MSM: A Multi-entity Scholarly Model for Systematic Understanding of Evolving Scholarly Networks

Lingkun Kong, Fengyu Deng, Jiaqi Liu, Jinghao Zhao, Luoyi Fu, Xinbing Wang
Shanghai Jiao Tong University, China

ABSTRACT

Scholarly networks contain massive scholarly information that can be mainly categorized into three entities, i.e., *paper*, *author* and *topic*, which exhibit a co-evolution over time. Although scholarly networks have attracted much attention over the past years, most works focus on single entity of the network, e.g. sub-networks generated by citation, co-authorship or topic relationship [1–5]; while few of them incorporate different entities into an entirety to provide a systematic understanding of scholarly networks at scale. We bridge this gap by proposing a multi-entity scholarly model (MSM) with strong theoretical guarantees, which amalgamates entities of paper, author and topic into one single framework to simulate interactions among different entities, and thus presenting the co-evolution within scholarly networks.

First of all, using real scholarly datasets — *Microsoft Academic Graph*[1] [6] with 126 million publications, we observe properties that belong exclusively to scholarly networks, such as varying and converging exponents of power-law distributions with time, degree densification in each of the three aforementioned entities. We also observe interesting evolving patterns like simultaneous co-evolution of all the three entities, faster growth rate of entities with larger size, etc. Based on our observations, we propose the MSM that jointly captures both intra and inter correlations among different entities during the evolving process. Through both theoretical analysis and empirical simulation, we further characterize the MCM as capable of reproducing evolving patterns of real multi-entity scholarly networks.

KEYWORDS

Data mining; Scholarly networks; Network Dynamics.

1 INTRODUCTION

Recent years have witnessed the rapidly growing scholarly information due to vast research works are undertaken in academia and industry [7]. Large collections of scholarly data such as papers, authors, citations, topics, etc., emerge by researchers' continuous working. All the information, when combined together, leads to the formation of the scholarly network that contains three major entities, i.e., *paper*, *author* and *topic*, [8] as will also be the concern of this paper. As can be seen from the Figure 1, the title of the publication, which we regard as paper, the authors of the publication, which we view as author, and the listed keywords which we consider as topic are strongly connected to each other in the sense that authors on the same paper exhibit collaborative relations (shown by arrows between As, which mean authors), with that paper (presented by arrows between A and P) further cited by other literatures (represented by arrows among Ps, which mean papers) belonging to some specific topics (denoted by arrows between Ks, which mean keywords, and meanwhile topics, and A). All the three entities interact with each other in the prescribed way as time goes by, which results in an evolving multi-entity scholarly network, in terms of both the entity size and its more complicated connection of inter & intra entities.

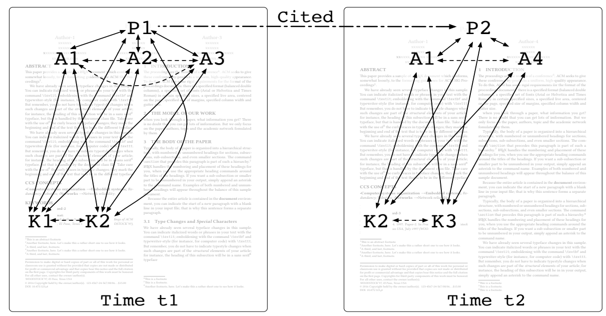


Figure 1: The heterogeneous scholarly network extracted from real papers.

As a matter of fact, studying properties of scholarly networks and getting insight of their evolving mechanism have important implications, as reported by a literature survey [9]. For example, by analyzing the citation relationships extracted from scholarly networks, we can evaluate the impact of a given paper or author, which can help to allocate reputations to scientists [10]. Similarly, by analyzing the co-author behaviors among scholars, we can find the distribution of scientific communities, which helps to build the map of sciences. Meanwhile, scholarly network analysis is not only important for academia but also helps in planning and development, i.e., for sociologists to understand researcher interactions [11], for policy makers to address knowledge and resources sharing, and for scientists, businesses, and the general public as a reference. Despite the importance of scholarly networks in many kinds of applications, most works, even already having realized the interaction between different entities in scholarly network, revolve around one entity of the scholarly network, e.g., sub-networks generated by citation, co-authorship or topic relationship[1–5]; while few of them merge multiple kinds of entities into one single fabric to obtain the understanding of scholarly networks from an overall perspective. In fact, three challenges are major hindrances.

First of all, due to the difficulty in acquisition of real scholarly data that contain complete information of all entities in scholarly networks, i.e., paper, author and topic, there is a lacking of studies that empirically explore co-properties among multiple entities. When considering corporate evolving patterns within all different entities, the paucity can be even more significant. Secondly, modeling the concurrent evolution among multiple entities in scholarly networks is far complicated than modeling certain extracted parts from the whole network (e.g., citation networks, co-author networks or topic networks) [1–4, 12]. Though the simplification

[1] http://acemap.sjtu.edu.cn/acenap/index.php/datasets.html

facilitates model design and analysis, the lacking of the integrality of scholarly networks might lead to information loss. For instance, the pattern of interactions between different scholarly entities is neglected as well as the co-evolution. Thirdly, due to inappropriate modeling methods, a large amount of previous studies get stuck in giving detailed mathematical proofs to support their model. In contrast, they employ simulating experiments or baseline evaluating methods to validate their models' brilliance, which in our eyes, is worthy of reference while not compelling enough.

Faced with these challenges, we are motivated to give the first model which incorporates entities of paper, author and topic into one single framework. To begin with, employing real-world datasets which have around 126 million papers, we comprehensively observe evolving patterns among different entities inside scholarly networks. Then, based on observations, we propose our multi-entity scholarly model (MSM) that is inspired from the structure of tripartite graph to roundly incorporate all entities and their correlation into an entirety. Last but not least, we mathematically characterize MSM as capable to reproduce observed patterns, along with further verification through simulations. Here, we illustrate our work and summarize our contributions by three aspects:

Observation: Our first contribution is to originally explore comprehensive properties in scholarly networks with the concern of multiple entities, i.e., paper, author and topic. Based on scholarly datasets provided by Microsoft, which contain about 126 million papers, we use data-mining and other big-data analyzing approaches to observe patterns in the growth of the scholarly network. On one hand, we observe some similar features to those that have already been discovered in many traditional social networks, such as power-law degree distribution, densification, etc. On the other hand, there also exists several unique features in scholarly networks, like faster growth rate of the entity that has a bigger size, varying and converging exponents in power-law distributions with time, and the simultaneous co-evolution of all entities, etc. All these evolving features, with the awareness of multiple entities, can be categorized into three types, i.e., inter-evolution, intra-evolution as well as the co-evolution on the whole. While deferring to Section 3 for more details, we remark that there is no prior work, other than ours, that have studied these patterns within multi-entity scholarly networks.

Modeling: Given empirical observations, our next significant contribution is for the first time establishing a comprehensive modeling of evolving scholarly networks. Combining entities of paper, author and topic in singe fabric, the proposed model captures both the inter-correlation and intra-correlation of the three entities during the evolving process. Particularly, inter-correlation is characterized through tripartite graph, whose evolving process follows the mode of preferential attachment prevalent in growth of real scholarly networks; Meanwhile, intra-correlation of nodes within each entity is described as intra-degree (which we define) power-law distribution, degree densification, etc.

