Diversity Index of Academic Community Ecosystem by Co-authorship Analysis with Granger Causality

Rui Wang
Zaozhuang University
Zaozhuang, China

***Abstract*—Compared to individual-level analysis, community-level analysis provides a new perspective to inspect network structure. It focuses on modeling the evolving relationships between communities. Intuitively, community-level analysis is a generalization of individual-level analysis. It reflects a macroscopic evolution of a network and reduces the overfitting of individual analysis to some degree. In this paper, we investigate the co-authorship characteristics between different affiliations in academic social networks and then adopt the weighted multigraph model to establish the coauthorships between communities. Subsequently, we define the Co-authorship Factor (CF) for each pair of communities and then propose the modified Shannon Co-authorship Diversity Index (SCDI) and Renyi Co-authorship Diversity Index (RCDI) to measure the diversity of co-authorship ecosystem of a certain community. Finally, we apply the Granger causality to model the mutual co-authorship influences between communities along time. We verify our proposed indexes on real dataset which is mainly based on the DBLP and Microsoft Academic Graph (MAG) datasets.**

Keywords—Community analysis; Granger causality; Academic social

I. INTRODUCTION

Networks are ubiquitous everywhere in the world. People are connected to each other by certain relationships, such as friends, family, colleagues, or relatives, etc. Correspondingly, different people can be distinguished into distinct groups or communities according to certain attributes or relationships. In more general sense, every connected object can be regarded as an entity in a network, and each entity may belong to one or more communities. Plenty of works has been conducted to analyze topologies of different networks (Boss et al. 2004; Ganesh, Massoulie, and Towsley ´2005; Rohden et al. 2014), finding influential entities (Aral and Walker 2012; Probst, Grosswiele, and Pfleger 2013; Trusov, Bodapati, and Bucklin 2010), and detecting communities (Fortunato 2010; Fortunato and Barthelemy 2007). Furthermore, researchers investigate the problem of information diffusion via influential users in social networks (Bakshy et al. 2012; Chen, Wang, and Yang 2009; Kempe, Kleinberg, and Tardos 2003). However, most existing works only focus on the relationships or connections between a pair of entities, few work has studied community-level relationships.

Recently, community-level analysis in social networks has shown new characteristics and drawn consistent interests from scholars. For instance, in order to analyze community level temporal information diffusion process, (Hu et al. 2015) introduces a unified latent framework to model topics and communities, and then extracts inter-community influence dynamics. Similarly, (Mehmood et al. 2013) proposes a novel information propagation model generalizing the classic Independent Cascade model to deal with communitylevel social influence analysis, and thus provides new influence patterns in large scale social networks. In this paper, we apply the community-level influence analysis to academic social networks. Specifically, the academic network involves an enormous number of scholars around the world and countless scientific literatures in every year. Therefore, it is essential to build a citation network between papers (Chen and Redner 2010; Dawson et al. 2014; Kajikawa et al. 2007)and scholar social network between researchers (tangjie). However, although there have been works dealing with the construction of academic networks and academic community detection (Tang, Zhang, and Yao 2007; Tang et al. 2008), few have investigated the relationship and influence between academic communities. (Chikhaoui, Chiazzaro, and Wang 2015) studies the citation influence evolution between different conferences in the field of Artificial Intelligence and Data Mining. Inspired by that, we study the co-authorship evolution between scholars in different academic affiliations. Based on the co-authorship statistics, we then propose an index to measure the diversity of research ecosystem of a certain affiliation. Finally, we present the Granger causality to model the evolution of mutual influence between two communities. Our contributions are summarized as follows:

- Proposing community-level influence model to analyze the academic co-authorships between different affiliations.
- Presenting different indexes to measure the diversity of academic ecosystem in terms of inter-community and intra-community co-authorships .
- Modeling inter-community co-authorships as time series and analyzing the reciprocal influences between two time series using Granger causality.

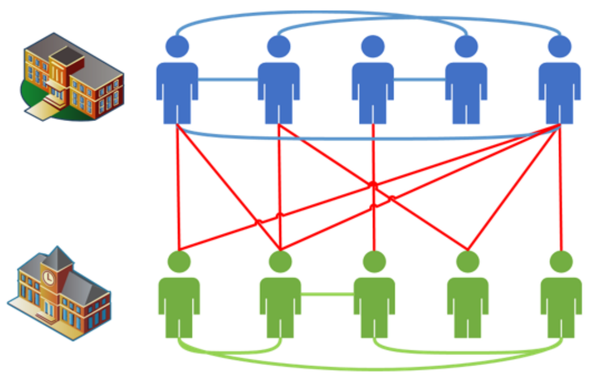

Fig. 1 Illustration of co-authorship between two different communities, i.e., affiliations. Each affiliation has self-coauthorships(i.e., co-authorships between authors in the same affiliation) and non-self-co-authorships (i.e., co-authorships between authors in different affiliations).

