2009

# Discovery of core-nodes in event-based social networks

Shaojie Yuan
*University of Wollongong*, sy395@uow.edu.au

Quan Bai
*University of Wollongong*, quan@uow.edu.au

Minjie Zhang
*University of Wollongong*, minjie@uow.edu.au

Khin Than Win
*University of Wollongong*, win@uow.edu.au

### Abstract

Most previous actor-node ranking algorithms for event-based social networks only consider how many events an actor participates in. However in event-based social networks, we should also consider the influence of events when we rank actor-nodes. In this paper we formally define event-based social networks and related concepts, then we propose rules to construct an event-based social network. Algorithms are presented to discover the activity and importance of each actor-node. We test the algorithms by analysing the DBLP data set. In the experiment actors in DBLP data set are ranked based on their activity, importance, and combination of activity and importance, respectively.

2009 Sixth International Conference on Fuzzy Systems and Knowledge Discovery

# Discovery of Core-Nodes in Event-Based Social Networks

Shaojie Yuan, Quan Bai, Minjie Zhang and Khin Than Win
*School of Computer Science & Software Engineering*
*University of Wollongong*
*Wollongong, Australia*
*email: {sy395, quan, minjie, win}@uow.edu.au*

***Abstract*—Most previous actor-node ranking algorithms for event-based social networks only consider how many events an actor participates in. However in event-based social networks, we should also consider the influence of events when we rank actor-nodes. In this paper we formally define event-based social networks and related concepts, then we propose rules to construct an event-based social network. Algorithms are presented to discover the activity and importance of each actor-node. We test the algorithms by analysing the DBLP data set. In the experiment actors in DBLP data set are ranked based on their activity, importance, and combination of activity and importance, respectively.**

***Keywords*-Knowledge Discovery, Event-Based Social Network, Data Mining, Ranking.**

## I. Introduction

In recent years, knowledge discovery and data analysis in dynamic social networks are becoming important topics in various application domains, including law enforcement, economic analysis and disease control. Researchers from different areas, such as sociology, mathematics and computer science, try to use various techniques to explore useful knowledge and reveal hidden patterns in dynamic social networks. Event-based social networks are a category of dynamic social networks. All the event-based social networks have a common feature that the events in the social network are immanently temporal. In other words, all the events have a time stamp to indicate when the events have happened [1]. So how to represent nodes and events, how to represent the relationship between the nodes and events and how to construct an event-based social network for efficient discovery important nodes become significant issues.

The motivation of this research is to find a method to construct a generic event-based social network for efficient knowledge discovery. This model should be able to not only analyse actors in the network, but also analyse events. Furthermore, the model could also be used to analyse relationships between actors and events.

The rest of this paper is arranged as follows. Formal definitions and network construction rules are proposed in Section II. In Section III, algorithms for evaluating importance and activity of actor-nodes and influence of event-nodes are introduced. A case study is presented in Section IV. In the case study, we use the DBLP data set as an example to demonstrate the use of the proposed methods. In Section V, related work and discussion are presented to highlight the motivation of this research. Finally, the paper is concluded and the future work is outlined in Section VI.

## II. Event-Based Social Network Structure

In this paper, we put forward an assumption that all network changes are caused by events in an Event-Based Social Network (*EBSN*). An actor in an *EBSN* may have a new relationship when an event happens. For example, Mark is Marry's husband after their marriage. The relationship between the two actors (i.e. Mark and Mary) is established after the event (i.e. get married) happens. In an *EBSN*, an event could create a (set of) new relationship(s) or change a (set of) relationship(s). Formal definitions of *EBSN*, *actor-node*s, *event-node*s, $lbAE$s, $lbEE$s and *core-node*s are given in Subsection II-A.

### A. Definitions in an Event-Based Social Network Structure

**Definition 1** *An **Event-Based Social Network (EBSN)** is a bipartite directed graph. It is defined by a two tuple* $EBSN = (N, C)$, *where*

- $N = \{n_1, ..., n_s\}$ is a finite set which contains all the vertices in the network.
- $C = \{c_1, ..., c_r\}$ is a finite set which contains all the arcs in the network, where $c_i = (n_j, n_k)$ with $n_j,\ \ n_k \in N$.