Analysis: Our third contribution is to offer detailed mathematical proof to consolidate the reasonability of our model. Based on the constructing methods of random arrival, preferential attachment, edge copying and the assumption of the affiliation relationship inside entities, we successfully obtain the growing rate of nodes' degree, power-law distributions inside or among multiple entities and the densification of the entire network. Further, we also implement simulating approaches to validate that our model can accurately reproduce real scholarly networks.

The paper is organized as follows. In Section 2, we discuss relevant literatures. In Section 3, We list properties in multi-entity scholarly networks. We present our design of MSM in Section 4 and characterize MSM mathematically in Section 5. Section 6 is our simulation and we conclude in Section 7.

2 RELATED WORK

Evolving networks have long been a significant research topic [13–15]. There are a lot of properties discovered from real-world network graphs, such as power-law degree distribution, small-world phenomenon, degree densification and etc., which are detected by a wide spectrum of evolving network models [15–19].

However, most of studies mainly focus on social networks, purposing models to reproduce observed features in social networks, including random graph model built by Chakrabarti and Faloutsos [20], preferential attachment model proposed by Barabasi et al. [21], edge-copy model created by R. Kumar et al. [22], and affiliation network model advanced by Silvio Lattanzi et al. [23], while ignoring general understanding of the scholarly network, which actually is also important in people's daily life.

Besides, some efforts have also been made to launch study on scholarly networks for diverse usage. For instance, Yang et al. [1] use a joint topic model to solve scholar ranking and predicting problems. Li et al. [3] propose a random walk model based on co-author networks for recommending new collaborations. Lin et al. [4] and Wang et al. [24] study topic evolution of scholarly networks. Sun et al. [5] explore the co-author relationship in scholarly networks based on networks' heterogeneous structure. However, due to theoretical and technical difficulties, even researchers have realized the existence of multiple entities in scholarly networks [1, 5, 24, 25], few of them generates modeling to study and simulate the patterns embedded in the interaction and co-evolution among all scholarly networks' entities.

Therefore, we propose a novel scholarly model that employs tripartite graph to depict multi-entity scholarly networks' inter-correlation while uses affiliation networks' structure to portray intra-correlation, which can well reproduce properties we observed in real database. Besides, by using preferential attachment and edge copying approaches [22, 26], we construct our sophisticated model with concise mathematical characterization.

3 OBSERVATION

In this section, we launch our observation based on *Microsoft Academic Graph (MAG)* which is an official and authoritative scholarly dataset containing massive scholarly information of publications such as titles, authors, conferences, fields of study and citations. Around 126 million papers from 53,834 subjects are included in this database, and the published years of them vary from 1800 to 2016. To prove our observation persuasive in scholarly networks, we browse in different scholarly fields. With deliberate selection, we extract four representative fields from MAG which contains about 4.7 million papers published from 1940 to 2016 to show the properties of evolving multi-entity scholarly networks. These fields are: Data Mining, Networks, Literature and Finance. The detailed information is listed in Table 1.

In each field, the generation time of a paper is defined by its published date. According to this data, the generation time of an author is defined by his first paper's publication date. Similarly, the generation time of a topic is determined. Based on these four datasets, we mainly study on two kinds of features:

Table 1: Statistics of Scholarly Datasets

Dataset	# of Papers	# of Authors	# of Topics
Data Mining	1042279	1703828	403
Networks	1093537	1391869	774
Literature	679350	939544	446
Finance	1949028	2716094	968

Structural Property: Structural properties, in our work, are different kinds of degree distributions, which contain both intra-degree in single entity and inter-degree between different entities.

Evolving Property: Evolving properties denote evolving patterns among entities, which we divide into three kinds. The first, named as *intra evolving properties*, studies intra relationships in a single scholarly entity, e.g., the evolving co-authorship in the entity of author. The second part, titled as *inter evolving properties*, includes inter correlation between entities such as evolution between the topic and paper. The third, which we call as *co-evolving properties*, presents patterns of the co-evolution among paper, author and topic.

3.1 Structural properties

Power-law distributed degree is a common feature of classic social networks, which is well studied by many existing literatures [17, 27, 28]. By observations on MAG, we ensure the existence of power-law degree in scholarly networks, and explore this feature by two categories: intra-degree and inter-degree.

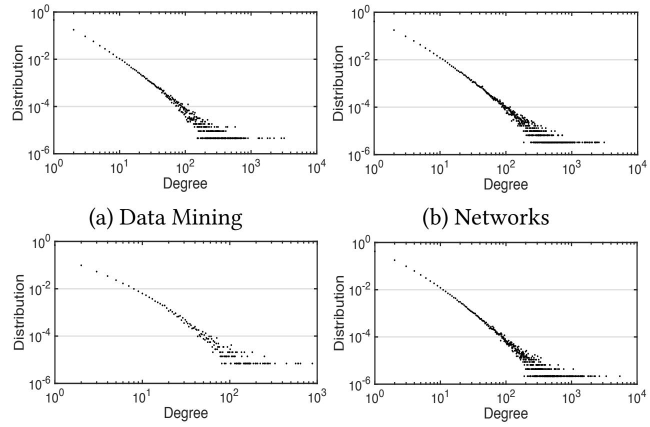
Intra-degree: Intra-degree signifies the element's degree in one single entity. To illustrate, in the entity of paper, publications are linked by citation, where the intra-degree represents the number of citations of one publication. In the entity of author, scholars are connected by co-authorship, where the intra-degree indicates the scholars' related researchers' count. Similarly, in the entity of topic, scientific communities are tied by topic networks, and the degree in this case indicates the number of one community's neighbors. We here simply take the intra-degree in the citation network, i.e., the entity of paper as an example to present our observations.

Recall that a random variable $x \in \mathbb{Z}^+$ follows a power-law distribution if:

$$\mathbb{P}\{x = k\} = \eta k^{-\varphi},$$
$$log(\mathbb{P}\{x = k\}) = -\varphi log(k) + log(\eta),$$

where φ is an exponent factor of the power-law distribution and η is a constant factor. $\mathbb{P}\{x = k\}$ is referred as the probability when $x = k (k \in \mathbb{Z}^+)$. From above equations, we know that the larger φ is, the greater probability x has to stay at a small value.

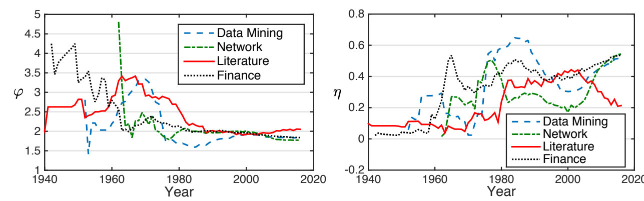
Figure 2 shows the degree distribution of citation networks in four datasets. Obviously, they all follow power-law degree distribution. We also study on the evolving process of factor φ and η. The result is illustrated in Figure 3. As time grows, in three datasets except Literature, φ keeps the value about 1.77. The φ of Literature is larger, which means that fewer papers have higher citations in this field.



(a) Data Mining (b) Networks

(c) Literature (d) Finance

Figure 2: Degree distribution of citation network.



