Characterizing the Evolution of Collaboration Network

Bin Wu
School of Computer Science
Beijing University of Posts and Telecommunications
Beijing 100876, China
wubin@bupt.edu.cn

Fengying Zhao
School of Computer Science
Beijing University of Posts and Telecommunications
Beijing 100876, China
hyzhaolt@gmail.com

Shengqi Yang
School of Computer Science
Beijing University of Posts and Telecommunications
Beijing 100876, China
yangshengqi@tseg.org

Lijun Suo
School of Computer Science
Beijing University of Posts and Telecommunications
Beijing 100876, China
suolj@yahoo.com.cn

Hongjiao Tian
School of Computer Science
Beijing University of Posts and Telecommunications
Beijing 100876, China
thq2004423@gmail.com

ABSTRACT

It has long been realized that social network of scientific collaborations provides a window on patterns of collaboration within the academic community. Investigations and studies about static and dynamic properties of co-authorship network have also been done in the recent years. However, the accent of most of the research is on the analysis about the macroscopic structure of the whole network or community over time such as distance, diameter shrinking and densification phenomenon and microscopic formation analysis of links,groups or communities over time. But in fact, how an individual or a community grows over time may not only offer a new view point to mine copious and valuable information about scientific networks but also reveal important factors that influence the growth process. In this paper, from a temporal and microscopic analytical perspective, we propose a method to trace scientific individual's and community's growth process based on community's evolution path combination with quantifiable measurements. During the process of tracing, we find out that it is the fact that the lifespan of community is related to the ability of altering its membership, but what's more and complementary, we find out that the lifespan of community is also related to the ability of maintaining its core members meaning that community may last for a longer lifespan if its core members are much more stable. Meanwhile, we also trace the growth process of research individuals based on the evolution of communities.

Categories and Subject Descriptors

H.2.8 [**Database management**]: Database applications—*Data mining*

SWSM'09, November 6, 2009, Hong Kong, China.

General Terms

Measurement Theory

Keywords

Social Network Analysis, Collaboration Network, Evolution of Community

1. INTRODUCTION

A social network is a collection of people, each of whom is acquainted with some subset of the others [20]. Such a network can be represented as a set of vertices denoting people and links denoting relationship among people, such as acquaintance, friendship and so on. In recent years, researchers have discovered interesting features about social networks, for example, social networks have the so called small world phenomenon, follow a scale-free power-law distribution [7] and exhibit higher clustering coefficient. Obviously, all the properties or features about social networks are static and macroscopic, and a common question captures researchers' interest including: how an individual or a community grows up; what kind of individual and community may grow up healthily and quickly; and what kind of factor conducts the evolution process.All these questions make guidance for the study about dynamic mechanism and features of social network in a wide rang from theoretical aspect to real application. Known in theoretical aspect is the graph generator model [22, 26] to find out the dynamic mechanism that controls the growth of social network and prediction model to predict the formation of membership [6] and tendency of collaboration [12] between praise of scientists. By contrast, on the application aspect, [6, 17, 15, 23] discussed what factors may lay an important influence on the formation of links, groups and the joint in an existing group or community in online social networks. And recently, more and more researchers also start to investigate how social network change over time from the evolution or detection of graph [6, 11], community [25, 10, 11], group [22, 6] and affiliation [26] in a stream of snap graphs.

Specifically, co-authorship network(a typical social network of scientists)in which two scientists are considered connected if they have co-authored at least a paper [20] arouses researchers' sufficient concern. [20, 19, 21] and [22, 8] investigated the static structure and the evolution properties of co-authorship network respectively. However, they paid much more emphases on the macroscopic properties such as distance, clustering coefficient and general discussion about evolution of co-authorship network. And Amit Goyal in [4] and Aris Anagnostopoulos in [5] investigated the influence of individuals during the formation of links and communities from a microscopic perspective. In this paper, from a temporal and microscopic analytical perspective, we propose a method to describe the growth process of scientific communities and research individuals by tracing the evolution of community combination with a series of quantifiable measurement. Our contribution is as follows:

1. Construct collaboration network considering not only the growth but also the diminishment of researchers and co-authorship over time.

2. Offer a microscopic insight into the growth process of scientific community and research individual based on community's evolution path combination with attributes of community and its membership.

3. We find out a complementary phenomena: Communities persist longer if they are capable of maintaining their core members, suggesting that the stability of core membership results in a longer lifespan for community.