An $EBSN$ is a bipartite graph as there are two types of nodes in the graph, i.e. actor-node and event-node.

**Definition 2** *An **actor-node** $a_m$ is a node which represents an actor in an EBSN. It can be defined by a three-tuple,* $a_m = (ID, activity, importance)$ *where*

- *ID* is a unique string which identifies different actor-nodes.
- *Activity* shows how active an actor-node is in a particular domain. It is defined by a four-tuple, $activity = (value, domain, time_0, time_1)$ where
  - *value* is a non-negative number which indicates the value of activity.
  - *domain* is a string that presents a specified field in which the activity is.
  - $time_0$ and $time_1$ are time tags which show the activity in period from $time_0$ to $time_1$.

978-0-7695-3735-1/09 $25.00 © 2009 IEEE
DOI 10.1109/FSKD.2009.335

- *importance* indicates an actor's influence in an *EBSN*. It is defined by a four-tuple, $importance = (value, domain, time_0, time_1)$, where
  - *value* is a non-negative number which expresses the value of importance.
  - *domain* is a string that presents a specified field in which the importance is.
  - $time_0$ and $time_1$ are time tags which show the importance in period between $time_0$ and $time_1$.

**Definition 3** *An* ***event-node*** $e_q$ *is a node which represents an event in an EBSN. It can be defined by a four-tuple,* $e_q = (ID, influence, domain, time)$*, where*

- *ID* is a unique string which identifies event-nodes,
- *influence* is a non-negative number which expresses the value of influence.
- *domain* is a string that presents to which field the event belongs.
- *time* is a time tag indicates when this event has happened.

**Definition 4** *A* ***lbAE (link between Actor-Node and Event-Node)*** *is a directed arc which can be defined by a two-tuple,* $lbAE = (a_i, e_j)$ *where*

- $a_i \in A\,(0 < i)$ where $A$ is a finite node set which contains all actor-nodes of an *EBSN*.
- $e_j \in E\,(0 < j)$ where $E$ is a finite node set which contains all event-nodes of an *EBSN*.

A lbAE represents an arc which connects an actor-node and an event-node in an *EBSN*. It indicates that actor $a_i$ participates in event $e_j$. We use LBAE to represent a set of lbAE which contains all the lbAE of an *EBSN*.

**Definition 5** A ***lbEE (link between Event-Node and Event-Node)*** *is another type of directed arcs which can be defined by a two-tuple,* $lbEE = (e_i, e_j)$ *where*

- $e_i,\ \ e_j \in E$; $e_i.time < e_j.time$ and $0 < i, 0 < j$

A lbEE connects two *event-nodes* in an *EBSN*. It represents that an event $e_j$ is affected by another event $e_i$. It directs from $e_i$ to $e_j$. We use LBEE to represent a set of lbEE which is composed of all the lbEE of an *EBSN*.

**Definition 6** *A* ***core-node*** $a_n$ *is an actor-node which has* $\left(a_n.activity.value - \overline{A.activity.value} > 0\right) \wedge \left(a_n.importance.value - \overline{A.importance.value} > 0\right)$ *where*

- $\overline{A.activity.value} = \frac{\sum_{i=1}^{\#(A)} a_i.activity.value}{\#(A)}$ where $\#(A)$ equals the number of elements in set $A$.
- $\overline{A.importance.value} = \frac{\sum_{i=1}^{\#(A)} a_i.importance.value}{\#(A)}$ where $\#(A)$ equals the number of elements in set $A$.

A core-node describes an actor-node which has higher value of both activity and importance than the average values. In other words, all core-nodes are active and important.

### B. Rules to Construct an EBSN

Most previous social network visualization approaches [2] only treat actors as nodes in a social network. In our *EBSN*, we treat actors and events as two types of nodes in the network to make the model carry more information. Figure 1 shows a simple *EBSN*. It can be seen in Figure 1 that there are two types of nodes, *actor-node*s and *event-node*s (refer to Definition 2 and Definition 3), and two types of arcs, *lbAE*s and *lbEE*s (refer to Definition 4 and Definition 5).