(a) Evolution of φ (b) Evolution of η

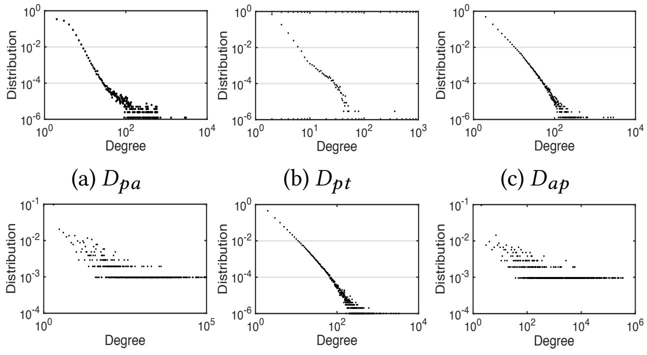
Figure 3: Evolution of power-law distribution factors

Inter-degree: The inter-degree represents the degree generated by connections between different entities, such as scientific community's degree in topic-paper sub-network which measures the number of published papers in this community, and scholar's degree in author-paper sub-network which scales how many theses the scholar has published. In our scholarly network, these sub-networks include elements in two entities as well as the links between them. Therefore, we have three sub-networks between every two entities, and six inter-degree distributions. For convenience, we denote inter-degree distributions of paper, author in paper-author sub-network and author-paper sub-network as D_{pa}, D_{ap}. By same methods, we define D_{ap}, D_{at}, D_{tp} and D_{ta}.

Figure 4 gives distributions of these degrees in Finance, which all follow power-law as expected. However, because three entities (paper, author, topic) differ by orders of magnitude in scholarly networks, i.e., a publication may only be related to a limited number of scientific communities while a community can link to thousands of publications. Thus, distribution factors are different, which are listed in Table 2. Compared with others, the φ of scientific communities' inter-degree is smaller, which means this community is more likely to include a large number of publications and scholars.

3.2 Intra evolving properties

In this subsection, we take the co-author network in the entity of author as an example to reveal intra evolving properties in scholarly networks. Looking into the evolution process of co-author network,



(a) D_{pa} (b) D_{pt} (c) D_{ap}

(d) D_{at} (e) D_{tp} (f) D_{ta}

Figure 4: Degree distribution between different entities

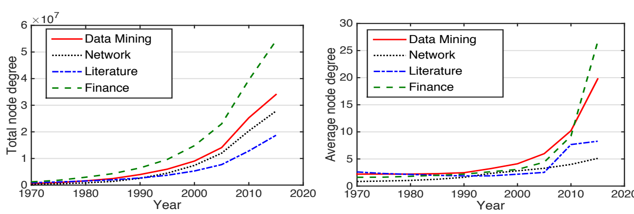
Table 2: Power-law distribution factors

Factors	D_{pa}	D_{pt}	D_{ap}	D_{at}	D_{tp}	D_{ta}
φ	1.67	3.11	2.38	2.32	0.18	0.13
η	1.70	0.34	0.52	0.32	2.32	2.48

we focus on two kinds of degree growths to show the densification [15] phenomenon.

To begin with, we observe the increasing mechanism of nodes' total degree in certain scholarly entity. As shown in Figure 5(a), the nodes', i.e., scholars' total degree is very small in the past century, while it grows rapidly in the 21st century. We reckon that in the new century, for the benefit of rapid development in science and technology, scholars have more opportunities to cooperate with others. Besides, it also can be found in Figure 5(b) that the nodes' average degree increases exponentially.

As a whole, both kinds of nodes' degree, i.e., intra-degrees comply with an exponential increasing pattern, which our model is proved to be able to reproduce; and detailed theoretical proof is given in Section 5.



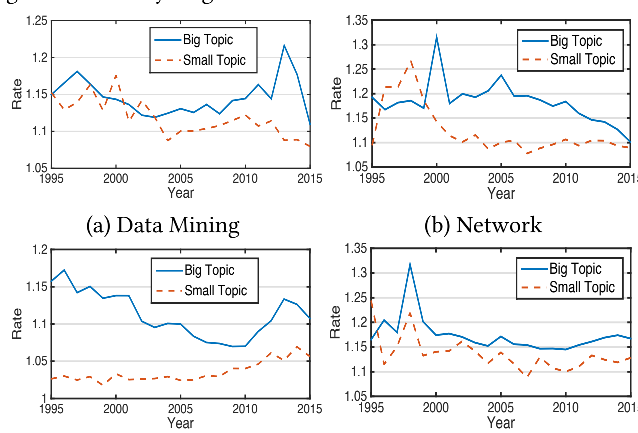
(a) Total node degree (b) Average node degree

Figure 5: Evolution of scholar's total degree and average degree in co-author network.

3.3 Inter evolving properties

Paper, author, and topic, in fact, are never separated from each other. Before we focus on the mechanism of the whole scholarly network, in this subsection, we study the relationship evolving over time between two entities. We refer this as the inter evolving property, and use the evolution of the size of scientific community in topic entity as an example to describe their connections.

According to previous definition, the size of a scientific community can be measured by how many publications it contains, or how many publications it links to in our multi-entity scholarly network. In datasets, we notice that some big communities can link to more than 100 thousand publications while some small communities only contain one hundred theses. Since they all grow up from a community with small amount of publications, there must exist difference in the process of their growth. What we observe is that communities with larger size grow with faster rate. In our observation, we divide scientific communities into two groups according to their sizes. Then we calculate the average growth rate of community size and plot the results in Figure 6. The reasonability of the observation holds because when a community gets bigger, it will be more likely to draw attention from researchers, who, as a result, might contribute new publications in this community. Consequently, the growth rate stays high.



(a) Data Mining (b) Network

(c) Literature (d) Finance

Figure 6: Evolution of topic's, i.e., scientific community's size, with the upper and lower threshold as 5,000 and 500

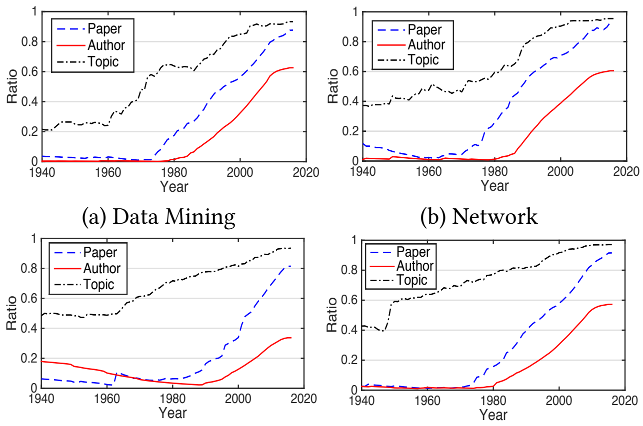
Moreover, the growth rate changes over time. In dataset of Networks, the size growth rate of big community is decreasing while small community's remains stable. This indicates that in the next few years, the size of the big community may increase slowly, which, to some extent, reveals the community with copious publications might not still be attractive in the foreseeable future.

3.4 Co-evolving properties

Apart from intra or inter evolving properties, what we are most curious about is the co-evolution of the scholarly network including all entities. We are eager to find a novel pattern, which is observed within all entities during the evolving process, to present the co-evolving structure of multi-entity scholarly networks.

