

# Social Capital in Friendship-Event Networks

Louis Licamele  
Computer Science Dept.  
University of Maryland  
College Park, MD 20742 USA  
licamele@cs.umd.edu

Lise Getoor  
Computer Science Dept.  
University of Maryland  
College Park, MD 20742 USA  
getoor@cs.umd.edu

## Abstract

*In this paper, we examine a particular form of social network which we call a **friendship-event network**. A friendship-event network captures both the friendship relationship among a set of actors, and also the organizer and participation relationships of actors in a series of events. Within these networks, we formulate the notion of **social capital** based on the actor-organizer friendship relationship and the notion of *benefit*, based on event participation. We investigate appropriate definitions for the social capital of both a single actor and a collection of actors. We ground these definitions in a real-world example of academic collaboration networks, where the actors are researchers, the friendships are collaborations, the events are conferences, the organizers are program committee members and the participants are conference authors. We show that our definitions of capital and benefit capture interesting qualitative properties of event series. In addition, we show that social capital is a better publication predictor than publication history.*

## 1. Introduction

Recently there has been a great deal of interest in research involving social networks, including both modeling and analyzing the networks. A social network describes actors and their relationships and, in some cases, events and actors' participation in events. A social network can be characterized by its relational structure; the underlying graph structure of the network dictates the structural properties. These include everything from the density of the graph and average degree of the nodes to the measure of centrality and information flow.

In this paper we will look at networks that are more complex than the classic 'who-knows-who' or friend-of-a-friend

(FOAF) networks. In addition to friendship networks, we are also interested in event networks [14]. The networks we propose, which we call *friendship-event networks*, combine information about friendship networks and information about events, including the organizers of an event and the participants in an event (these may be overlapping). We present a general formulation of these friendship-event networks.

To measure interesting structural properties of these networks, we define the notions of *capital* and *benefit*. Capital is a measure of an actor's social capital. It is defined in terms of the number of event organizers with whom an actor is friends. Benefit is defined from the perspective of an event organizer, in terms of how much benefit they give their friends and from the perspective of an event participant in terms of their participation in events. Depending on context, benefit may be perceived positively (as in the more benefit that exists in a network the greater the benefit for everyone in the network) or negatively (in terms of bias). Here we view them simply as descriptive properties useful for understanding the data.

Events naturally have a time associated with them and it is possible for relationships, positions and roles to change over time. These changes will in turn affect the social capital of an individual as well as benefit received and benefit given. To be more specific, events can occur at different times, the organizers of events change over time, and a different set of actors might participate in each event. In order to analyze temporal trends in capital and benefit properly, we must model these temporal aspects in our networks.

Building on our models of temporal friendship-event networks, we propose a predictive model for benefit, based on characteristics of the structure of the friendship-event network. We look at the problems of predicting an actor's participation in an event and predicting a group's participation. We describe how both benefit history and social capital can be used as predictors.

To demonstrate the usefulness of the measures that we have developed, we apply them to academic collaboration networks. These networks describe researchers and their collaborations. In addition to researches and collaborations, we also have conference events along with their organizers (program committee (PC) members) and participants (authors). Collectively, we will refer to these friendship-event networks as academic collaboration networks. In this example dataset, an authors friend is defined as someone with whom an author has coauthored, and social capital is the number of these friends who serve on the program committee for the conference in which the author publishes. Benefit given is expressed as the number of papers that the friends of a PC member publish in the conference, and benefit received is the number of papers that an author publishes in a conference.

While this domain might seem very specific, other examples can be seen in the political and corporate domains. In politics, a large number of events occur because of who you know and with whom you associate. An example is the processing of bills by, say, a senate subcommittee. In this case, the actors are the senators. Friendships can be defined in several ways; for example having cosponsored a bill together in the past. The event is a session of the subcommittee and it is characterized by the set of bills that make it through the subcommittee during that session. The committee members can be seen as the organizers of the event. A senator’s social capital is the number of friends he or she has on the committee. Benefit is received if a senator’s bill makes it through the subcommittee; benefit is given if a friend’s bill is successful.

This model can be extended to incorporate the next step of the proposal process, namely submitting the bill to the general body. This domain could be presented as a hierarchy of events and organizers, each level corresponding to a different amount of capital. A similar example can be seen in the corporate domain; the event here being selection as an executive of the company. The organizers are the directors of the board. Friendship would be defined as having worked together in the past. These are just two simple examples of the generality of the models described here; we believe there are many others. However, to avoid making the presentation overly abstract, and because it is a domain with which we are all familiar, we will continue to focus on the academic collaboration domain for the rest of this paper.