The rest of the paper is organized as follows: In Section 2, we introduce the related work. In Section 3, we propose the problem formulation in terms of weighted multigragh model, and then we analyze the co-authorship between communities and present different indexes to measure the diversity of academic ecosystem for different academic communities. Subsequently, we propose the Granger causality test to analyze the evolution of co-authorship. The experimental results based on real academic social networks are presented in Section 4. Finally, the paper is concluded in Section 5.

II. Related Work

A. Community Level Influence Analysis

With the rapid growth of data volume, community detection has become an important task in AI research. It aims to distinguish data into groups by similar attributes or features. The data in the same community normally either share common properties or have closer relations to each other. Correspondingly, the community detection methods include detection by data source attributes or by data interaction features (Lancichinetti and Fortunato 2009). For that reason, the community detection issue can be treated as finding subgraphs with certain network structure patterns (Fortunato 2010). In contrast to individual analysis, the research of a community shows collective alteration of a group of correlated data. Given the ever growing volume of big data, community-level data mining becomes more necessary and also more rational. For example, with the growth of population, it is more reasonable to analyze the life habit, citizen behavior or economical development in a certain district such as a city or a county because of geographical influence. Similarly, when investigating the citation influence of academic literatures, researchers currently are studying conference-level or topic-level citation influence (Chikhaoui, Chiazzaro, and Wang 2015; Liu et al. 2010), which generates a macroscopic vision of academic development of a target cluster. Intuitively, community-level influence analysis manifests generalized characteristics of objects and avoids individual overfitting to some degree.

B. Academic Co-authorship Networks

In academia, scholars all around the world cooperate with others in research and publish scientific literatures together. We consider two authors have an co-authorship if they appear in the same paper. Therefore, we will have a coauthorship network that connects authors together. There have been works analyzing the co-authorship networks, research community detection and author productivity and so on (Glanzel and Schubert 2004; Liu et al. 2005). It is impor- ¨tant to discover the information in co-authorship networks.

C. Diversity Index

A diversity index is a measurement that quantifies two aspects of a specific dataset. Firstly, it manifests the number of types of data in a dataset. Secondly, it simultaneously takes into account how evenly every type of data is distributed. Moreover, the value of a diversity index increases both when the number of types increases and when the evenness of distribution of each type increases. Furthermore, when the number of types is fixed, the value of a diversity index is maximized when all types are equally allocated. In the context of our work, given an academic community, the diversity index indicates the number of communities it coauthors with and the evenness of cooperation with these communities. The value of diversity index provides certain insights of the community's academic influence in a specific research area such as Computer Science.

D. Granger Causality

The Granger causality, first proposed in (Granger 1969), is a statistical hypothesis test to investigate whether one time series helps to predict another one. Specifically, if one time series T1 makes the prediction of another time series T2 more accurate, i.e., T1 can provide useful information in predicting future value of T2, then we define that T1 Granger causes

T2. Granger causality is first presented in the field of economics and then extended to various areas attributed to its computational simplicity and robustness (Ioannis, David, and Wan). In our paper, we adopt the Granger causality to analyze the mutual influences of the co-authorship time series between two communities.

III. Problem Formulation

In this section, we first present a weighted multigraph model to establish community-level coauthoring relationship among different affiliations. Then we propose several measurements to quantify the contribution of one community to another in terms of co-authorship counts. Based on the proposed measurements, we then propose the modified Shannon Co-authorship Diversity Index (SCDI) and Renyi ´ Co-authorship Diversity Index (RCDI) to scale the diversity of co-authorship ecosystem of a community. Subsequently, we present another index named Balance Index (BI) to quantify the balance between internal connections and external cooperations for a certain community in terms of coauthorship. Finally, we use the Granger causality to analyze the potential reciprocal contributions between communities. Hopefully, these can provide some insights for universities to adjust their academic strategies.

A. Weighted multigraph

Since different authors in an affiliation may coauthor with different authors from another affiliation or the same affiliation in different publication years, thus, there will be multiple edges between a pair of affiliations. Given the characteristics of co-authorship networks, we need a graph that is able to capture multiple co-authorships between a pair of communities. For that reason, we propose the weighted multigraph to manifest the co-authorship network.