For a promising field, there is no doubt that more scholars would like to set foot in this field and launch research in relevant scientific communities, which leads to more published papers. Consequently, it will elicit growing citations, stronger co-authorship and more diverse communities in this field. As a result, the connectivity of scholarly networks will become denser. Based on this intuition, we find the pattern about *Giant Component* from the perspective of connectivity in our observations.

Giant Component [29] is the largest connected component of a given graph. If a giant component contains C_g nodes and the total nodes in the graph is C_G, then the connectivity of this graph can be measured by the ratio of C_g and C_G: $ratio = C_g/C_G$. We calculate this ratio in each entity at every time slot, and present the result in Figure 7. According to this result, the ratio in each entity of the scholarly network has strong connection with other entities and they grow with the same pattern simultaneously. It indicates a concurrent evolution among all entities in the scholarly network.



(a) Data Mining (b) Network

(c) Literature (d) Finance

Figure 7: Evolution of ratio (C_g/C_G) in three sub-graphs of the scholarly network

4 MODELING OF MULTI-ENTITY SCHOLARLY NETWORKS

Based on above observations, we design a novel model to capture these properties and we name it as: Multi-entity Scholarly Model (MSM). In this section, we first introduce MSM and then describe the evolving process of it.

4.1 Multi-entity scholarly model

In MSM, the scholarly network, depicted as a heterogeneous graph in which entities are theoretically characterized as node sets, is denoted as $G(N_p, N_a, N_t)$. In each node set, nodes represent elements in this entity, e.g., authors in the entity of author. We use tripartite graph to present the inter-correlation between entities. Also, we focus on the intra-features in each entity. The framework of MSM is illustrated by Figure 8. It contains:

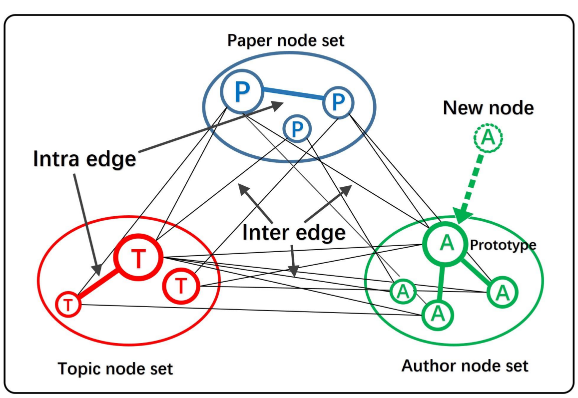


Figure 8: Structure of Multi-entity scholarly model

(1) Three node sets: Paper node set N_p, author node set N_a and topic node set N_t. The node in each node set is marked as n_p, n_a and n_t.

(2) Three inter-edge sets: We denote the edges between every two different node sets as inter-edge sets, and the graph has three inter-edge sets. For example, we refer all edges between paper node and author node as $E_{pa}(E_{ap})$, then an edge $e_{n_p n_a}$ which belongs to E_{pa} means author n_a writes the paper n_p.

(3) Three intra-edge sets: Intra-edge is the edge in the same node set, and our graph has three intra-edge sets, which we refer as E_{pp}, E_{aa}, and E_{tt}. If an edge $e_{n_a^i n_a^j} \in E_{aa}$, then we know that author n_a^i and n_a^j have cooperation.

In Figure 8, nodes are illustrated as colored circles in each node set, and intra edge and inter edge are labeled. A new author node is trying to preferentially attach himself with some heavily linked authors nodes (distinguished by their sizes) that have already been in the author set. With these nodes and edges of the model, we can well extract the structure of scholarly networks. We present notations in Table 3 for later convenience and describe the evolving process of the proposed model in the following subsection.

Table 3: Notations and Definations

Notations	Definations
N_p, N_a, N_t	Node set of paper, author and topic
E_{ij}	Inter-edge set between nodes in N_i and N_j
E_{ii}	Inner-edge set of N_i
α_i	Probability that a new node arrives in N_i
β_{ij}	Probability that an edge added in set E_{jj}
c_{ij}	Number of edges added to E_{ij} at one time slot
$G(N_p, N_a, N_t)$	Graph of our evolving scholarly model
$B(N_i, N_j)$	Bipartite graph with sets N_i, N_j, and E_{ij}

4.2 Evolving process

While we defer the detailed evolving process of the MSM to Algorithm 1[2], we would also like to provide a corresponding brief summary of the process. We first fix parameters including α_i, β_{ij} and c_{ij} where $i \neq j \in \{p, a, t\}$, and then assume that the evolution starts from an initial case that can be modeled as an initial graph, showing that each node in the graph is linked to a number of nodes in other node sets. After initialization, for every time slot, we classify the process into five main steps: **1)** A new node, which can be randomly designated as a paper, an author, or a topic, is added to the graph. For clarity, here we take the arrival of a new author as the example for explanation of the subsequent steps. The symmetry also holds for paper and topic. **2)** With probability proportional to degree in $B(N_a, N_p)$ and $B(N_a, N_t)$, two author nodes n_a^p and n_a^t are chosen as prototypes for the new node n_a. **3)** c_{ap} neighbors (n_p^1 ,..., $n_p^{c_{ap}}$) of n_a^p in N_p and c_{at} neighbors (n_t^1 ,..., $n_t^{c_{at}}$) of n_a^t in N_t are randomly chosen to have connections with node n_a. **4)** $c_{ap}c_{at}$ edges are added between n_p^1 ,..., $n_p^{c_{ap}}$ and n_t^1 ,..., $n_t^{c_{at}}$. **5)** Edges between every two author nodes are added with probability β_{ap} (β_{at}) if they have a common paper (topic).

For a better understanding of this evolving process, let us consider the arrival of a new author as an instance. The new author is

[2] http://acemap.sjtu.edu.cn/acenap/index.php/algorithms.html

Algorithm 1: Evolving Process

Simulated time steps: T
Fix probability α_i that a new node arrives in N_i.
Fix parameters $\beta_{ij} \in (0, 1)$ and integers $c_{ij} > 0$ where $i \neq j \in \{p, a, t\}$.
Initialization: In initial graph, the node in each set has neighbors with other two node sets. For example, a paper node n_p connects to at least c_{pa} author nodes. Meanwhile, an author node n_a has at least c_{ap} paper neighbors. So the inter-edge set E_{pa} has at least $c_{ap}c_{pa}$ edges in the beginning.
for $1 \leq t \leq T$ **do**
1) ***Node arrival:*** According to α_p, α_a, α_t, we decide the type of node to join the graph. In later discussion, we take the arrival of a new author node n_a as example, and the symmetry also holds for paper and topic.
2) ***Preferentially chosen prototype:*** A node $n_a^p \in N_a$ is chosen as prototype for the new node, with probability proportional to its degree in $B(N_a, N_p)$.In the same way, another node $n_a^t \in N_a$ is chosen as prototype according to its degree in $B(N_a, N_t)$.
3) ***Edge copying:*** c_{ap} edges are copied from n_a^p, that is, c_{ap} neighbors of n_a^p, denoted by n_p^1 ,..., $n_p^{c_{ap}}$ in N_p are chosen uniformly at random, and the edges (n_a^p, n_p^1), ..., $(n_a^p, n_p^{c_{ap}})$ are added to the graph. Follow the same method, c_{at} edges (n_a^t, n_t^1), ..., $(n_a^t, n_t^{c_{at}})$ are added to the graph.
4) ***Indirect evolution:*** $c_{ap}c_{at}$ edges between nodes p_1 ,..., $p_{c_{pa}} \in N_p$ and t_1 ,..., $t_{c_{pt}} \in N_t$ are added to the graph $B(N_p, N_t)$.
5) ***Evolution inside:*** For every two nodes n_a^x and n_a^y $(x \neq y)$, if they have a common author (topic), then with probability $\beta_{ap}(\beta_{at})$, a edge (n_a^x, n_a^y) is added in E_{aa}.
end

likely to learn from an influential author, who is thus selected as a prototype and influences the new author on choosing research topics. Similarly, when conducing the new research, this new author probably looks up to some pioneers who have already made many publications as a prototype, from whom he chooses some papers for references. The symmetry also holds for the arrival of a new paper and a new topic, as they also select corresponding prototypes for guides.