The rest of the paper is organized as follows: Section 2 reviews the related work. Section 3 gives symbols and definitions used in this paper. Section 4 introduces our method of collaboration network construction, establishment of evolution relationship between communities and the growth process of individuals, finally with statistics about features of communities and individuals during the process of evolution. Section 5 presents our findings based on our five data sets and Section 6 concludes.

2. RELATED WORK

Recently more and more researchers are interested in the study and utilization of academic social networks: Jie Tang et.al released their online academic researcher finding system ArnetMiner [2] with focus on academic researcher social network building, search, and mining;Theodoros Lappas proposed method in the latest research for finding a team of experts with special skills in academic social network[16] effectively. Both of the application and study show that the analysis of academic social network formed by researcher's co-authorship may bring us an effective and different perspective to analyze large and complex data sets in academia.

M.E. Newman in [20] presented a first look at the scientific collaboration network of scientists and revealed significant patterns about collaboration network such as short distance, higher clustering, power-law distribution and thus a small world phenomenon [21, 19]. Meanwhile A.L. Barabasi took a different but complementary approach with accent on dynamics that control the evolution and topology of collaboration networks [8] and investigated the evolution of community. What's more in detail, Jian Huang and Ziming Zhuang [12] conducted a longitudinal analysis of collaboration network in Computer Science at network, community and individual level respectively.

In respect to dynamic graph mining, Christos et al. in [13, 14] presented statistical properties of graph over time and a parameter-free method to trace the evolution of large graphs. Gergely Palla in [22], Tanya Y. Berger-Wolf and Jared Saia in [9] established evolution relationship between two communities based on overlapped nodes of communities in time t - 1 and in time t. Yi Wang et al. in [25] proposed a algorithm CommTracker based on the overlapped core nodes to establish evolution relationship between communities. Specifically, in our work, we analyze the growing process of communitiy and individual based on the evolution of community with help of algorithm CommTracker.

Above all, analysis about the evolution of collaboration networks may be put into three categories: structure analysis about collaboration networks [19] and communities [15]; formation of links, group/community [17, 18]; establishment of evolution relationship between communities in time step t and t -1 [9, 8].Complementary to the analysis about the evolution of community is the microscopic analysis of individual's influence on the others in social networks, such as identifying [4], measuring [5] and quantifying [24] the strength of individual's social influence respectively.Specifically, our work is similar to [9] and [8], but different from that, we establish evolution relationship between communities based on the overlap rate of core nodes at first, and then we characterize the evolution of collaboration networks with new methods such as active life span, core members' stability considering core members as backbone of community. Meanwhile we analyze the growth process of individuals with new quantifiable measurements such as active life span, activity, core rate etc. Furthermore, we investigate the relationship between community and its members based on these measurements.And finally, we find out that communities persist longer if they are capable of maintaining their core members,suggesting that the stability of core membership results in a longer lifespan for communities. And we'll show the phenomena from our data sets in Section 5.

3. PRELIMINARIES

3.1 Symbol & Description

In this section we formally introduce all the necessary symbols used in this paper in Table 1. And then give all the definitions of quantifiable measurements such as activity of authors,individual's active life span and core rate, community's member stability and core member stability etc.

3.2 Definitions

Activity of authors: $\forall a_i \in A$, we define the first time when the author a_i published a paper p_j in our data sets as the birth time of a_i denoting by $Birth(a_i)$. By analogy, $Death(a_i)$ is defined as the last time when the author a_i published a paper p_i in our data sets a_i. At any time $t \in [Birth(a_i), Death(a_i)]$, we argue that if the author a_i publishes a paper at t, then an collaboration event occurs and he/she participates in the event at that time. We define activity to describe the overall extent of activity of author a_i inspired by Barabasi's definition of the weight of link between two nodes in [22].

$$LV(a_j^{(t)}) = \sum_i Contr(p_i, a_j) * exp(-\lambda|t - t_i|Contr(p_i, a_j)) \tag{1}$$