Our specific contributions in this paper are as follows. First, we introduce a novel class of social networks which we refer to as friendship-event networks. These networks have a structure which captures many commonly occurring dynamic, temporal social networks. Next we give a quantitative definition of social capital in these networks. We show how the definition can be used to define a notion of

benefit given and benefit received, which captures the transfer of social capital in the network. Finally, we present results on a real-world dataset, showing the utility of our measures both for descriptive purposes and, perhaps more interestingly, also as a predictor for future event participation.

We begin by describing some of the related work in Section 2. In Section 3, we give a general definition for the family of friendship and event networks that we study and show the mapping to the academic collaboration networks. In Section 4, we define capital and benefit and in Section 5, we further extend our definitions with the important element of time. In Section 6 we explain the participation prediction task. Finally, in Section 7 we describe some preliminary results applying these measures to three different computer science conferences over a 10 year time period.

## 2 Related Work

A large portion of the work in mining social networks has focused on analyzing structural properties of the networks. For recent surveys, see Newman [13] and Jensen [6]. Much of the work has been descriptive in nature, but recently there has been more work which uses structural properties for prediction. Within this category, a number of papers focus on the spread of influence through the network (e.g., [4, 8]). These papers attempt to identify the most influential nodes in the network. Domingos and Richardson [4] use a global, probabilistic model that employs the joint distribution of the behavior over all the nodes. Kempe et al. [8] use a diffusion process that begins with an initial set of active nodes and uses different weighting schemes to determine whether or not a neighbor should be activated. McCallum et al. have proposed role discovery in social networks by looking at messages sent and received between entities [12]. Liben-Nowell and Kleinberg [9] attempt to predict future interactions between actors using the network topology. In addition, Palmer et al. [16] propose an efficient method for approximating the connectivity properties of a graph.

Even though social capital is defined slightly differently in different contexts such as sociology and economics, most definitions agree that social capital is a function of ties between actors in a social network whereas human capital refers to properties of individual actors. Degenne and Forse [3] trace the idea back to Hobbes who said “to have friends is power” [5]. However, the term itself and its systematic studies are relatively recent [1, 11, 2]. Portes argues that a systematic treatment of social capital must distinguish between the “possessor of the capital” (actors who receive benefits), “sources of the capital” (actors who give benefits), and the resources that have been received or given [18]. In our analysis, the “sources of the capital” are the organizers

of the events. Two related notions in social network analysis are position and role; White et al.[19] have defined position as subsets of actors who have similar ties to other actors, and role as patterns of relationships between these actors or subsets.

Like O'Madadhain et al. [15, 14] and others [9], we are interested in capturing temporal aspects of social networks. Our work differs from O'Madadhain in that we have a richer model of events, which includes information about organizers and participants, and our focus is on characterizing the social capital of the networks. A number of link prediction tasks have been studied in academic collaboration networks. Both O'Madadhain et al. and Liben-Nowell et al. predict collaborations in co-authorship networks. Popescul [17] examine citation prediction. Here we look at conference publication prediction, both for a single author and for a collection of authors. The work presented here expands on an earlier, unpublished workshop paper [10].

### 3 The Friendship-Event Network

We begin with a generic description of a family of social networks which we refer to as friendship-event networks. These networks have the following sets of entities:

**actors:** a set of actors  $A = \{A_1, \dots, A_n\}$

**events:** a set of events  $E = \{E_1, \dots, E_m\}$

and the following sets of relationships:

**friends:**

$$F(A_i, A_j) = A_i \text{ is friends with } A_j$$

**organizers:**

$$O(A_i, E_k) = A_i \text{ is an organizer of event } E_k$$

**participants:**

$$P(A_i, E_k) = A_i \text{ is a participant in event } E_k$$

We use  $f(A_i)$  to denote the friends of actor  $A_i$ , i.e.,

$$f(A_i) = \{A_j \mid F(A_i, A_j)\},$$

and  $o(E_k)$  to denote the organizers of event  $E_k$ , i.e.,

$$o(E_k) = \{A_i \mid O(E_k, A_i)\},$$

and  $p(E_k)$  to denote the participants in event  $E_k$ , i.e.,

$$p(E_k) = \{A_i \mid P(E_k, A_i)\}.$$

In some cases, it makes sense to allow an actor to participate in an event more than once. In these cases, for each  $E_k$ , we define an associated set of subevents,

$$se(E_k) = \{E_{k1}, \dots, E_{kp}\},$$

and define a participant subevent relationship:

$$p(A_i, E_k, E_{kj}) = A_i \text{ is a participant in subevent } E_{kj} \text{ of } E_k$$