4.3 Potential extra application of MSM

In this paper, MSM is proposed to characterize evolving patterns of the multi-entity scholarly network, and thereby helping to provide a comprehensive understanding for scholarly networks. However, though MSM is designed for evolving scholarly networks, it has the potential for evolving pattern prediction in other homogeneous social networks.

For instance, in an evolving social network that consists of users, groups and interests, entities can be extracted as user, group and interest. With mapping the entity of paper in MSM to user, author to group and topic to interest, a duplication of MSM can be generated and leveraged to predict both intra and inter evolving patterns of this social network. Some additional examples include market networks that comprise products, customers and commercial agents, contact networks that involve different roles of users, etc.

5 THEORETICAL ANALYSIS

In this section, we provide theoretical guarantees for MSM and confirm that our model can well reproduce properties in the real-world multi-entity scholarly network.

5.1 Growth of node degree

According to our model, we divide the nodes' degree into two types – the first is the *inter-degree*, i.e., the node degree between node sets, related with the growth of E_{ij}, we call it d^{ir}, which represents the inter-correlation in our model, and the second is the *intra-degree*, i.e. the node degree inside node set, related with the growth of E_{ii}, we call it d^{ia}, which reflects the intra-correlation of scholar networks.

Growth of inter-degree: Assuming node n arrives at node set N_i at time t_0 with initial inter-degree $d_i^{ir}(t_0)$, the inter-degree of n at time $t > t_0$ is

$$d_i^{ir}(t) = \left(\frac{t}{t_0}\right)^{\lambda_i} d_i^{ir}(t_0),$$

where $\lambda_i \in (0, 1)$ is a constant.

In fact, two implications can be deduced by this result.

(1) The inter-degree $d_i^{ir}(t)$ grows with polynomial rate in time t, following the power $\lambda_i \in (0, 1)$.
(2) The two components of vector $d_i^{ir}(t)$ are in the same order. For instance, $d_{pa}^{ir}(t) = \Theta\left(d_{pt}^{ir}(t)\right)$.

The first implication gives the growth rate of node's inter-degree; while the second one implies the similarity of a certain node's degree over two different node sets, which indicates that an influential node (node with large inter-degree) also plays an important role in all the other bipartite relationship network and vice versa. This complies with properties of our scholarly networks. For example, an author who studied in multiple fields is more likely published more papers and vice versa. The detailed proof is given in Theorem 5.1.

Growth of intra-degree: Again, we set beginning time as t_0 and the intra-degree of node set N_i at time $t > t_0$ is $d_i^{ia}(t)$, then

$$d_i^{ia}(t) = \Theta\left(\max\left\{t^{\frac{1}{\lambda_j}+1}, t^{\frac{1}{\lambda_k}+1}\right\}\right),$$

where λ_j, λ_k represent the constant λ in N_j and N_k which are neighbors of N_i, and the max is the maximum of two formulas.

The equation reveals that, in our model, the intra-degree of a node set actually is related with the inter-degree's growing rate variable λ. As in equation the intra-degree is positively related with the growing with time slot t, we can say the intra-degree also grows with time. The detailed proof is given in Theorem 5.2.

Combining these two results together, it can be easily viewed in our scholarly network graph $G(N_p, N_a, N_t)$ that the degree of the node in graph grows with polynomial rate in time t, and the growth rate differs from inter-degree to intra-degree of the node.

THEOREM 5.1. *For graph $G(N_p, N_a, N_t)$ generated after t time slots ($t \geq t_0$), with the initial condition that a certain node $n \in N_p$ is added to node set N_p at time t_0 with the degree $d^{ir}(t_0)$ from N_p to N_a and N_t, the inter-degree of n at time t satisfies*

$$d_p^{ir}(t) = \left(\frac{t}{t_0}\right)^{\lambda_p} d_p^{ir}(t_0).$$

This result also holds for $n \in N_a$ and $n \in N_t$ with symmetrical expressions.

PROOF. At each time slot t, the inter-degree of node $n \in N_p$ in $B(N_p, N_a)$, i.e. $d_{pa}^{ir}(t)$, may increase in two cases:

(1) A new node arrives at N_a and is connected to n, which results in $d_{pa}^{ir}(t) = d_{pa}^{ir}(t-1) + 1$.
(2) A new node arrives at N_t and is connected to n, then n will connect to c_{ta} neighbors of the new node in N_a which results in $d_{pa}^{ir}(t) = d_{pa}^{ir}(t-1) + c_{ta}$.

In edge copying, we choose the prototype node according to its inter-degree, while the endpoint of any edge is chosen with equal probability. Thus, the probability that a new added edge in $B(N_p, N_a)$ points to a certain node n is $\frac{d_{pa}^{ir}(t-1)}{s_{pa}(t-1)}$, where $s_{pa}(t-1)$ denotes the sum number of edges in $B(N_p, N_a)$ at time $t-1$, and we have

$$s_{pa}(t-1) = (\alpha_p c_{pa} + \alpha_a c_{ap} + \alpha_t c_{tp} c_{ta})(t-1).$$

And $s_{pt}(t-1)$ as well as $s_{at}(t-1)$ can be obtained by same method.

Combining the above two cases, we get

$$d_{pa}^{ir}(t) - d_{pa}^{ir}(t-1) = \alpha_a c_{ap} \frac{d_{pa}^{ir}(t-1)}{s_{pa}(t-1)} + \alpha_t c_{tp} c_{ta} \frac{d_{pt}^{ir}(t-1)}{s_{pt}(t-1)},$$

and similarly,

$$d_{pt}^{ir}(t) - d_{pt}^{ir}(t-1) = \alpha_t c_{ta} \frac{d_{pt}^{ir}(t-1)}{s_{pt}(t-1)} + \alpha_a c_{ap} c_{at} \frac{d_{pa}^{ir}(t-1)}{s_{pa}(t-1)}.$$

With the initial condition that

$$d_p^{ir}(t) = \left[\left(\frac{t}{t_0}\right)^{\lambda_p} d_{pa}^{ir}(t_0), \left(\frac{t}{t_0}\right)^{\lambda_p} d_{pt}^{ir}(t_0)\right], \tag{1}$$

where

$$\lambda_p = \frac{\sqrt{\Delta} + \alpha_a c_{ap} s_{pt} + \alpha_t c_{tp} s_{pa}}{2 s_{pa} s_{pt}}, \tag{2}$$

here, $s_{pt} = \frac{s_{pt}(t)}{t}$ is a constant, and according to the calculation result,

$$\Delta = (\alpha_t c_{tp} s_{pa} - \alpha_a c_{ap} s_{pt})^2 + 4\alpha_a \alpha_t c_{ap} c_{tp} c_{at} s_{pa} s_{pt}.$$

By same approach we can obtain the expression result of $d_p^{ir}(t)$ for nodes in N_a and N_t, thus we complete the proof. □

In fact, the proof of Theorem 5.1 also reflects that in N_p, $d_{pa}^{ir}(t)$ and $d_{pt}^{ir}(t)$ have the same order, as Equation (1) shows $d_{pa}^{ir}(t)$ and $d_{pt}^{ir}(t)$ have same growing function, i.e. $C(\frac{t}{t_0})^{\lambda_p}$, where C is a constant. Symmetrically, this property also holds in N_a and N_t.