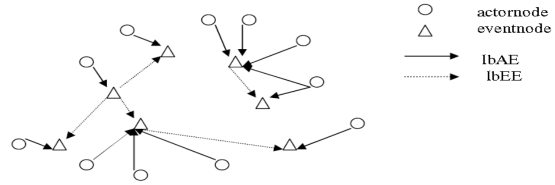



Figure 1. Event-Based Social Network

In this research, we define a set of rules to construct an *EBSN*.

**Rule 1:** An *actor-node* could only directly connect to an *event-node(s)*.

In an *EBSN*, an *actor-node* could connect to an *event-node(s)* via *lbAE(s)*. In an event-based social network, people might have a number of types of relations with others, but the relations among them do not exist at the beginning. In this paper, the interactions are events which are represented by *event-node*s. The relations are created by interactions. Rule 1 only allows an actor-node to connect to an *event-node(s)*. For example, Bob does not know Winly until an event happened, e.g. both of them attended a same conference.

**Rule 2:** The relation between an *actor-node* and an *event-node* could only direct from an *actor-node* to an *event-node*. In other words, there are only unidirectional relations in an *EBSN*.

The reason behind the rule is that *actor-node*s could choose whether or not to participate in an event-node.

**Rule 3:** An *event-node* $e_m$ may affect another *event-node* $e_n$, if so, then $(e_m, e_n) \in LBEE$, and $e_m.time < e_n.time$.

**Rule 4:** An *event-node* has to be connected by at least one *actor-node*.

This can be explained as each event happens because of involvement of at least one actor.

## III. Algorithms to Discover Core-Nodes

To discover *core-node*s in an *EBSN*, we need to rank *actor-node*s based on their activity and importance. The next subsection introduces algorithms to discover the values of the activity and importance of an *actor-node*.

### A. Knowledge Discovery Based on Actor-Nodes

In this section, based on previous definitions we propose an algorithm to discover the activity and importance of an *actor-node*.

1) *An algorithm to discover the activity of an actor-node.*

Intuitively, the value of *activity* should be proportionally the number of *event-node*s that the *actor-node* connecting to via $lbAE$ between $t_0$ and $t_1$. Equation 1 shows the formula to calculate *activity.value* of *actor-node* $a_m$ in $domain_0$.

$$\begin{cases} a_m.activity.value = \#\left(E_n\right) \\ a_m.activity.domain = domain_0 \\ a_m.activity.time_0 = t_0 \\ a_m.activity.time_1 = t_1 \end{cases} \quad (1)$$

In Equation 1, $\#\left(E_n\right)$ is the number of elements in set $E_n$. $E_n$ can be obtained according to Equation 2.

$$\begin{aligned} E_n &= \{\forall e_i \in E_n | e_i.time > t_0 \wedge e_i.time < t_1 \wedge \\ &\quad (a_m, e_i) \in LBAE \wedge e_i.domain = domain_0\} \end{aligned} \quad (2)$$

2) *An algorithm to discover the importance of an actor-node*.
The importance value relates to the influence of every *event-node*s that the *actor-node* connects with. Equation 3 introduces how to calculate *importance* of *actor-node* $a_m$.

$$\begin{cases} a_m.importance.value = \sum_{e_i \in E_m} \alpha_i \bullet \\ \qquad e_i.influence \\ a_m.importance.domain = domain_0 \end{cases} \quad (3)$$

In Equation 3, $\alpha_i$ is the coefficient of $e_i.influence$ and $\alpha_i \in [0, 1]$, $E_m$ is a set of events. The details of the set $E_m$ are demonstrated in Equation 4.

$$\begin{aligned} E_m = \{\forall e_i \in E_m | \left(a_m, e_i\right) \in LBAE \wedge \\ e_i.domain = domain_0\} \end{aligned} \quad (4)$$

## B. Knowledge Discovery Based on Event-Nodes

An *event-node* $e_q$ can be represented by a four-tuple, $e_q = (ID, influence, domain, time)$ where $influence$ is a non-negative number indicating the capability of influence in $e_q.domain$. Equation 5 exhibits how to calculate the value of *influence*.

$$\begin{cases} e_q.influence = \beta_0 \bullet \#\left(E_m\right) + \beta_1 \bullet \#\left(A_n\right) + \\ \qquad initialvalue \\ e_q.domain = domain_0 \end{cases} \quad (5)$$