Then the participants can be defined in terms of the subevent relation:

$$p(E_k) = \{A_i \mid \exists E_{kj} \in se(E_k) \text{ s.t. } P(A_i, E_k, E_{kj})\}$$

In terms of the academic collaboration example, the actors are the researchers (both authors and PC members) and the events are the conferences. The friendship relation is defined based on whether two researchers have co-authored a paper together. In this case the friendship relationship is symmetric, but this may not be true in other domains. The organizers of an event are the PC members and the participants in the event are the set of authors that have papers published in the conference. Since authors may have more than one publication in a conference, the subevent relationship is authorship of a paper (the subevent) in a conference. An illustration of an academic collaboration network is given in Figure 1.

### 4 Event-Specific Capital and Benefit

Next we introduce the notions of capital and benefit. *Social capital* is a measurement of the amount of "good-will" available to an actor based on the actor's friendship relationships. We begin by defining social capital in the context of a single event  $E_k$ .

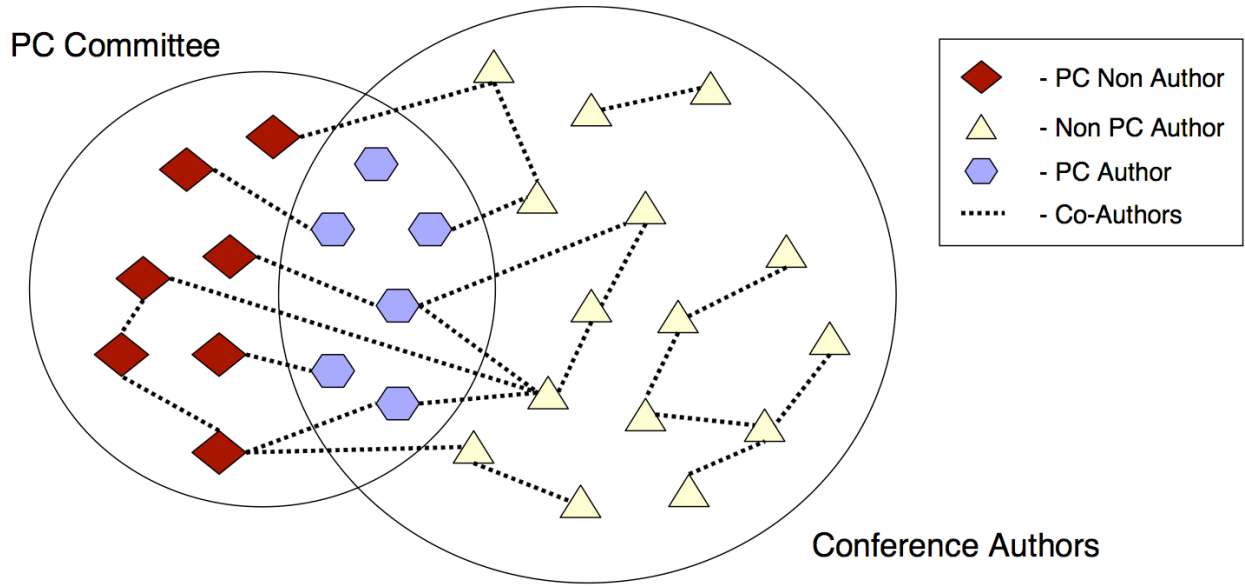
**Definition 1** *Social Capital:* The social capital of an actor  $A_i$  in an event  $E_k$  is the number of organizers with whom the actor is friends:

$$SC(A_i, E_k) = \sum_{A_j \in o(E_k)} I(F(A_i, A_j))$$

where  $I$  is an indicator function which is 1 when the relation holds.<sup>1</sup>

The definition is based on Hobbes's idea that it is more important to have powerful friends than to have numerous powerless friends [5]. Therefore, we define an actor's capital in terms of organizer friends rather than simply friends. We also define the notion of the *social capital ratio* which is the proportion of the organizing committee with whom an actor is friends:

<sup>1</sup>To improve readability, we will drop the  $I$  in the definitions that follow, but throughout the intended interpretation is that we are counting the number of times a relation or expression holds.