Table 1: Symbols

Symbol	Description
$CN^{(t)}$	The network of collaboration at time t
A	The set of authors/researchers
$A^{(t)}$	The set of authors/researchers at time t
a_i	The author i
$L^{(t)}$	The set of co-authorship at time t
$l_{i,j}$	$\{(a_i, a_j)\|a_i \in A \wedge a_j \in A \wedge a_i, a_j$ are co-authors$\}$
p_i	The paper i
$c_i^{(t)}$	The community i in snapshot at time t
$C^{(t)}$	The set of community at time t
$c_i^{(t)} \rightarrow c_i^{(t+1)}$	$c_i^{(t+1)}$ is the successor of $c_i^{(t)}$
$c_i^{(t)} \Rightarrow c_i^{(t+1)}$	$c_i^{(t+1)}$ is the descendant of $c_i^{(t)}$
$\|c_i^{(t)}\|$	The size of c_i in snapshot at time t
$Edge(c_i^{(t)})$	The edge of c_i in snapshot at time t
$\|Edge(c_i^{(t)})\|$	The edge size of c_i in snapshot at time t
$Mem(c_i^{(t)})$	The members of c_i in snapshot at time t
$\|Mem(c_i^{(t)})\|$	The size of c_i at time t
$Core(c_i^{(t)})$	The cores of c_i in snapshot at time t
$\|Core(c_i^{(t)})\|$	The size of cores of c_i at time t

Specifically, there are three parts of $LV(a_j^{(t)})$:

1 $Contr(p_i, a_j)$: the contribution of a_j to paper p_i published at time t. We assume that if a paper has n co-authors, any of the co-author's contribution to the paper is $1/n$.

2 t_i: the nearest time point before t, and author a_j participates at least an event at t_i.

3 λ: a decay time characteristic/constants for different collaboration networks.

Obviously, when an author doesn't publish any papers after a specified time point, his/her life value may decay over time until he/she publishes new papers.

Collaboration network: In our study, we construct collaboration networks at each time step (one year). And each year's collaboration network is defined as:$CN^{(t)} =< A^{(t)}, L^{(t)} >$.Besides,$A^{(t)} = \{a|Birth(a) <= t \wedge Death(a) >= t\}$, $L^{(t)} = \{(a_i, a_j)|a_i \in A^{(t)} \wedge a_j \in A^{(t)}\}$ and a_i, a_j have co-authored at least one paper before or at the current time t.

Community's processor/successor in evolution process: $\forall c_i^{(t)}$, if $c_i^{(t)}$ and $c_i^{(t+1)}$ meet the following conditions:

$$Core(c_i^{(t)}) \cap Mem(c_i^{(t+1)}) \neq \phi \tag{2}$$

$$Core(c_i^{(t+1)}) \cap Mem(c_i^{(t)}) \neq \phi \tag{3}$$

$$\frac{Mem(c_i^{(t)}) \cap Mem(c_i^{(t+1)})}{Mem(c_i^{(t)}) \cup Mem(c_i^{(t+1)})} \geqslant \beta(\beta \in (0, 1]) \tag{4}$$

Then $c_i^{(t)}$ is defined as the processor of $c_i^{(t+1)}$, and $c_i^{(t+1)}$ is the successor of $c_i^{(t)}$, and denoting by $c_i^{(t)} \rightarrow c_i^{(t+1)}$. Furthermore, $c_i^{(t)}$ is the ancestor of $c_i^{(t+k)}$ (or $c_i^{(t+k)}$ is the descendant of $c_i^{(t)}$) if $c_i^{(t)} \rightarrow c_i^{(t+1)} \rightarrow ...c_i^{(t+k)}$. Specifically, if $\neg\exists c_i^{(t-1)}$ makes $c_i^{(t-1)} \rightarrow c_i^{(t)}$, then $c_i^{(t)}$ is a new born community at time step t denoting by $Birth(c_i^{(t)})$, and if $\neg\exists c_i^{(t+1)}$ makes $c_i^{(t)} \rightarrow c_i^{(t+1)}$, then $c_i^{(t)}$ is a dead community denoting by $Death(c_i^{(t)})$. Finally, a community $c_i^{(t)}$'s age may be calculated by $Death(c_i^{(t)}) - Birth(c_i^{(t)}) + 1$. Besides, the definition of core members and the method of core members detecting is described in [25].

Stability of Members and Core Members: $\forall c_i^{(t)} \in C^{(t)}$, we define the core member and member stability of $c_i^{(t)}$ at time $t+1$ compared to time t as $CS(c_i^{(t)})$ and $MS(c_i^{(t)})$,

$$CS(c_i^{(t)}) = \frac{|(Core(C_1^{t+1}) \cup ... \cup Core(C_k^{t+1})) \cap Core(C_i^t)|}{|Core(C_1^{t+1}) \cup ... \cup Core(C_k^{t+1}) \cup Core(C_i^t)|} \tag{5}$$

$$MS(c_i^{(t)}) = \frac{|(Mem(C_1^{t+1}) \cup ... \cup Mem(C_k^{t+1})) \cap Mem(C_i^t)|}{|Mem(C_1^{t+1}) \cup ... \cup Mem(C_k^{t+1}) \cup Mem(C_i^t)|} \tag{6}$$

and $c_i^t \rightarrow c^{(t+1)_j}$, $j \in [1, k]$, $k \geqslant 1$.