In Equation 5, $initialvalue$ is a constant, $\beta_0$ is a coefficient of $\#\left(E_m\right)$ and $\beta_1$ is a coefficient of $\#\left(A_n\right)$. The value of $initialvalue$ could vary due to different domain areas, and $\beta_0, \beta_1 \in [0, 1]$. $E_m$ is a set of *event-node*s which introduced in Equation 6 and $A_n$ is a set of *actor-node*s which is elaborated in Equation 7, respectively.

$$\begin{aligned} E_m = \{\forall e_i \in E_m | \left(e_i, e_q\right) \in LBEE \wedge \\ e_i.domain = e_q.domain\} \end{aligned} \quad (6)$$

$$A_n = \{\forall a_i \in A_n | \left(a_i, e_q\right) \in LBAE\} \quad (7)$$

If an event $e_r$ happens after $e_q$ and is affected by $e_q$, we need to update the value of $e_q.\left(influence, domain_0\right)$. We can use $e_q.\left(influence', domain_0\right)$ to represent the updated value. Equation 8 shows how to calculate the updated value.

$$\begin{cases} e_q.influence' = e_q.influence + \beta_3 \bullet e_r.influence \\ e_q.domain = e_r.domain \end{cases} \quad (8)$$

In Equation 8, $\beta_3 \in [0, 1]$ and $\left(e_q, e_r\right) \in LBEE$.

# IV. Case Study

In this section, the DBLP data set is analysed. The most active actors, most important actors and top core actors in the *EBSN* are discovered.

## A. Problem Description

The DBLP data set contains information about authors, their publications (include publication title, author/authors, editor/editors, publish year, publisher), and citations. The information could be used to construct an academic *EBSN*.

An academic *EBSN* is a bipartite directed graph which includes authors, publications, authorships and citations. In the *EBSN*, let an *actor-node* represents an author where $a_m.ID =$ author's name. A publication could be represented by an *event-node* where $e_q.time =$ publish year. If an author $a_m$ is the author or editor of the publication $e_q$ then $lbAE = \left(a_m, e_q\right) \in LBAE$. If a paper $e_i$ is cited by another paper $e_j$ then $lbEE = \left(e_i, e_j\right) \in LBEE$. The experiment is introduced in the next subsection.

## B. Experiment Setup

In this experiment, a DBLP data set which contains 658132 authors and 1087683 publications was downloaded from http://dblp.uni-trier.de/xml/. The time range of the data set is from year 1936 to year 2009. Most of the publication topics are related to 'Database'. In the experiment we set both the *event-nodes* and *actor-nodes* $domain = database$. We use *actor-nodes* to represent authors where *actor.ID*=author's name, *actor.activity.value*=*actor-node*'s publications. The *importance* of an *actor* could be calculated using Equation 3 where we set $\alpha_i = 1$.

As we explained before, the *event-node*s represent publications. The *influence* of an *event-node* could be calculated using Equation 5. In this case, we set $\beta_0 = 1$, $\beta_1 = 0$ and $initialvalue = 1$. In other words, we believe the publication *influence* is related to the cited number but not the co-authors publications.

The earliest publication in the DBLP data set is published in 1936 and the latest publication in DBLP data set is published in 2009. In order to simplify the process, we separate the publication time into seven periods. The first period starts from year 1936 and ends at year 1945. After that, every ten years is a period until year 2005. Since there