**Figure 1. An event in the friendship-event network for academic collaboration. The actors in the network are PC members and authors. The edges in the network indicate co-authorship links (friendship). The organizers are the PC members (the set on the left), and the participants are the authors (the set on the right). Note that these sets need not be disjoint; i.e. a PC member can be an author as well. The three categories of actors are: PC-Non-Authors, PC-Authors, and Non-PC-Authors. If we name the sets as  $PC$  and  $CA$  from left to right, these categories refer to the sets  $PC \setminus CA$ ,  $PC \cup CA$ , and  $CA \setminus PC$  respectively.**

**Definition 2 Social Capital Ratio:** The social capital ratio of an actor  $A_i$  in an event  $E_k$  is the proportion of organizers with whom  $A_i$  is friends:

$$SCR(A_i, E_k) = \frac{\sum_{A_j \in o(E_k)} F(A_i, A_j)}{|o(E_k)|}$$

**Definition 3 Group Social Capital:** The social capital of a group of actors in a subevent  $E_{kj}$  in an event  $E_k$  is defined by taking some statistical aggregation over the social capital of each of the individual actors in the group. An obvious example would be to use the sum of the social capital values:

$$SC(E_{kj}) = \sum_{A_i \in p(A_i, E_k, E_{kj})} SC(A_i, E_k)$$

In our example domain, we can use this definition of group social capital to refer to the social capital of a particular paper, which is the subevent. The set of all of the actors that participate in this subevent is the group of the authors of the paper and the group social capital of the subevent is the sum of their social capital.

Next we turn to a definition of *benefit*. We can look at benefit from both the perspective of an event participant and

an event organizer. In our model, participation in an event is considered beneficial. As mentioned earlier, we may consider participation to be a binary yes/no relationship, or, alternatively, actors may participate in an event more than once, and the more an actor participates, the more benefit they receive. Given our motivating example, the latter definition is more appropriate, so we use it in our definition of benefit below.

**Definition 4 Benefit Received:** Actors receive benefit when they participate in events. The benefit received by an actor  $A_i$  in event  $E_k$  is:

$$BR(A_i, E_k) = \sum_{E_{kj} \in se(E_k)} S(A_i, E_k, E_{kj})$$

In the context of the academic collaboration network the benefit an author receives for a given conference is the number of publications the author has in the conference. We also define the benefit received ratio as the proportion of conference paper authorships for a particular conference (where a paper with 3 authors counts as 3 paper authorships):

**Definition 5** *Benefit Received Ratio:* The benefit received ratio for an actor  $A_i$  in event  $E_k$  is:

$$BRR(A_i, E_k) = \frac{BR(A_i, E_k)}{\sum_{A_j \in A} BR(A_j, E_k)}$$

From the perspective of an event organizer, we measure the benefit given. Benefit given is the benefit that an event organizer's friends receive.

**Definition 6** *Benefit Given:* The benefit given by an organizer  $A_o$  of an event  $E_k$  is:

$$BG(A_o, E_k) = \sum_{A_i \in f(A_o)} BR(A_i, E_k)$$

and the benefit given ratio is the percentage of all conference benefit that an organizer is responsible for:

**Definition 7** *Benefit Given Ratio:* The ratio of benefit given by an organizer  $A_o$  of an event  $E_k$  is:

$$BGR(A_o, E_k) = \frac{BG(A_o, E_k)}{\sum_{A_i \in o(E_k)} BG(A_i, E_k)}$$

We can also look at benefit from the event perspective by aggregating these measure over events:

**Definition 8** *Average Benefit Received Ratio and Average Benefit Given Ratio:* The average benefit received ratio for an event  $E_k$  is:

$$ABRR(E_k) = \frac{\sum_{A_i \in p(E_k)} BRR(A_i, E_k)}{|p(E_k)|}$$

And the average benefit given ratio for an event  $E_k$  is:

$$ABGR(E_k) = \frac{\sum_{A_o \in o(E_k)} BRG(A_o, E_k)}{|o(E_k)|}$$

## 5 Temporal Aspects

Social networks are dynamic so time obviously plays an important role. We look at two temporal components of our networks.

### 5.1 Event Series

It is often the case that there is not just a single event, but that multiple events form an event series. The conferences in our academic collaboration network are one example, but others include regularly scheduled meetings, a book or movie series or a series of sporting events.