Finally, the change rate of members and core members is defined as $1 - MS(c_i^{(t)})$ and $1 - CS(c_i^{(t)})$ respectively.

Author's Activity Stability: The author's stability is defined based on the evolution relationship between communities at different time slice. When we trace the growth process of an author, we may find out all the communities in which he/she participates at each time slice during his/her life time, and then detect what kind of communities he/she participates in. So we define an author's activity stability as $AS(a_i^{(t)})$, and $AS(a_i^{(t)}) = min(\frac{C^{(t)} \cap C^{(t+1)}}{C^{(t)}}, \frac{C^{(t)} \cap C^{(t+1)}}{C^{(t+1)}})$. Specifically, $C^{(t)}$ and $C^{(t+1)}$ are the set of communities in which the author a_i participates in time t and $t+1$ respectively; $|C^{(t)} \cap C^{(t+1)}|$ is the count of community $c_i^{(t)}$ ($c_i^{(t)} \in C^{(t)} \wedge \exists c_i^{(t+1)} \in C^{(t+1)} \wedge c_i^{(t)} \rightarrow c_i^{(t+1)}$).

Author's Core Rate: During the life time of an author, he/she may participate in many communities, and may be the core member of a community or may not. We define the author's core rate as $ACR(a_t^{(t)})$. $ACR(a_t^{(t)}) = \frac{|C_c^{(t)}|}{|C_m^{(t)}|}$, specifically, $C_c^{(t)} = \{c_j^{(t)}|a_i \in Core(c_j^{(t)})\}$ and $C_m^{(t)} = \{c_j^{(t)}|a_i \in Mem(c_j^{(t)}) \wedge a_i \notin Core(c_j^{(t)})\}$.

4. EVOLUTION OF CO-AUTHORSHIP NETWORK

In this section, our work is conducted by the following steps:

1 Construct each year's collaboration network according to the co-authorship among researchers and the birth and death of individuals.

2 Utilize community detection algorithm based on clique to find out communities of each year.

3 For a given community in a specified year, we trace the evolution of the community forwards and backwards based on its core membership and make statistics such as member stability, core stability, active life span etc.

4 Find out and analyze essential factors that influence the growth of community with experiment results in five data sets showed in Table 2.

4.1 Evolution of Communities

In this section, we focus on the analysis of the evolution of community including the detection of community in each time step, establishment of evolution relationship between communities in different time steps, statistics about features of community such as $MS(c_i^{(t)})$, $CS(c_i^{(t)})$ defined in section 3.2 and average active life span of members and core members during evolution of community, and finally with comparison and analysis of factors which are related to the lifespan of community.

4.1.1 Community detection and emerging

As to the detection of community, we utilize algorithm based on clique at each time step. Then we use the algorithm CoreDetecion [25] to calculate centricity of every member in community. Based on that, we investigate that many communities consists of less than five members and with an obvious core member, so we emerge these small community into another community as follows: $\forall c_i^{(t)}, c_j^{(t)} \in C^{(t)}$, if $c_i^{(t)}, c_j^{(t)}$ meet at least one of the conditions: (1) $Core(c_i^{(t)}) \subseteq Core(c_j^{(t)})$; (2) $Core(c_j^{(t)}) \subseteq Core(c_i^{(t)})$; (3) $\frac{|Core(c_i^{(t)}) \cap Core(c_j^{(t)})|}{|Core(c_i^{(t)}) \cup Core(c_j^{(t)})|} \geq 0.5$.

4.1.2 Establishment of evolution relationship

With the information of members and core members of communities in each time step, we start to establish the evolution path forwards or backwards based on the algorithm CommTracker proposed in [25]. Besides we add a complementary constrain about overlap rate of nodes between two communities $c_i^{(t)}$ and $c_i^{(t+1)}$ just as defined in Section 3.1.2. Specifically, we argue that $c_i^{(t)}$ is a new born community if it doesn't have any predecessor, by analogy, $c_i^{(t)}$ is a dead one if it doesn't have any successor. So as to a given community $c_i^{(t)}$ we may trace its ancestor $\{C^{(t-k)}, C^{(t-k+1)}, ..., C^{(t-1)}\}$ recursively until $\forall c_i^{(t-k)}$, $c_i^{(t-k)} \in C^{(t-k)}$, $c_i^{(t-k)}$ doesn't have any predecessor. Also, we can also trace $c_i^{(t)}$'s descendant $\{C^{(t+1)}, C^{(t+2)}, ..., C^{(t+h)}\}$ recursively, and define $EVOL(c_i^{(t)}) = \{C^{(t-k)}, ..., C^{(t-1)}, c_i^{(t)}, C^{(t+1)}, ..., C^{(t+h)}\}$ as the set of community at each time during the evolution of community $c_i^{(t)}$.