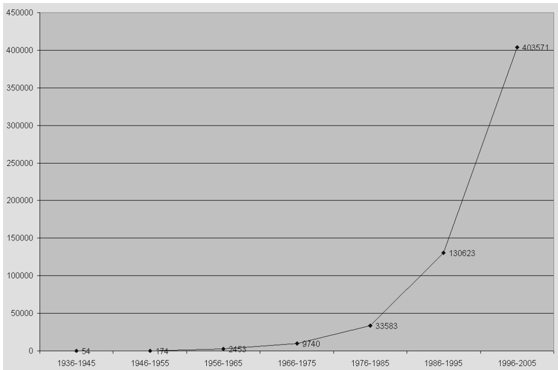



Figure 2. The numbers of authors in DBLP in different periods

| Year | First | Second | Third |
|---|---|---|---|
| 1996 | A. L. S.-Vincentelli | K. G. Shin | R. Tamassia |
| 1997 | K. G. Shin | S. Jajodia | J. R. Hancock |
| 1998 | M. Potkonjak | H. Garcia-Molina | I. Pomeranz |
| 1999 | E. R. Hancock | A. N. Choudhary | M. Potkonjak |
| 2000 | B. Hancock | T. S. Huang | E. R. Hancock |
| 2001 | H. Zhang | T. S. Huang | H. Yan |
| 2002 | M. T. Kandemir | E. R. Hancock | S. K. Das |
| 2003 | H. Zhang | W. Li | H. Seidel |
| 2004 | W. Gao | H. Zhang | M. Li |
| 2005 | M. T. Kandemir | W. Gao | L. Jiao |

Table I
TOP THREE ACTIVE AUTHORS IN DIFFERENT YEARS

| Year | First | Second | Third |
|---|---|---|---|
| 1996 | A. L. S.-Vincentelli | K. G. Shin | S. M. Reddy |
| 1997 | K. G. Shin | E. R. Hancock | E. Bertino |
| 1998 | P. S. Yu | T. S. Huang | H. Garcia-Molina |
| 1999 | E. R. Hancock | T. S. Huang | E. Bertino |
| 2000 | B. Hancock | T. S. Huang | E. R. Hancock |
| 2001 | T. S. Huang | Ho. Zhang | E. R. Hancock |
| 2002 | Ma.t T. Kandemir | S. M. Reddy | E. R. Hancock |
| 2003 | W. Li | M. T. Kandemir | H. Seidel |
| 2004 | W. Wang | W. Gao | C. Chang |
| 2005 | C. Chang | M. T. Kandemir | P. S. Yu |

Table II
TOP THREE IMPORTANT AUTHORS IN DIFFERENT YEARS

| Year | First | Second | Third |
|---|---|---|---|
| 1996 | A. L. S.-Vincentelli | K. G. Shin | S. M. Reddy |
| 1997 | K. G. Shin | E. R. Hancock | S. Jajodia |
| 1998 | H. Garcia-Molina | S. M. Reddy | P. S. Yu |
| 1999 | E. R. Hancock | T. S. Huang | E. Bertino |
| 2000 | B. Hancock | T. S. Huang | E. R. Hancock |
| 2001 | T. S. Huang | H. Zhang | E. R. Hancock |
| 2002 | M. T. Kandemir | E. R. Hancock | S. M. Reddy |
| 2003 | W. Li | M. T. Kandemir | H. Seidel |
| 2004 | W. Gao | W. Wang | H. Zhang |
| 2005 | C. T. Kandemir | C. Chang | W. Gao |

Table III
TOP THREE CORE AUTHORS IN DIFFERENT YEARS

are only four years from year 2006 to year 2009 which less than a period, the information from 2006 to 2009 is not used.

### C. Size of the EBSN

Figure 2 shows the numbers of *actor-nodes* in different periods. From this figure we can see that the number of *actor-nodes* in the *EBSN* grows fast, especially from the period 1986-1995 to the period 1996-2005. The number of *actor-nodes* increases from 130623 to 403571 which means that the number of researchers studying on 'database' in the period 1996-2005 increases more than three times than that of in the period 1985-1996.

### D. Actor-Node Ranking

An *actor-node* in the *EBSN* represents an author in the DBLP data set. There are two numerical attributes of an *actor-node*, *activity.value* and *importance.value*. Since in the period of 1996-2005 the *EBSN* has more actors and events than in other periods, we focus on ranking in that period.

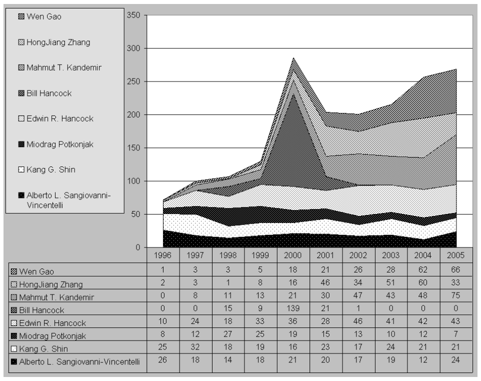
Table I describes the *actor-node* ranking based on *activity* from year 1996 to year 2005.