THEOREM 5.2. *For graph $G(N_p, N_a, N_t)$ generated after t time slots ($t \geq t_0$), with the condition that inter-degree in node set N_a and N_t growing with the power λ_a and λ_t, the intra-degree of $n \in N_p$ at time t satisfies*

$$d_p^{ia}(t) = \Theta\left(\max\left\{t^{\frac{1}{\lambda_a}+1}, t^{\frac{1}{\lambda_t}+1}\right\}\right).$$

This result also holds for $n \in N_a$ and $n \in N_t$ with symmetrical expressions.

PROOF. The intra-degree in N_p is generated by common neighbors in N_a and N_t independently.

When a certain node $a \in N_a$ has node degree x from N_a to N_p, it has exactly x neighbors in N_p. Thus, the expected intra-degree in N_p added by this node is $2\gamma_{pa}\binom{x}{2}$, where γ_{pa} is the linking probability when two nodes inside node set N_p have a common neighbor node in N_a; and the number of nodes in N_a who have x neighbors in N_p is expected as $|N_a|\mathbb{P}\left\{d_{ap}^{ir}(t) = x\right\}$ where $\mathbb{P}$ denotes the probability that node in N_a having x neighbors in N_p exists and $|N_a|$ denotes the total nodes in N_a. Therefore, the intra-degree generated by nodes with x neighbors in N_a is

$$\text{Contribution}_a(x) = 2\gamma_{pa}\binom{x}{2}|N_a|\mathbb{P}\left\{d_{ap}^{ir}(t) = x\right\}. \tag{3}$$

Considering we add certain number of nodes with a certain probability in the node set, we get $|N_a| = \Theta(t)$. Thus, combining the result of Theorem 5.3, we get the intra-degree $d_{pa}^{ia}(t)$ in node set N_p contributed by node set N_a is

$$\begin{aligned} d_{pa}^{ia}(t) &= \sum_{x=1}^{\max_a} \text{Contribution}_a(x) \\ &= \sum_{x=1}^{\max_a} 2\gamma_{pa}\binom{x}{2}|N_a|\mathbb{P}\left\{d_{ap}^{ir}(t) = x\right\} \\ &= \Theta\left(\sum_{x=1}^{\max_a} x^2 x^{-\frac{1}{\lambda_a}-1} t\right) \\ &= \Theta\left(\sum_{x=1}^{t} x^{-\frac{1}{\lambda_a}+1} t\right), \end{aligned}$$

where $\max_a$ presents the maximum inter-degree from N_a to N_p which satisfies $\max_a = \Theta(t)$. By using the sum of *p-series*, we get

$$\sum_{x=1}^{t} x^{-\frac{1}{\lambda_a}+1} = t^{1-(1-\frac{1}{\lambda_a})}.$$

Therefore, we have $d_{pa}^{ia}(t) = \Theta\left(t^{\frac{1}{\lambda_a}+1}\right)$.

Considering the symmetric contribution of N_t to intra-degree in N_p, we get

$$\begin{aligned} d_p^{ia}(t) &= d_{pa}^{ia}(t) + d_{pt}^{ia}(t) \\ &= \Theta\left(\max\left\{t^{\frac{1}{\lambda_a}+1}, t^{\frac{1}{\lambda_t}+1}\right\}\right). \end{aligned}$$

By same approaches, we can also obtain the expression result of d_p^{ia} for nodes in N_a and N_t, thus we complete the proof. □

5.2 Power-law distribution

Power-law distribution is a classical node degree distribution which can be widely found in social network structure. By our design, we can also find power-law distribution in our scholarly network model and give proper mathematical proof.

Also, we analyze nodes' power-law distribution in two cases – inter and intra-degree respectively. However, there is little difference from the formal proof in 5.1, that we secondly do not study the intra-degree's case as its detailed distribution expression is hard to obtain, but study the general case when combining inter-degree and intra-degree together, i.e. the total degree of the nodes since when time slot $t \to \infty$.

Distribution of inter-degree: For the node $n \in N_i$ in $G(N_p, N_a, N_t)$ with $t \to \infty$, the inter-degree distribution of it follows

$$\mathbb{P}\left\{d_{ij}^{ir}(t) = x\right\} \propto x^{-\frac{1}{\lambda_i}-1}.$$

And we find that the inter-degree d^{ir} follows the power-law distribution with exponent $-\frac{1}{\lambda_i} - 1$.

Results show our model well capture the power-law distribution of nodes' inter-degree, which are proved in Theorem 5.3 and verified by simulating measurements.

Distribution of total degree: For the node n in $G(N_p, N_a, N_t)$ with $t \to \infty$, the total degree distribution of $n \in N_i$ follows

$$\mathbb{P}\{d_i(t) = x\} \propto x^{-\omega_i},$$

where ω_i is a constant which describes the exponential factor in power-law distribution.

This means our model well simulates the power-law distribution of nodes' total degree. This is proved in Theorem 5.4.

Theorem 5.3. *For graph $G(N_p, N_a, N_t)$ generated after t time slots, when $t \to \infty$, the inter-degree sequences of $n \in N_p$ in $B(N_p, N_a)$ and $B(N_p, N_t)$ both follow power-law distribution that*

$$\mathbb{P}\left\{d_p^{ir}(t) = x\right\} \propto x^{-\frac{1}{\lambda_p}-1},$$

where x is one node's total degree and $\mathbb{P}$ presents the probability. This result also holds for node $n \in N_a$ and $n \in N_t$ as they share symmetrical expressions.