4.1.3 Statistics about features of community in evolution path

In this section, we aim to investigate important factors that influence the growth of community. Under our construction of collaboration network considering both the birth and death of individuals, we also find that community which lasts for a long lifespan has a strong ability to alter its members just as described in [22]. Based on that and inspired by that, we wonder to investigate what's the mechanism that conducts the growth of community even though with many members' participation in and leaving off. Based on statistics and observation of our data sets, we find out that community's lifespan is also related to the ability of maintaining its core members, and a higher stability of core members may result in a longer lifespan of community.

Active life span of member. Similar to birth, death, life span of an individual relative to the global collaboration network, we define a local active time span of author a_i relative to community $c_j^{(t)}$ denoted by $ALS(a_i, c_j^{(t)})$, and $ALS(a_i, c_j^{(t)}) = Death(a_i, c_j^{(t)}) - Birth(a_i, c_j^{(t)}) + 1$, and $Birth(a_i, c_j^{(t)})/Death(a_i, c_j^{(t)})$: the first/last time when individual participates in at least one of $c_j^{(t)}$'s ancestor/descendant.

Table 2: Data Sets

Data Sets	Time Period	Author Num	Max Link Size
DBLP	1959 ∼ 2008	516,113	484,713
BUPT	1998 ∼ 2004	1,847	3,074
WanFang	1987 ∼ 2007	108,383	65,794
Cornell	1993 ∼ 2006	58,212	58,757
Cond-mat	1993 ∼ 2006	68,612	52,738

Then we put all the members of communities appear in the evolution of $c_j^{(t)}$ into three categories, new members, all members and core members in each time step.

$$A_{mem} = \bigcup_{t''} Mem(C^{(t'')}) \quad (7)$$

$$A_{new} = \bigcup_{t''} NewMem(C^{(t'')}) \quad (8)$$

$$A_{core} = \bigcup_{t''} Core(C^{(t'')}) \quad (9)$$

Specifically, $t \in$ [Birth of community $c_j^{(t)}$, Death of community $c_j^{(t)}$], $C^{t''} \in EVOL(c_j^{(t)})$ and $NewMem(C^{t''}) = \{a_i | a_i \in Mem(C^{t''}) \wedge Birth(a_i, c_j^{(t'')}) = t'' \wedge c_j^{(t'')} \in C^{t''}\}$. And we define $AVG(ALS(A, c_j^{(t)}))$ as the average active time span of all individuals $a_i (a_i \in A)$. Finally, we find out that $AVG(ALS(A_{new}, c_j^{(t)})) < AVG(ALS(A_{mem}, c_j^{(t)})) < AVG(ALS(A_{core}, c_j^{(t)}))$. It shows that, it is the new born members that result in the altering membership of the original community, but most of the new born ones have a short active time span, by contrast, the core members have a longer active time span during the evolution of community.

But from a overall perspective, we calculate the distribution of all the members' age, and we find that it's most similar to the pow-law distribution indicating that most of the member may die or leave off the community, especially the new members. Based on this, we continue to investigate the other factors that influence the growth of community.

4.2 Evolution of Individuals

For a given researcher, we find out all the communities in which he/she participates in each year combination with the activity of the researcher and describe the researcher's growth with three different phrases: turning up and growing, growing stronger and diminishing. Obvious behavior of an individual in the first phrase is preferential attachment to other individuals who have possessed much more links and not being a member of any community or being member of very few communities without being core membership. After that the individual may have a try to participate in other communities and become core member of these communities and in some extent recruits more and more new members into the community to which he/she already belongs.

Accompany with the overall investigation of an individual's growth process, we propose activity $LV(a_i^{(t)})$ to define the extent of activity of an individual over time. Compared to traditional statistics about the output of a researcher, the utilization of $LV(a_i^{(t)})$ to describe researcher's activity manifests the following merits:

1 $LV(a_i^{(t)})$ is a function taking three factors into account: output, contribution to a paper and time-the most important one in evolution analysis of individual.

