, EPJ Data Sci. 3 (1) (2014) 1–25.
[29] G.M. Salton, A. Wong, C.S.A. Yang, A vector space model for automatic indexing, Commun. ACM 18 (11) (1974) 613–620.
[30] D. Radhakrishnan, V. Jawahar Senthil Kumar, Evaluating the performance of similarity measures used in document clustering and information retrieval, Integr. Intell. Comput. 0 (2010) 27–31.
[31] R. Feldman, J. Sanger, The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data, Cambridge University Press, 2007.
[32] M.E. Newman, Fast algorithm for detecting community structure in networks, Phys. Rev. E 69 (6 Pt 2) (2004) 066133.
[33] V.D. Blondel, J.L. Guillaume, R. Lambiotte, E. Lefebvre, Fast unfolding of communities in large networks, J. Stat. Mech. 2008 (10) (2008) 155–168.
[34] A. Clauset, C.R. Shalizi, M.E.J. Newman, Power-law distributions in empirical data, SIAM Rev. 51 (4) (2009) 661–703.
[35] M. Hlynka, Probability and statistics by example. Volume 2: Markov chains: A primer in random processes and their applications by Yuri Suhov; Mark Kelbert, Technometrics 52 (2) (2010).
[36] E. Bruckner, W. Ebeling, A. Scharnhorst, Stochastic dynamics of instabilities in evolutionary systems, Syst. Dyn. Rev. 5 (2) (1989) 176–191.
[37] G. Gajewski, K. Suchecki, J.A. Holyst, Multiple propagation paths enhance locating the source of diffusion in complex networks, Physica A 519 (2019) 34–41.
[38] J. Choloniewski, J. Sienkiewicz, G. Leban, J.A. Holyst, Modeling of temporal fluctuation scaling in online news network with independent cascade model, Physica A 523 (2019) 129–144.