Mathematically, we use G = (V,E,Y,W) to denote the weighted multigraph, where V represents the set of nodes (i.e., affiliations), E represents the set of edges indicating the existence of co-authorship between different nodes, Y represents the publication year of the coauthored literature, and W is the weight function of a certain edge e.

Given a certain publication year, each affiliation will have a co-authorship vector with the other ones. Since coauthorships between authors are mutual, the edges between a pair of affiliations are indirected. Nevertheless, different affiliations may have distinct cooperation affiliations, thus the weight of each edge is different. For that reason, we define the following measurements to distinguish the co-authorship weights of different affiliations.

First we propose the Co-authorship Factor (CF) for a node u, which is defined as

$$CF(u \to v, y_i) = \frac{N_{u,v}(y_i)}{N_u(y_i)}\,, \tag{1}$$

where Nu, v is the number of co-authorships between all the authors in community u and v in time yi, and Nu(yi) is the total number of co-authorships of all the authors in community u in time yi. CF(u → v,yi) indicates the contribution of community v to the research of community u in a specific field in a given year. Naturally, we can divide all the coauthorships of community u into two parts: co-authorships with authors inside and outside community u. Accordingly, we propose the Self-Co-authorship Index (SCI) and NonSelf-Co-authorship Index (NSCI) as following

$$SCI_u(y_i) = \frac{N_{u,u}(y_i)}{N_u(y_i)}, \tag{2}$$

$$NSCI_u(y_i) = \frac{\sum_{v\in V(u), v\neq u} N_{u,v}(y_i)}{N_u(y_i)} = 1 - SCI_u(y_i), \tag{3}$$

where Vu is the set of communities that u has co-authorships with in time yi.

B. Co-authorship Diversity Index

A diversity index provides intuitive knowledge of the coauthorship ecosystem of a community. Specifically, it both reflects the number of different communities it coauthors with and meanwhile takes into account the evenness of these co-authorships.

According to (Jost 2006), the generalized co-authorship diversity index can be defined as

$$^qD = \frac{1}{M_{q-1}} = \left(\sum_{v\in V(u)} (CF(u\to v, y_i))^q\right)^{1/(1-q)}, \tag{4}$$

where Mq−1 is the weighted mean of the proportions of coauthored communities. q is the order of the diversity. Based on the generalized diversity equation, we subsequently present the following two specific co-authorship indexes.

Shannon Co-authorship Diversity Index (SCDI) According to the definition of Shannon entropy in information theory and its application to measure diversity (Jost 2006), we propose a new measurement to calculate the diversity of co-authorship ecosystem of a certain community. Specifically, the diversity of ecosystem is high when a community has co-authorships with multiple communities and the weights of co-authorships with different communities are close to each other. Intuitively, we claim a community has diverse academic ecosystem when it builds a wide variety of connections with various communities and meanwhile, maintains its own academic strength. Inspired by that, we present the Shannon Co-authorship Diversity Index (SCDI) of a community u at time yi as

$$\mathrm{SCDI}_u(y_i) = -\sum_{v\in V_u} \mathrm{CF}(u\to v, y_i)\ln \mathrm{CF}(u\to v, y_i), \tag{5}$$

where $\sum_{v\in V_u} \mathrm{CF}(u\to v, y_i) = 1$. The value of SCDI of community u achieves its maximum when u has equal Co-authorship Factors with all the communities it coauthors with, including itself. In this case, we claim that community u has the optimal diversity of co-authorship ecosystem in a specific research area.

Renyi Co-authorship Diversity Index (RCDI) ´ Renyi entropy is regarded as a generalization of the Shannon entropy by assigning different values to the diversity order q. (Jost 2006). The general definition is given by

$$^qR = \ln(^qD) = \frac{1}{q-1}\ln\left(\sum_{v\in V_u} (\mathrm{CF}(u\to v, y_i))^q\right). \tag{6}$$

For simplicity, we assign the diversity order q = 2, and then we have the Renyi Co-authorship Diversity Index (RCDI): ´

$$RCDI_u(y_i) = \ln\left(\sum_{v\in V_u} (\mathrm{CF}(u\to v, y_i))^2\right). \tag{7}$$

Balance Index (BI) Furthermore, in order to measure the balance degree between the internal connections and external cooperations of a community, we propose a simplified co-authorship diversity index named Balance Index (BI), which is defined as following

$$BI_u(y_i) = -SCI_u(y_i)\ln SCI_u(y_i) - NSCI_u(y_i)\ln NSCI_u(y_i). \tag{8}$$

C. Granger Causality

In addition to the static diversity analysis, we are also interested in the dynamic evolution of co-authorship between communities, which exhibits the development of mutual influences between communities. Based on the above analysis, the co-authorship between two communities can be formulated as a problem of time series evolution. Specifically, for community u and v, we have two stationary time series CF(u → v,y),y ∈ Y and CF(v → u,y),y ∈ Y . Intuitively, we conjecture that these two time series are correlated to each other to some degree. In order to analyze how one time series can impact the other one, in this section, we propose the Granger Causality model as the assessment.