2 Decay constants λ with time: with the help of λ, we may clearly investigate the extent of activity of a researcher decaying if he/she doesn't participate in event during a period of time.

Finally we find that only these individuals that experience the 'growing stronger' phrases can they have a longer lifespan.

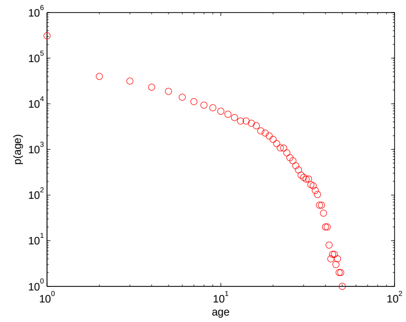
5. EXPERIMENTS AND RESULTS

5.1 Data Sets

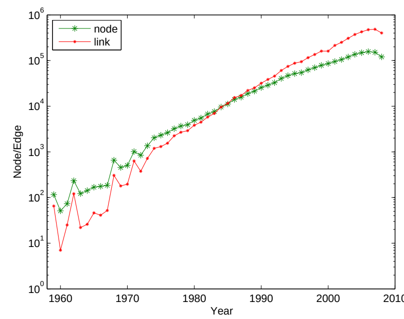
In our study, we collect five different data sets of academic collaboration shown in Table 2. Specifically, *DBLP*, *Cornell*, *Cond-mat* data sets are public data sets about collaboration networks obtained from [1] and [3] respectively. The data set of *BUPT* is obtained according to the published papers from 1998 to 2004 indexed by *SCI*, *EI* and *ISTP* in Beijing University of Posts and Telecommunications. And *WanFang* is about the co-authorship between researchers in Life Science area from 1987 to 2007.

5.2 Construction of collaboration networks

According to the definition of an individual's birth and death, we calculate the age of each individual. And after the death of an individual we delete all the links which links to it. Just as displayed in Figure 1(a), we discovered that the age distribution of individuals follows a power-law distribution which indicates that most of the individual have a short life span and only a few have a longer one. So from this temporal perspective, we may take both of the growing and diminishing of individuals and co-authorship. Figure 1(b) shows the collaboration network of DBLP from 1959 to 2008. And we may see that the collaboration network's individuals and links are not always growing but diminishing in some time (before 1975) which may reflect the evolution process of community in a much more practical way.



(a) Individual's age distribution in DBLP



(b) Collaboration network of DBLP

Figure 1: Construction of collaboration networks

5.3 Community Evolution Analysis

5.3.1 *Community Age Distribution*

After the construction of collaboration network we detect community in each year and establish the evolution relationship between two communities in time t and t +1. Before the analysis of features about community's member, we may investigate the age distribution of community during evolution to get an overall insight.

And we find out that most of the communities have short life spans, and a few with much longer ones just as described in Figure 2. And inspired by this, we wonder why some community may last a longer life span and some not, and in Section 5.3.2 and 5.3.3 we may address the issue and show our findings.

5.3.2 *Active life span analysis*

From Figure 2 we may see the age distribution of communities clearly. And from this section we start to investigate the factors which influence the evolution of community. Based on the evolution path of each community in a given year, we make statistics about the active life span of new born members and core members. And we find out that most of the new born members have a much shorter active life span which indicates that new born members aren't stabile during the evolution of community. By contrast, the average active life span of core members is much longer than the new ones indicating the stability of core members may play a much more important role leading community lasting for a long life span.

5.3.3 *Changing Rate of members and core members in community evolution*

In order to find out the factors that influence the evolution of a community, we investigate the change rate of members and core members respectively with combination of community's age. Finally, we find out that communities may last for a long life span with strong ability of altering its membership and maintaining its core members just as shown in Figure 4. We can see clearly that the change rate of members are obviously high if a community's life span is longer. By contrast,the change rate of core members shows an opposite influence, that is ,stable core member composition may result in a longer life span of community.