Table II shows the *actor-node* ranking based on *importance* from year 1996 to year 2005. Comparing Table I with Table II, we find that only in year 2000 the ranking result of top three active *actor-node*s are same as top three important *actor-node*s. Actually, in year 2000 the fourth active *actor-node* is different from the fourth important *actor-node*. Hence, we can conclude that in most years the ranking result of important *actor-node* and active *actor-node* are different.

Table III exhibits the top three core-nodes in different years. Table III is different from Table I and Table II. The core-nodes ranking is based on both the importance and the activity ranking results. We believe that the core-nodes rank is an important indicator because it combines both the activity and importance information in a domain.

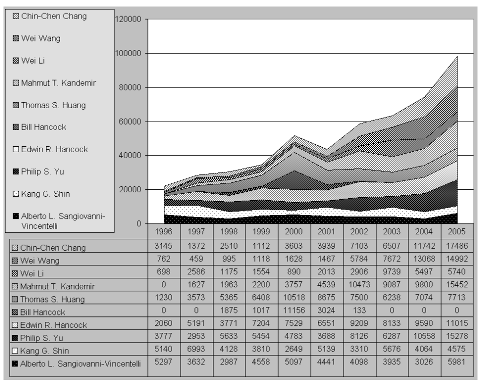
Figure 3 records the *activity* values of top active *actor-node*s from the year 1996 to the year 2005. In the graph, the stacked area indicates the author's *activity* value, and the x-axis is the year. The change of stacked area reveals the author's research "trend". If the stacked area increases, then we call the author in the developing period, otherwise we call the author in the developed period. There are only eight *actor-node*s in Figure 3. Refer to Table I, We can see two of them, i.e. HongJiang Zhang and Mahmut T. Kandemir, are two years' top active *actor-node*s.

Figure 4 demonstrates the *importance* values of top important *actor-nodes* from the year 1996 to the year 2005. In the figure, the stacked area presents author's *importance* value. If the stacked area increases, we say that the author's influence is in a developing period. This will give us some hints if we need to choose a researcher for a project. If the candidates have the same importance value, then we choose the one whose influence is in a developing period.



| | 1996 | 1997 | 1998 | 1999 | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 |
|---|---|---|---|---|---|---|---|---|---|---|
| Wen Gao | 1 | 3 | 3 | 5 | 18 | 21 | 26 | 28 | 62 | 66 |
| HongJiang Zhang | 2 | 3 | 1 | 8 | 16 | 46 | 34 | 51 | 60 | 33 |
| Mahmut T. Kandemir | 0 | 8 | 11 | 13 | 21 | 30 | 47 | 43 | 48 | 75 |
| Bill Hancock | 0 | 0 | 15 | 9 | 139 | 21 | 1 | 0 | 0 | 0 |
| Edwin R. Hancock | 10 | 24 | 18 | 33 | 36 | 28 | 46 | 41 | 42 | 43 |
| Miodrag Potkonjak | 8 | 12 | 27 | 25 | 19 | 15 | 13 | 10 | 12 | 7 |
| Kang G. Shin | 25 | 32 | 18 | 19 | 16 | 23 | 17 | 24 | 21 | 21 |
| Alberto L. Sangiovanni-Vincentelli | 26 | 18 | 14 | 18 | 21 | 20 | 17 | 19 | 12 | 24 |

Figure 3. Top active authors in different years



| | 1996 | 1997 | 1998 | 1999 | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 |
|---|---|---|---|---|---|---|---|---|---|---|
| Chin-Chen Chang | 3145 | 1372 | 2510 | 1112 | 3603 | 3939 | 7103 | 6507 | 11742 | 17486 |
| Wei Wang | 762 | 459 | 995 | 1118 | 1628 | 1467 | 5784 | 7672 | 13068 | 14992 |
| Wei Li | 698 | 2586 | 1175 | 1554 | 890 | 2013 | 2906 | 9739 | 5497 | 5740 |
| Mahmut T. Kandemir | 0 | 1627 | 1963 | 2200 | 3757 | 4539 | 10473 | 9087 | 9800 | 15452 |
| Thomas S. Huang | 1230 | 3573 | 5365 | 6408 | 10518 | 8675 | 7500 | 6238 | 7074 | 7713 |
| Bill Hancock | 0 | 0 | 1875 | 1017 | 11156 | 3024 | 133 | 0 | 0 | 0 |
| Edwin R. Hancock | 2060 | 5191 | 3771 | 7204 | 7529 | 6551 | 9209 | 8133 | 9590 | 11015 |
| Philip S. Yu | 3777 | 2953 | 5633 | 5454 | 4783 | 3688 | 8126 | 6287 | 10558 | 15278 |
| Kang G. Shin | 5140 | 6993 | 4128 | 3810 | 2649 | 5139 | 3310 | 5676 | 4064 | 4575 |
| Alberto L. Sangiovanni-Vincentelli | 5297 | 3632 | 2987 | 4558 | 5097 | 4441 | 4098 | 3935 | 3026 | 5981 |