In the following, we present both the null hypothesis H0 and hypothesis H1:

$$H_0 : \mathrm{CF}(u \to v, y) = \sum_{l=1}^{L} a_l \mathrm{CF}(u \to v, y-l) + \epsilon_1,$$

$$\begin{aligned} H_1 : \mathrm{CF}(u \to v, y) &= \sum_{l=1}^{L} a_l \mathrm{CF}(u \to v, y-l) \\ &+ \sum_{l=m}^{n} b_l \mathrm{CF}(v \to u, y-l) + \epsilon_2, \end{aligned} \tag{10}$$

where L is the maximal time lag for autoregression, m and n are the shortest and longest lag length for which the past values of CF(v → u,y) provide significant explanatory power to the regression, al and bl are the regression variable coefficients, and $\epsilon_1, \epsilon_2 \sim \mathcal{N}(0; \sigma^2)$ are the independent and identically distributed residual variables.

According to the definition of Granger causality, if model H1 achieves better performances than model H0, then we claim that CF(v → u,y) Granger-causes CF(u → v,y).

IV. EXPERIMENTAL RESULTS

A. Dataset

In the experimental section, we obtain important statistical results from our dataset which is a combination of the DBLP and Microsoft Academic Graph (MAG) datasets. Our dataset contains approximately 126,000,000 scientific literatures and 114,000,000 authors in the field of Computer Science from 1900 to 2016, including information of the authors in each literature. Based on this dataset, without loss of generality, we choose 10 representative academic affiliations to conduct our experiments, half of which are from US and the other half from China. In this case, the diversity of affiliations can be guaranteed with both national and international universities included. Moreover, we choose all the 10 affiliations from US News Ranking (usn 2016) which is publicly recognized. Basically we choose one out of every five affiliations. The selected 10 affiliations are as follows: Massachusetts Institute of Technology (MIT), California Institute of Technology (CIT), University of Illinois at UrbanaChampaign (UIUC), Columbia University (CU), Duke University (DU), Shanghai Jiao Tong University (SJTU), Tsinghua University (THU), Nanjing University (NJU), University of Science and Technology of China (USTC) and Fudan University (FU). In each affiliation, we select the first 3000 authors with the highest citations.

B. Statistics

Table 1, 2, and 3 are statistical co-authorship counts between any two affiliations from 2012-16, 2007-11, 2002-06 respectively. In the experiment, we analyze the total co-authorship counts in 5 years instead of a single year so that the statistics can be regarded as stationary time series. From the tables, we can have an intuitive knowledge of the co-authorship evolution in these fifteen years in the field of Computer Science.

For instance, in Table 1, we can see that there are coauthorships between all the affiliations. Basically, an affiliation has the highest co-authorship counts with itself except for the case of MIT and CIT. It is normal to have high co-authorships with authors within the same affiliation since they belong to the same community and naturally have more interactions with each other. However, it seems that MIT and CIT have extremely frequent academic cooperations in year 2012-2016, with a total co-authorship counts of 17892, which is approximately twice the number of selfco-authorship counts. Figure 2 shows the number of total co-authorship, which maintains increasing with time. In addition, MIT has the highest increase in total and UIUC has the lowest by contrast. It is also noticeable that THU almost has the highest co-authorship counts in all three time intervals, which possibly indicates that THU is the most active community of the ten in 15 years from 2002-2016 in Computer Science.

Figure 3 shows an example of the co-authorship relationship between UIUC and another three communities. Each edge in this figure represents the contribution of a community to UIUC in terms of co-authorship frequencies. For example, the connection (2007-11,0.105) between UIUC and MIT represents that in 2007-11, the proportion of coauthorships with MIT accounts for approximate 10.5% of all the co-authorships of UIUC, i.e., the contribution of MIT to UIUC is 0.105.