Our explanation of the phenomenon is as follows: in scientific collaboration network, the main research group is formed in research institutions of colleges, universities and

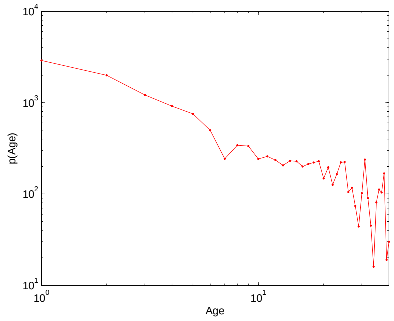


Figure 2: Age Distribution of Community in 2008 of DBLP

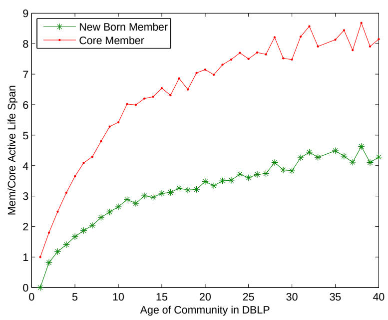


Figure 3: Community age and member's active life span

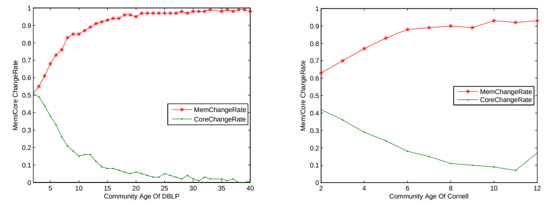
industrial enterprise; at the establishment or birth of these researcher institutions to which few people are familiar, so displays lower new member recruiting ability; meanwhile higher changing rate of core members at the beginning indicates that core or authoritative researchers play a leading role in both establishment of research institutions and a new research area. During the growth of research groups, more and more students or new researchers may participate in, but students may graduate from colleges and new researchers may leave off and continue to try other research area, but the original research groups or institutions are still existing with a stable core members indicating a stable research area.Meanwhile,the publications also exist in the original research institutions.

5.4 Establishment of individual's lifecycle

In this section, we illustrate the growing process of individuals in terms of active life time in a microscopic perspective and investigate the changing of his/her participation in communities. Here we may show several quantifiable measurements, such as activity, change rate and core rate in different communities, to show the growth process of an individual with three different phrases.

From Figure 5 we see individual's growth process with individual's activity, change rate and core rate in red, blue and green colored line respectively. From both of the Figure 5(a) and Figure 5(b) we may find out that the activity of individual is lower at the birth and sometime after that time and isn't core member of any community shown by zero value of individual's core rate(green line).But as the rise of activity,the new born individual may become one of the core members indicating that the individual has entered into a growing stronger phrase. And in this phrase, the individual becomes more and more active and the core rate is obvious high and start to change over time. Meanwhile, the change rate of individual indicates individual's new participation in new communities.

Finally, from Figure 5(b) we may see after 2000 year, the active of individual decay over time indicating the third growth phrase - diminishing.



(a) DBLP (b) Cornell

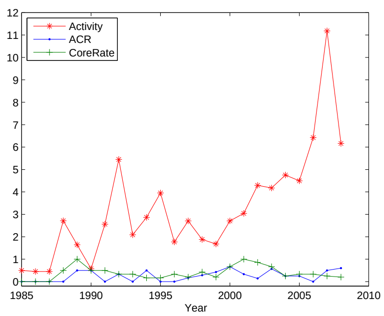
(c) Cond-mat (d) WanFang

Figure 4: Change Rate of Members and Core Members

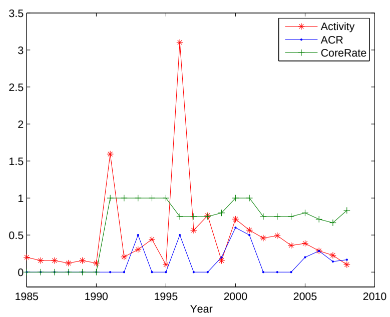
6. CONCLUSIONS

In this paper, from a temporal and microscopic perspective, we construct the collaboration network considering both the growth and diminishment of individuals and co-authorship. And then detect all communities at each time step and establish evolution relationship between communities at different time step. Meanwhile, we propose a quantifiable measurement - activity of individuals to describe his/her extent of activity overall his/her life time with three import phrase of growing up-birth and growing, growing up stronger and diminishing based on the analysis of evolution of community. What's more important is that core members of community play an important role conducting the growth of community with strong stability and a much longer active life span during evolution of community.

In the future work, our aims may focus on a much more microscopic analysis of growth process both on community and individual including: modeling the evolution (growth process) of community, such as establishment of community's life cycle, important turning point detection with network's structure features and individual's attributes; evaluating the quality and predicting future growth phrase of community; analyzing individual's influence on his/her friends based on individual growth process.Specifically, we may try more quantitative measurements to characterize the evolution of communities and individuals and to investigate whether the dynamical processes of scientific research and collaboration share some common characteristics.