We introduce the notion of an event series by adding a time index to our events:

- **event series:** an event series  $E_k(T)$  is composed of a set of events  $E_k(t_1), \dots, E_k(t_q)$

The notions of benefit received and benefit given defined above can easily be extended to event series. For example, the overall average benefit received ratio for a conference series  $E_k(T)$  is:

$$OABRR(E_k(T)) = \frac{\sum_{t=t_1}^{t_q} ABRR(E_k(t))}{q}$$

and similarly we can define  $OABGR(E_k(T))$ , the overall average benefit given ratio for a conference series  $E_k(T)$ .

## 5.2 A Temporal Definition of Friendship

Now that we have a notion of time associated with events, clearly we must update our definition of friendship so that we only consider current friends and not future friends in our calculations. We modify the definition of friendship to include a temporal argument:  $f(A_i, A_j, t)$  means that  $A_i$  and  $A_j$  are friends at time  $t$ . In the case of our academic collaboration network, we say that  $A_i$  and  $A_j$  are friends at time  $t$  if they co-authored a paper which was published at or before time  $t$ .

Friendships evolve over time. In addition to the above definition which defines friendship at a particular time, we also introduce a time window, which allows us to consider only friendships within a certain recency window. For the academic collaboration network, we say that  $A_i$  and  $A_j$  are friends at time  $t$  if they co-authored a paper which was published within a time window of size  $n$  before time  $t$ .

**Definition 9** *Friendship:* Two authors are considered friends at time  $t$  if they have co-authored a paper within last  $n$  years.

$$F(A_i, A_j, t, n) \Leftarrow \exists t' CoAuthor(A_i, A_j, t') \wedge 0 \leq t - t' \leq n$$

## 6 Participation Prediction

Given the above definitions, there are a number of predictive tasks of interest. Here, we focus on benefit, or predicting future participation, based on both past benefit and social capital. Let  $p(A_i, E_k(t))$  denote the random event that actor  $A_i$  participates in event  $E_k(t)$ . Then one quantity of interest is  $Pr(p(A_i, E_k(t)))$ , the probability that actor  $A_i$  has participated in event  $E_k(t)$ . We will refer to this prediction task as simply *participation prediction*.

Another quantity of potential interest is, given the participation of an actor in *some* event at time  $t$ , which is the most likely event? We denote the random event that actor  $A_i$  has

participated in some event  $E_k(t) \in E$  by  $p(A_i, t)$ , and then we are interested in

$$\operatorname{argmax}_{E_k(t) \in E} Pr(p(A_i, E_k(t)) | p(A_i, t))$$

Similarly, if there are subevents, we may be interested in the probability that a subevent  $E_{k'j}$  is a subevent of  $E_k(t)$ , given that subevent  $E_{k'j}$  occurred at time  $t$  in *some* event  $E_k(t) \in E$ , denoted  $p(E_{k'j}, t)$ . We write this as follows:

$$\operatorname{argmax}_{E_k(t) \in E} Pr(E_{k'j} \in se(E_k(t)) | p(E_{k'j}, t))$$

We refer to these predictions as *event-participation* prediction.

Intuitively, either form of participation will depend on past participation, and there is a question of whether it will depend on the social capital of the actors involved. Ideally, if our definition of social capital is useful, it should serve as a useful predictor for future participation.

In order to quantify past participation for an actor, we choose some temporal window  $n$  and measure participation at each point  $t - 1, t - 2, \dots, t - n$ . We refer to this as the participation history. We explored more complex models of the time series, but this simple model performed best. In order to quantify past participation for a subevent, we measure the participation for the actors in the subevent at each point  $t - 1, t - 2, \dots, t - n$ . We evaluate the following measures of actors' participation: *min*, *max*, *mean* and *total* participation.

In order to quantify social capital history for an actor, we again choose some temporal window  $n$  and measure social capital at each point in the window. In order to quantify social capital history for an event, we measure the social capital for the actors in the event at each time point. As above, we evaluate the following measures of the actors' social capital: *min*, *max*, *mean* and *total*.

We evaluate classifiers which use various combinations of these features in standard off-the-shelf learning implementations. As we will see, *participation* prediction is quite challenging, using either participation or social capital history. However *event-participation* prediction is feasible, and both participation and social capital are accurate predictors.

## 7 Experimental Results

We explore how our proposed descriptive statistics for social capital and benefit apply to several real academic friendship-event networks. We measured friendship, capital and benefit on a dataset describing publication information and program committee members for five major conferences of a subfield of computer science. There are 11,644

unique papers from 1959 till 2004, and these papers contain 11,554 unique authors. There are 1,821 distinct program committee members. Because two of the conferences have missing data for PC members, we leave them out for the capital and benefit analysis, but use their publications for defining friendships.