Proof. We also divide the proof into two parts, i.e. prove the power-law in $d_{pa}^{ir}(t)$ and in $d_{pt}^{ir}(t)$ separately. First of all, we consider the distribution of $d_{pa}^{ir}(t)$ which denotes the degree of node $n \in N_p$ in $B(N_p, N_a)$. According to Equation (1), the cumulative distribution function of $d_{pa}^{ir}(t)$ can be calculated as

$$\begin{aligned} \mathbb{P}\left\{d_{pa}^{ir}(t) < x\right\} &= \mathbb{P}\left\{d_{pa}^{ir}(t_0)\left(\frac{t}{t_0}\right)^{\lambda_p} < x\right\} \\ &= \mathbb{P}\left\{t_0 > \left(\frac{d_{pa}^{ir}(t_0)}{x}\right)^{\frac{1}{\lambda_p}} t\right\} \\ &= 1 - d_{pa}^{ir}(t_0)^{\frac{1}{\lambda_p}} x^{-\frac{1}{\lambda_p}}. \end{aligned}$$

Then, the probability density function of $d_{pa}^{ir}(t)$ can be calculated using $\mathbb{P}\left\{d_{pa}^{ir}(t) = x\right\} = \frac{\partial \mathbb{P}\{d_{pa}^{ir}(t)<x\}}{\partial x}$. Also, it can be expressed as

$$\mathbb{P}\{d_{pa}^{ir}(t) = x\} = \frac{x^{-\frac{1}{\lambda_p}-1}}{\sum_{x=1}^{n} x^{-\frac{1}{\lambda_p}-1}},$$

where $\sum_{x=1}^{n} x^{-\frac{1}{\lambda_p}-1}$ is a constant normalization coefficient. Therefore, we get

$$\mathbb{P}\left\{d_{pa}^{ir}(t) = x\right\} \propto x^{-\frac{1}{\lambda_p}-1},$$

By same approaches, we can also calculate the distribution of $d_{ij}^{ir}(t)$, where $i \neq j \in \{p, a, t\}$ and thus the proof is complete. □

Theorem 5.4. *For graph $G(N_p, N_a, N_t)$ generated after t time slots, when $t \to \infty$, the nodes' total degree sequences of $n \in N_p$ follow power-law distribution that*

$$\mathbb{P}\{d_p(t) = x\} \propto x^{-\omega_p},$$

where x is one node's total degree, $\mathbb{P}$ presents the probability and ω_p is a constant. This result also holds for node $n \in N_a$ and $n \in N_t$ as they share symmetrical expressions.

Proof. The proof uses the result of *Silvio Lattanzi and D. Sivakumar*'s research work. [23]. In their work, the model's bipartite network's structure is similar to our model's bipartite networks' which are deconstructed from $G(N_p, N_a, N_t)$.

And by Theorem 4 and Theorem 8 in their paper, they fully prove the total degree distribution is similar to the inter-degree distribution when time slot $t \to \infty$. Which means the total degree is also power-law distributed.

Therefore, the total degree distribution in our model follows

$$\mathbb{P}\{d_p(t) = x\} \propto x^{-\omega_p},$$

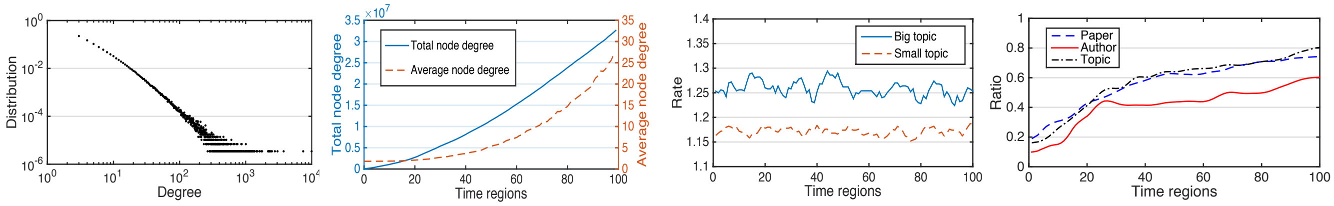
where ω_p is a constant.

Using same methods, we can obtain the distribution for node $n \in N_a$ and $n \in N_t$ and thus complete the proof. □

5.3 Densification

Now we turn to analyze the property of densification, defined as the phenomenon that the network's density of $G(N_i|N_j)$ – generated graph of N_i obtained from $B(N_i, N_j)$, increases with time if $\lambda_j \geqslant \frac{1}{2}$. The detailed proof is presented in Theorem 5.5.

Theorem 5.5. *For graph $G(N_p, N_a, N_t)$ which is generated after t time slots. The ratio of edges to nodes in $G(N_i|N_j)$ is*



(a) *Degree distribution* (b) *Growth of node degree* (c) *Topic-size evolution* (d) *Factors of power-law distributions*

Figure 9: Property verification of evolving scholarly model

$$\frac{|E|}{|V_i|} = \begin{cases} \Theta(1), & 0 < \lambda_j < \frac{1}{2} \\ \Theta(\log t), & \lambda_j = \frac{1}{2} \\ \Theta\left(t^{2-\frac{1}{\lambda_j}}\right), & \frac{1}{2} < \lambda_j < 1. \end{cases}$$

Proof. According to the definition of $G(N_i|N_j)$, each node $v \in N_j$ in $B(N_i, N_j)$ becomes a clique where all neighbors of v are connected with probability γ_{ij}. Therefore, the average number of edges in $G(N_i|N_j)$ is

$$|E| = \sum_{k=1}^{n} n \frac{k^{-\frac{1}{-\lambda_j - 1}}}{G_j} \gamma_{ij} C_k^2, \tag{4}$$

where $n\frac{k^{-\frac{1}{-\lambda_j-1}}}{G_j}$ is the average number of nodes with degree k in N_j. Since $|N| = n$, we can use the sum of *p-series* to get final results:

$$\lim_{n\to\infty} \sum_{x=1}^{n} \frac{1}{x^p} = \begin{cases} \Theta(1), & p > 1 \\ \Theta(\log n), & p = 1 \\ \Theta(n^{1-p}), & 0 \leqslant p < 1. \end{cases} \tag{5}$$

Therefore, we complete our proof. □

6 SIMULATION

Upon theoretical analysis of our model, in this section, we present simulations to verify that our model can correctly extract the properties of real multi-entity scholarly networks. According to the ratio of paper, author, and topic count in four datasets, we set $\alpha_p = 0.4136$, $\alpha_a = 0.5862$, $\alpha_t = 0.0002$; while other parameters are set as: $c_{ij} = 2$, $\beta_{ij} = 0.2 (i \neq j)$. Then after 2 million time slots, we get 827636 papers, 1171977 authors and 386 topics.

In real scholarly networks, we notice a tendency that during each year, more nodes are added to the network than last year. However, in our simulation, every equal time gap, we add the same number of nodes. To revise the influence of this factor and for the convenience of analysis, we divide the 2 million time slots into 100 time regions unequally. Each time regions can be reckoned as a "year" compared to real datasets. Based on the synthetic dataset, we verify the validity of our model and the result is listed below:

(1) Structural properties: Figure 9(a) shows the power-law distribution of paper degree in citation sub-network with $\varphi = 1.82$ and $\eta = 0.21$. As for the inter-degree, they all follow the power-law distribution and we give the final factors in table 4, which conforms to our observations on structural properties of scholarly networks in subsection 3.1.

Table 4: Simulation of power-law distribution factors

Factors	D_{pa}	D_{pt}	D_{ap}	D_{at}	D_{tp}	D_{ta}
φ	2.19	3.61	2.30	3.82	0.23	0.26
η	0.28	2.01	0.09	0.65	2.18	2.08

(2) Intra evolving properties: As shown in Figure 9(b), total node degree and average node degree in co-author network, i.e., the entity of author, both increase exponentially in the evolving process, which meets our expectation of patterns of intra-evolution in subsection 3.2.

(3) Inter evolving properties: From Figure 9(c), we again, acknowledge that big topic has the ability to increase its size faster than small topic. In our simulation, the average growth speed rate of big topic is 1.26 while the small topic is 1.17. This result also complies with inter-evolving properties explored in subsection 3.3.