(a) Author A



(b) Author B

Figure 5: Activity of individuals in DBLP

7. REFERENCES

[1] http://kdl.cs.umass.edu/data/dblp/dblp-info.html.

[2] Arnetminer. http://www.arnetminer.org/.

[3] Newman network data. http://www-personal.umich.edu/ mejn/netdata/.

[4] F. B. Amit Goyal and L. V. S. Lakshmanan. Discovering leaders from community actions. In *CIKM '08: Proceeding of the 17th ACM conference on Information and knowledge management*, pages 499–508, 2008.

[5] R. K. Aris Anagnostopoulos and M. Mahdian. Infuence and correlation in social networks. In *KDD'08: Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 7–15, 2008.

[6] L. Backstrom, D. Huttenlocher, J. Kleinberg, and X. Lan. Group formation in large social networks: membership, growth, and evolution. In *KDD '06: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 44–54, 2006.

[7] A. L. Barabasi and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, October 1999.

[8] A. L. Barabasi, H. Jeong, Z. Neda, E. Ravasz, A. Schubert, and T. Vicsek. Evolution of the social network of scientific collaborations. *Physica A: Statistical Mechanics and its Applications*, 311(3-4):590–614, August 2002.

[9] T. Y. Berger-Wolf and J. Saia. A framework for analysis of dynamic social networks. In *KDD '06: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 523–528, 2006.

[10] J. Hopcroft, O. Khan, B. Kulis, and B. Selman. Natural communities in large linked networks. In *KDD '03: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 541–546, 2003.

[11] J. Hopcroft, O. Khan, B. Kulis, and B. Selman. Tracking evolving communities in large linked networks. *Proc Natl Acad Sci U S A*, 101 Suppl 1:5249–5253, April 2004.

[12] J. Huang, Z. Zhuang, J. Li, and C. L. Giles. Collaboration over time: characterizing and modeling network evolution. In *WSDM '08: Proceedings of the international conference on Web search and web data mining*, pages 107–116, 2009.

[13] S. P. Jimeng Sun, Christos Faloutsos and P. S. Yu. Graphscope: parameter-free mining of large time-evolving graphs. In *KDD '07: Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 687–696, 2007.

[14] J. K. Jure Leskovec and C. Faloutsos. Graphs over time: densication laws, shrinking diameters and possible explanations. In *KDD '05: Proceeding of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*, pages 177–187, 2005.

[15] R. Kumar, J. Novak, and A. Tomkins. Structure and evolution of online social networks. In *KDD '06: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 611–617, 2006.

[16] T. Lappas, K. Liu, and E. Terzi. Finding a team of experts in social networks. In *KDD '09: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 467–476, 2009.

[17] J. Leskovec, L. Backstrom, R. Kumar, and A. Tomkins. Microscopic evolution of social networks. In *KDD '08: Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 462–470, 2008.

[18] L. Licamele and L. Getoor. Social captial in friendship-event networks. In *ICDM'06:Sixth IEEE International Conference on Data Mining.*, pages 959–964, 2006.

[19] M. Newman. Who is the best connected scientist?a study of scientific coauthorship networks. *Complex Networks*, pages 337–370, January 2004.

[20] M. E. Newman. The structure of scientific collaboration networks. *Proc Natl Acad Sci U S A*, 98(2):404–409, January 2001.

[21] M. E. Newman. Coauthorship networks and patterns of scientific collaboration. *Proc Natl Acad Sci U S A*, 101 Suppl 1:5200–5205, April 2004.

[22] G. Palla, A.-L. Barabasi, and T. Vicsek. Quantifying social group evolution. *Nature*, 446(7136):664–667, April 2007.

[23] M. A. Serrano, M. Boguñá, and A. Vespignani.

Extracting the multiscale backbone of complex weighted networks. *Proc Natl Acad Sci U S A*, 106(16):6483–6488, April 2009.

[24] J. Tang and Z. Yang. Social influence analysis in large-scale social networks. In *KDD'09:Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 807–816, 2009.

[25] Y. Wang, B. Wu, and X. Pei. Commtracker: A core-based algorithm of tracking community evolution. In *ADMA '08: Proceedings of the 4th international conference on Advanced Data Mining and Applications*, pages 229–240, 2008.

[26] E. Zheleva, H. Sharara, and L. Getoor. Co-evolution of social and affiliation networks. In *KDD '09: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1007–1016, 2009.