Figure 4. Top important authors in different years

## V. Related Work and Discussion

Generally, there are three ways to analyse an dynamic social network [3]. The first method is to analyse a snapshot of a time point. This method only analyses very limited information of the network. It could hardly get high accurate results. The second method is to analyse an aggregate network which aggregate all events into a single network. This method actually treats the event-based social network as a static network, and achieves a statistic summary for the network. It could not catch the sequence of the events. The third method is to analyse series of aggregated networks. This method is suitable for event-based social networks because it analyses all information an event-based social network contains, and it respects the temporal elements in networks. In the area of social network analysis, most current studies use the first two methods [3]. Fiaidhi and Mohammed [4] ranked authors in DBLP based on their publications, and they use "*author collaboration score*" to rank authors which is a relative number. Their model makes use of the second method which treats the network as a static network. O'Madadhain et al. [1] defined a model for event-based ranking. Their model used the third method to construct the social network which analyses a series of aggregate networks. Their model could catch the dynamic features of the network, but they only focused on the actors. The events should also be concerned as events are the power of evolution in the event-based social network. Different from these methods, we firstly use the third method to construct an event-based social network. Then the authors are ranked based on their publications and the citation of their publications. In other words, we rank an *actor-node* based on the *influence* of the *event-nodes* which the actor-node participates in. Furthermore, we analysed the network as a dynamic network, and ranked authors in different years relied on their activity value and importance value through the proposed approach. We defined *core-nodes* which consider both the importance ranking result and activity ranking result. Table III shows the top three *core-nodes*. We also exploit "developing period" and "developed period" to describe an author's trends. Such information can be used to evaluate core researchers in different fields, and the trends of their impact. This work could be employed in other event-based social networks. The case study provides an example of how to use the *EBSN*.

## VI. Conclusion

In this paper we formally defined an *Event-Based Social Network (EBSN)*, an *actor-node*, an *event-node*, a *core-node*, a *lbAE* and a *lbEE*, and introduced how to construct an *EBSN*. We proposed algorithms to discover the *activity* and *importance* of an *actor-node*, and the *influence* of an *event-node*. Then we used the definitions and algorithms to construct an *EBSN* which is built on the DBLP data. After that we ranked the authors based on their publications and the number of citations of their publications in different years. We also used "developing period" and "developed period" to describe an author's trend. The *EBSN* could also be borrowed to analyse other event-based social network, such as email network. The future research will focus on pattern discovery and link mining in *EBSN*.

## References


[1] J. O'Madadhain, J. Hutchins, and P. Smyth, "Prediction and ranking algorithms for event-based network data," *SIGKDD Explor. Newsl.*, vol. 7, no. 2, pp. 23–30, 2005.

[2] J. Zhang, J. Tang, and J. Li, "Expert finding in a social network," in *Advances in Databases: Concepts, Systems and Applications*, ser. LNCS, vol. 4443/2008. Springer Berlin / Heidelberg, 2007, pp. 1066–1069.

[3] C. Tantipathananandh, T. Berger-Wolf, and D. Kempe, "A framework for community identification in dynamic social networks," in *KDD '07: Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*. New York, NY, USA: ACM, 2007, pp. 717–726.

[4] J. Fiaidhi and S. Mohammed, "Detecting some collaborative academic indicators based on social networks: A dblp case study," in *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2008*. Chesapeake, VA: AACE, January 2008, pp. 6150–6157. [Online]. Available: http://go.editlib.org/p/29234