C. Co-authorship Diversity Index

Figure 4 shows the proposed Shannon Co-authorship Diversity Index (SCDI) for each community between 200206, 2007-11, and 2012-16 respectively. We can see that for every community, there is an increase in SCDI along time, indicating that the entire co-authorship ecosystem becomes more and more diverse than past years. Probably it is attributed to the development of communication technology and higher degree of globalization which facilitates the cooperations between communities around the world and thus contributes to more frequent interactions.

Moreover, universities from US generally have higher increasing rates than those from China. The explanation of this phenomena may be that with more and more researchers or potential researchers in China going to US for higher degree or research cooperation, US researchers become active instead of being passive in academic cooperations with China.

More generally, it is a process of resource re-allocation between resourceful communities and resourceless ones and the result is becoming more homogeneous.

Figure 5 shows the proposed Renyi Co-authorship Di-´ versity Index (RCDI). From (7) we know RCDI is a measurement of diversity with order 2. The experimental results show similar characteristics with SCDI. Figure 6 is the Balance Index (BI) of communities. BI is the measurement of balance between self-co-authorship and non-selfco-authorship, indicating the ratio of internal and external cooperations of communities. From the figure, we can see that, different from SCDI and RCDI, BI is not increasing monotonically with time. Some communities such as NJU and USTC have the opposite trend with monotonically decreasing value of BI. This may be caused by higher ratio of co-authorships with the other communities and relatively lower ratio of self-co-authorships. However, for most com-

munities, such as MIT, CIT, UIUC, CU, DU, SJTU, THU and FU, the value of BI reaches highest in 2007-11 and falls lower in 2012-16. Probably the phenomena results from the process of globalization. The balance is obtained and then lost with international cooperations increasing from a lower level. Hopefully, in the next phase, the balance can be regained with self-co-authorship and non-self-co-authorship levels become close to each other. Furthermore, BI also provides information for academic affiliations to make adjustments on their research strategy, i.e., either strengthening internal connections or seeking for external cooperations.

D. Granger Causality Analysis

In order to analyze the influence of co-authorship between two communities. We use the Granger Sargent test (Granger 1969) to determine the direction of contribution between communities. The test is given by

$$F = \frac{(RSS_1 - RSS_2)/L}{RSS_2/(n-2L)}, \qquad (11)$$

where RSSi is the residual sum of squares under model Hi,i = 0,1 as defined in Eq. (9) and (10). L and n−2L are the degree of freedom for each model. The null hypothesis H0 is rejected if the F − statistic calculated from the data is greater than a threshold. In order to analyze the direction of contribution, we calculate the reciprocal F − statistic between a pair of communities and reject H0 with higher F − statistic and lower false probability.

We extract the data of CF(u → v) for each year from 2006-2016 between MIT, UIUC, SJTU and THU. Then we calculate the F-statistic and probability for each pair. The results are shown in Table 4. For example, MIT −to→ UIUC represents the ratio of the contribution to UIUC by MIT in terms of co-authorship counts. From the table, we can see that MIT −to→ UIUC has higher F-statistic than UIUC −to→ MIT. Thus, there is a higher probability that MIT contributes to UIUC more than the opposite. The results only show an specific perspective of analyzing the direction of contribution between communities and can only be viewed as a reference.

TABLE I CO-AUTHORSHIP COUNTS BETWEEN TEN AFFILIATIONS FROM 2012 TO 2016

Affiliation	MIT	CIT	UIUC	CU	DU	SJTU	THU	NJU	USTC	FU
MIT	9308	17892	2338	3591	2248	2717	3327	3484	2803	3257
CIT		8214	1386	2193	1356	1674	2176	2031	1350	1861
UIUC			3851	1153	843	282	390	490	620	276
CU				7745	7188	1331	2006	3107	4507	1578
DU					4732	832	1117	1779	2505	910
SJTU						13759	3499	4012	3459	3835
THU							18938	4518	5417	4573
NJU								9650	2834	2726
USTC									9690	2834
FU										13006

TABLE II CO-AUTHORSHIP COUNTS BETWEEN TEN AFFILIATIONS FROM 2007 TO 2011

Affiliation	MIT	CIT	UIUC	CU	DU	SJTU	THU	NJU	USTC	FU
MIT	8130	5486	1121	1102	700	842	1002	1039	927	1111
CIT		6239	624	792	378	598	795	673	612	669
UIUC			6923	227	282	258	356	335	338	259
CU				5079	597	312	361	396	506	333
DU					4689	253	269	261	304	236

Table II, cont

SJTU						14183	3098	3560	3246	2983
THU							20621	4640	5561	4674
NJU								9893	2368	2298
USTC									9075	1899
FU										11522

TABLE III CO-AUTHORSHIP COUNTS BETWEEN TEN AFFILIATIONS FROM 2002 TO2006 .