(4) Co-evolving properties: As we can see from Figure 9(d), the connectivity of each entity changes simultaneously in a same pattern, which accords to co-evolving properties introduced in subsection 3.4. Thus, we can draw a conclusion that our model is able to reproduce the co-evolution among entities of paper, author, and topic within scholarly networks.

In general, the results based on our evolving scholarly model perform well in the simulation process, which additionally verifies MSM's capability of capturing the properties of multi-entity scholarly networks during its evolution.

7 CONCLUSIONS

In this paper, using Microsoft's datasets, we start with observing several interesting phenomena in the structure and evolution of real scholarly networks. Based on observations, we generate the MSM to capture both intra and inter correlations of paper, author and topic during the evolving process. Further, through theoretical analysis and simulating evaluations, we validate that our model can accurately reproduce both structural and evolving patterns of real multi-entity scholarly networks.

We believe that our work is one of the first steps in modeling and analyzing evolving scholarly networks from a multi-entity perspective, and that there remain some future directions which wait for exploration. On one hand, the structure of our MSM can be mapped into counterparts to model other kinds of multi-entity social networks, as mentioned in subsection 4.3. On the other hand, since the number of entities in social networks might be extremely large or even also evolving, there is room for our current model to improve to accommodate more complicated multi-entity social networks.

REFERENCES

[1] Zaihan Yang, Liangjie Hong, and Brian D Davison. Academic network analysis: A joint topic modeling approach. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pages 324–333. ACM, 2013.

[2] Yuya Kajikawa, Junko Ohno, Yoshiyuki Takeda, Katsumori Matsushima, and Hiroshi Komiyama. Creating an academic landscape of sustainability science: an analysis of the citation network. *Sustainability Science*, 2(2):221, 2007.

[3] Jing Li, Feng Xia, Wei Wang, Zhen Chen, Nana Yaw Asabere, and Huizhen Jiang. Acrec: a co-authorship based random walk model for academic collaboration recommendation. In *Proceedings of the 23rd International Conference on World Wide Web*, pages 1209–1214. ACM, 2014.

[4] Yu-Shan Lin. Topic evolution of innovation academic researches. *Journal of Small Business Strategy*, 26(1):25, 2016.

[5] Yizhou Sun, Rick Barber, Manish Gupta, Charu C Aggarwal, and Jiawei Han. Co-author relationship prediction in heterogeneous bibliographic networks. In *Advances in Social Networks Analysis and Mining (ASONAM), 2011 International Conference on*, pages 121–128. IEEE, 2011.

[6] Microsoft academic graph, 2016.

[7] Jason Priem. Scholarship: Beyond the paper. *Nature*, 495(7442):437–440, 2013.

[8] Jie Tang, Jing Zhang, Limin Yao, Juanzi Li, Li Zhang, and Zhong Su. Arnetminer: extraction and mining of academic social networks. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 990–998. ACM, 2008.

[9] Feng Xia, Wei Wang, Teshome Megersa Bekele, and Huan Liu. Big scholarly data: A survey. *IEEE Transactions on Big Data*, 2017.

[10] Mark Newman. Networks: an introduction. 2010. *United Slates: Oxford University Press Inc., New York*, pages 1–2, 2010.

[11] Wei Wang, Jiaying Liu, Shuo Yu, Chenxin Zhang, Zhenzhen Xu, and Feng Xia. Mining advisor-advisee relationships in scholarly big data: A deep learning approach. In *Digital Libraries (JCDL), 2016 IEEE/ACM Joint Conference on*, pages 209–210. IEEE, 2016.

[12] Claudia Colicchia, Alessandro Creazza, and Fernanda Strozzi. Citation network analysis for supporting continuous improvement in higher education. *Studies in Higher Education*, pages 1–17, 2017.

[13] Martin Atzmueller, Andreas Ernst, Friedrich Krebs, Christoph Scholz, and Gerd Stumme. On the evolution of social groups during coffee breaks. In *Proceedings of the 23rd International Conference on World Wide Web*, pages 631–636. ACM, 2014.

[14] Yanhong Wu, Naveen Pitipornvivat, Jian Zhao, Sixiao Yang, Guowei Huang, and Huamin Qu. egoslider: Visual analysis of egocentric network evolution. *IEEE transactions on visualization and computer graphics*, 22(1):260–269, 2016.

[15] Jure Leskovec, Jon Kleinberg, and Christos Faloutsos. Graphs over time: densification laws, shrinking diameters and possible explanations. In *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*, pages 177–187. ACM, 2005.

[16] Elena Zheleva, Hossam Sharara, and Lise Getoor. Co-evolution of social and affiliation networks. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1007–1016. ACM, 2009.

[17] Michalis Faloutsos, Petros Faloutsos, and Christos Faloutsos. On power-law relationships of the internet topology. In *ACM SIGCOMM computer communication review*, volume 29, pages 251–262. ACM, 1999.

[18] Duncan J Watts and Steven H Strogatz. Collective dynamics of fismall-worldfinetworks. *nature*, 393(6684):440–442, 1998.

[19] Jon Kleinberg. The small-world phenomenon: An algorithmic perspective. In *Proceedings of the thirty-second annual ACM symposium on Theory of computing*, pages 163–170. ACM, 2000.

[20] Deepayan Chakrabarti and Christos Faloutsos. Graph mining: Laws, generators, and algorithms. *ACM computing surveys (CSUR)*, 38(1):2, 2006.

[21] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.

[22] Ravi Kumar, Prabhakar Raghavan, Sridhar Rajagopalan, D Sivakumar, Andrew Tomkins, and Eli Upfal. Stochastic models for the web graph. In *Foundations of Computer Science, 2000. Proceedings. 41st Annual Symposium on*, pages 57–65. IEEE, 2000.

[23] Silvio Lattanzi and D Sivakumar. Affiliation networks. In *Proceedings of the forty-first annual ACM symposium on Theory of computing*, pages 427–434. ACM, 2009.

[24] Xiaolong Wang, Chengxiang Zhai, and Dan Roth. Understanding evolution of research themes: a probabilistic generative model for citations. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1115–1123. ACM, 2013.

[25] Jiaqi Liu, Yuhang Yao, Xinzhe Fu, Luoyi Fu, Xiao-Yang Liu, and Xinbing Wang. Evolving k-graph: Modeling hybrid interactions in networks. In *Proceedings of the 18th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, page 29. ACM, 2017.

[26] Albert-László Barabási, Réka Albert, and Hawoong Jeong. Mean-field theory for scale-free random networks. *Physica A: Statistical Mechanics and its Applications*, 272(1):173–187, 1999.

[27] Lev Muchnik, Sen Pei, Lucas C Parra, Saulo DS Reis, José S Andrade Jr, Shlomo Havlin, and Hernán A Makse. Origins of power-law degree distribution in the heterogeneity of human activity in social networks. *arXiv preprint arXiv:1304.4523*, 2013.

[28] Cong Xie, Ling Yan, Wu-Jun Li, and Zhihua Zhang. Distributed power-law graph computing: Theoretical and empirical analysis. In *Advances in Neural Information Processing Systems*, pages 1673–1681, 2014.

[29] Michael Molloy and Bruce Reed. The size of the giant component of a random graph with a given degree sequence. *Combinatorics, probability and computing*, 7(03):295–305, 1998.