Affiliation	MIT	CIT	UIUC	CU	DU	SJTU	THU	NJU	USTC	FU
MIT	5035	431	192	175	163	87	80	84	70	69
CIT		3179	82	102	61	63	79	70	82	61
UIUC			4510	151	148	273	314	280	247	242
CU				2745	100	93	136	120	111	89
DU					2902	93	119	87	101	79
SJTU						6663	1041	1198	1143	1123
THU							10763	1835	2663	1943
NJU								3855	796	789
USTC									3672	632
FU										3754

TABLE IV CONTRIBUTION DIRECTION ANALYSIS BETWEEN AFFILIATIONS USING GRANGER CAUSALITY TEST.

Affiliation	F-statistic	Probability	Contribution direction
MIT −to→ UIUC UIUC −to→ MIT	29.9259 2.2599	5.94e-04 0.1712	MIT −to→ UIUC
MIT −to→ SJTU SJTU −to→ MIT	22.7334 10.2355	0.0014 0.0126	MIT −to→ SJTU
MIT −to→ THU THU −to→ MIT	2.644 7.402	0.1426 0.0262	THU −to→ MIT
UIUC −to→ SJTU SJTU −to→ UIUC	77.9742 42.6863	2.13e-05 1.82e-04	UIUC −to→ SJTU
THU −to→ UIUC UIUC −to→ THU	30.6092 5.1029	5.52e-04 0.0538	THU −to→ UIUC
THU −to→ SJTU SJTU −to→ THU	1.9439 1.2403	0.2008 0.2978	THU −to→ SJTU

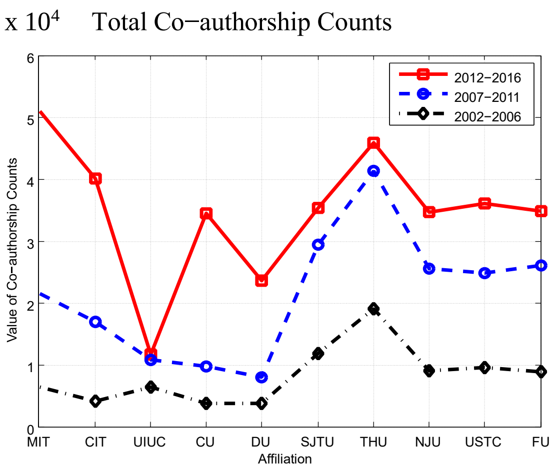

Fig. 2 Total number of co-authorship counts for each of the ten affiliations from year 2002-2016.

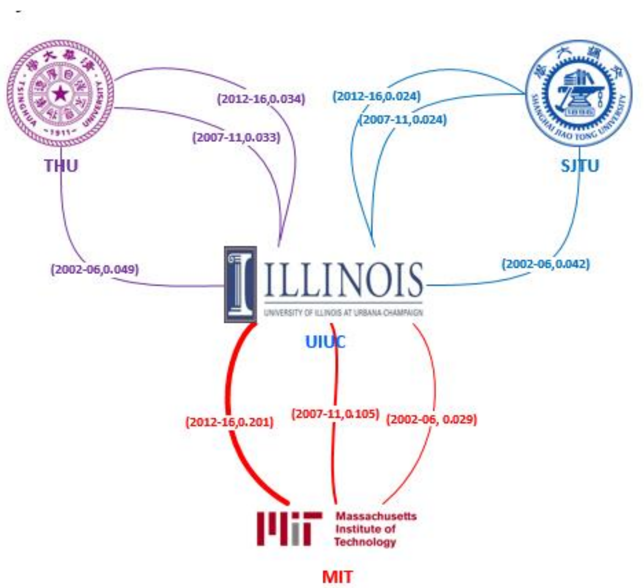

Fig. 3 Example of a weighted multigraph exhibiting the co-authorship relationship between UIUC and MIT, SJTU, and THU respectively. Every edge in the multigraph is denoted by a tuple, for instance, the tuple (2012-16, 0.201) between UIUC and MIT represents the co-authorship weight of MIT for UIUC is 0.201 in 2012-2016, i.e., 20.1% of the coauthors of UIUC are from MIT in 2012-2016.

V. Conclusion

In this paper, we study the co-authorship evolution at community level in academic social networks. First, we model the co-authorships between different research communities with weighted multigraph. Then we propose indices for intra-community and inter-community co-authorships. Subsequently, in order to quantify the diversity of co-authorship ecosystem of a certain community, we present the modified Shannon Co-authorship Diversity Index (SCDI) and Renyi Co-authorship Diversity Index (RCDI) to scale both´ the quantity and evenness of coauthored communities. In addition, we also analyze the evolution of co-authorship time series and use the Granger causality to examine the recip-Shannon Co−authorship Diversity Index (SCDI)

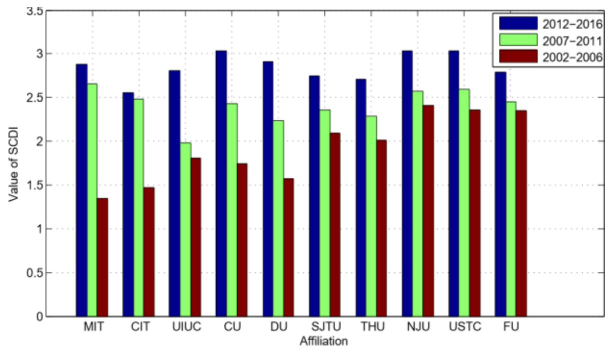

Fig. 4 Shannon Co-authorship Diversity Index (SCDI) for each of the ten affiliations from year 2002-2016.

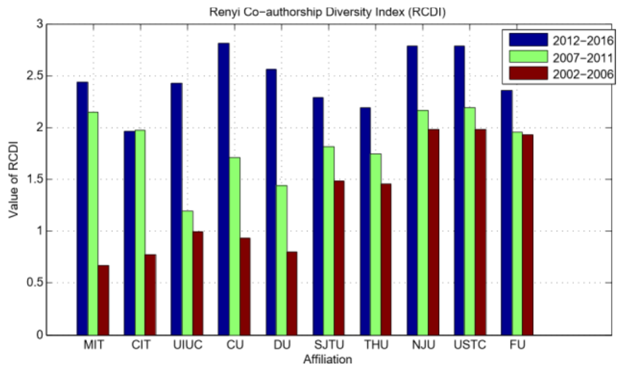

Fig. 5 Renyi Co-authorship Diversity Index (RCDI) for´ each of the ten affiliations from year 2002-2016.

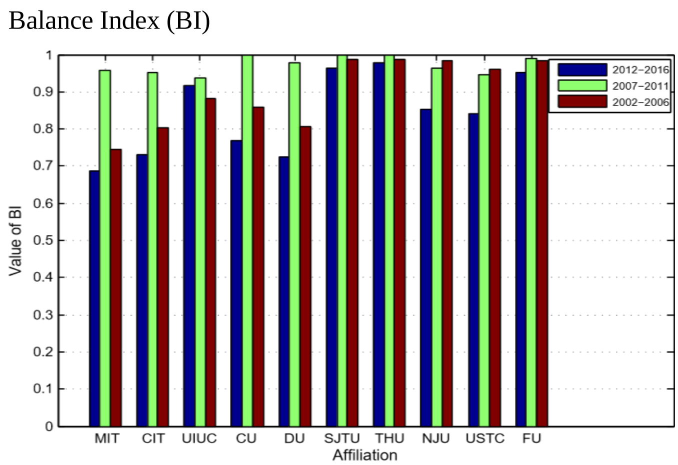

Fig. 6 Balance Index (BI) for each of the ten affiliations from year 2002-2016.

Rocal influences between each other. Finally, we verify our proposed indices on our co-authorship dataset which is a combination of the DBLP and MAG datasets. In the future work, we intend to incorporate authors from the cited papers and calculate a weighted co-authorship index for each community and justify our results on more datasets in different research areas.

REFERENCES

[1] Aral, S., and Walker, D. 2012. Identifying influential and susceptible members of social networks. Science 337(6092):337–341.

[2] Bakshy, E.; Rosenn, I.; Marlow, C.; and Adamic, L. 2012. The role of social networks in information diffusion. In Proceedings of the 21st international conference on World Wide Web, WWW ’12, 519–528. New York, NY, USA: ACM.

[3] Boss, M.; Elsinger, H.; Summer, M.; and Thurner 4, S. 2004. Network topology of the interbank market. Quantitative Finance 4(6):677–684.

[4] Chen, P., and Redner, S. 2010. Community structure of the physical review citation network. Journal of Informetrics 4(3):278–290.

[5] Chen, W.; Wang, Y.; and Yang, S. 2009. Efficient influence maximization in social networks. In Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, 199–208. New York, NY, USA: ACM.

[6] Chikhaoui, B.; Chiazzaro, M.; and Wang, S. 2015. A new granger causal model for influence evolution in dynamic social networks: The case of dblp. In AAAI, 51–57. AAAI. Press.

[7] Dawson, S.; Gaseviˇ c, D.; Siemens, G.; and Joksimovic, S.´ 2014. Current state and future trends: A citation network analysis of the learning analytics field. In Proceedings of the Fourth International Conference on Learning Analytics And Knowledge, LAK ’14, 231–240. New York, NY, USA: ACM.

[8] Fortunato, S., and Barthelemy, M. 2007. Resolution limit in community detection. Proceedings of the National Academy of Sciences 104(1):36–41.

[9] Fortunato, S. 2010. Community detection in graphs. Physics reports 486(3):75–174.

[10] Ganesh, A.; Massoulie, L.; and Towsley, D. 2005. The effect ´ of network topology on the spread of epidemics. In Proceedings IEEE 24th Annual Joint Conference of the IEEE Computer and Communications Societies., volume 2, 1455– 1466. IEEE.

[11] Glanzel, W., and Schubert, A. 2004. Analysing scientific ¨ networks through co-authorship. In Handbook of quantitative science and technology research. Springer. 257–276.

[12] Granger, C. . W. J. 1969. Investigating causal relations by econometric models and cross-spectral methods. Econometrica 37(3):424–438.

[13] Hu, Z.; Yao, J.; Cui, B.; and Xing, E. 2015. Community level diffusion extraction. In Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data, SIGMOD ’15, 1555–1569. New York, NY, USA: ACM.

[14] Ioannis, A.; David, A.; and Wan, M. M. Non-linear granger causality in the currency futures returns. Economics Letters 68(1):25–30.

[15] Jost, L. 2006. Entropy and diversity. Oikos 113(2):363–375.

[16] Kajikawa, Y.; Ohno, J.; Takeda, Y.; Matsushima, K.; and Komiyama, H. 2007. Creating an academic landscape of sustainability science: an analysis of the citation network. Sustainability Science 2(2):221–231.

[17] Kempe, D.; Kleinberg, J.; and Tardos, E. 2003. Maximizing ´the spread of influence through a social network. In Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, KDD ’03, 137–146. New York, NY, USA: ACM.

[18] Lancichinetti, A., and Fortunato, S. 2009. Community detection algorithms: a comparative analysis. Physical review E 80(5):056117.

[19] Liu, X.; Bollen, J.; Nelson, M. L.; and Van de Sompel, H. 2005. Co-authorship networks in the digital library research community. Information processing & management 41(6):1462–1480.

[20] Liu, L.; Tang, J.; Han, J.; Jiang, M.; and Yang, S. 2010.

[21] Mining topic-level influence in heterogeneous networks. In Proceedings of the 19th ACM international conference on Information and knowledge management, CIKM ’10, 199– 208. New York, NY, USA: ACM.

[22] Mehmood, Y.; Barbieri, N.; Bonchi, F.; and Ukkonen, A. 2013. Csi: Community-level social influence analysis. In Joint European Conference on Machine Learning and Knowledge Discovery in Databases, 48–63. Springer.

[23] Probst, F.; Grosswiele, D.-K. L.; and Pfleger, D.-K. R. 2013. Who will lead and who will follow: Identifying influential users in online social networks. Business & Information Systems Engineering 5(3):179–193.

[24] Rohden, M.; Sorge, A.; Witthaut, D.; and Timme, M. 2014. Impact of network topology on synchrony of oscillatory power grids. Chaos: An Interdisciplinary Journal of Nonlinear Science 24(1):013123.

[25] Tang, J.; Zhang, J.; Yao, L.; Li, J.; Zhang, L.; and Su, Z. 2008. Arnetminer: extraction and mining of academic social networks. In Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, KDD ’08, 990–998. New York, NY, USA: ACM.

[26] Tang, J.; Zhang, D.; and Yao, L. 2007. Social network extraction of academic researchers. In Proceedings of the 7th IEEE International Conference on Data Mining, ICDM ’07, 292–301. IEEE.

[27] Trusov, M.; Bodapati, A. V.; and Bucklin, R. E. 2010. Determining influential users in internet social networks. Journal of Marketing Research 47(4):643–658.

[28] 2016. Us news ranking. http://www.usnews.com/ rankings.