The summary statistics for the data are given in Table 1. The mean  $\mu$  and standard deviation  $\sigma$  are computed for the last 10 years of the data, i.e. from 1994 to 2003. As we can see, C1 and C2 can be considered similar in terms of having a relatively large number of papers, a large number of authors and relatively large PC. C3 on the other hand, is significantly smaller. It turns out that C1 and C2 are two flagship conferences for the area, and are more applied, while C3 has a more theoretical bent.

Overall aggregate statistics for the conferences are also shown in Table 1. Here we are using a friendship window size of 5 years. Interestingly, despite the difference in the sizes of the friendship-event networks for the three conferences, the aggregate structural statistics are surprisingly similar. The statistics are not significantly different for all three conferences; the means are all less than one standard deviation away from one another. The only significant difference is in the standard deviations in benefit given (BG) for conference C1 and C2 as compared to conference C3.

### 7.1 Event Series Analysis

The notion of friendship and capital can allow for insights to be made when comparing different conferences. We can also look at trends in a specific conference by examining how friendship and capital change throughout the years. We present a more detailed inspection of conference C1 in order to demonstrate how these new notions can be used in this exploratory data mining process.

We begin by examining how the levels of friendship compare between the different categories of actors (PC-Authors, Non PC-Authors and PC Non-Authors). This information was calculated for the last 23 years of conference C1 and is shown in Figure 2(a). The friendship window is held fixed at 5 years. As can be seen, the friendship levels increase over time. These values are an average over all the individuals involved, so it is not skewed by the increase in the number of authors or the size of the program committee over time.

The amount of friendship for the PC-Authors is one of the first things that stands out in this graph. It appears that the PC-Authors have more than double the number of friends than both the Non-PC-Authors and the PC-Non-Authors. One explanation for the difference in friendship between PC-Authors versus Non-PC-Authors is that we might assume that PC members have more friends and that

**Table 1. For each conference series, average number of papers, average number of authors, average PC size and aggregate statics for friendship, capital, benefit given (BG) and benefit received (BR) for the past 10 years.**

Conf.	Papers		Authors		PC		Friendship		Capital		BR		BG	
	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$
C1	78.90	9.45	223.20	25.24	32.60	5.87	8.29	2.50	0.55	0.20	1.16	0.46	4.64	10.28
C2	87.00	23.75	237.70	85.89	69.62	23.30	7.45	1.20	0.71	0.42	1.09	0.34	3.13	8.41
C3	29.20	2.94	66.30	9.87	9.30	2.87	8.37	2.31	0.57	0.29	1.10	0.33	3.15	4.64

is why they are chosen to be on the program committee. In that case, we would expect for the friendship levels of all PC members, not just the ones who are authors, but also the PC-Non-Authors, to be higher than the friendship levels of the Non-PC-Authors. The PC-Non-Authors have a slightly higher friendship level than the Non-PC-Authors but it is still much lower than PC-Authors' friendship levels.

To better understand these differences, we examined the amount of capital of each group. Given the number of friends that a person has, and assuming that each friend has an equal chance of being on the program committee, we would expect to find similar patterns in the capital values. The capital values are shown in Figure 2(b). The same overall upward trend that was seen for the friendship values is present. The PC-Authors' social capital is still more than double the values of the other two groups. In many ways this is not surprising because they had the largest number of friends. The interesting results in this graph are those of the PC-Non-Authors. Though it was shown that they have more friends than Non-PC-Authors, it appears that they have less friends on the program committee.

One possible scenario that would lead to an increase in capital over time would be if the size of the program committee increased each year, which in many cases it does. To check if this trend exists in this data, we calculate the ratio of total number of program committee members to the total number of authors per year. As it turns out, the size of the program committee grows at a slower rate than the total number of authors overall and over the last ten years this ratio has stayed somewhat static.

Another way to look at the difference in trends between the friendship and capital values is to examine the ratio of capital over friendship. This ratio is shown in Figure 2(c) for all actor groups. Overall, the PC-Authors have the highest percentage of friends that are on the program committee with them. The Non-PC-Authors have the next highest percentage of friends on the program committee. The PC-Non-Authors have a much lower percent of their friends that are on the program committee. Perhaps this is why they are

much less likely to publish in a conference where they serve on the PC.

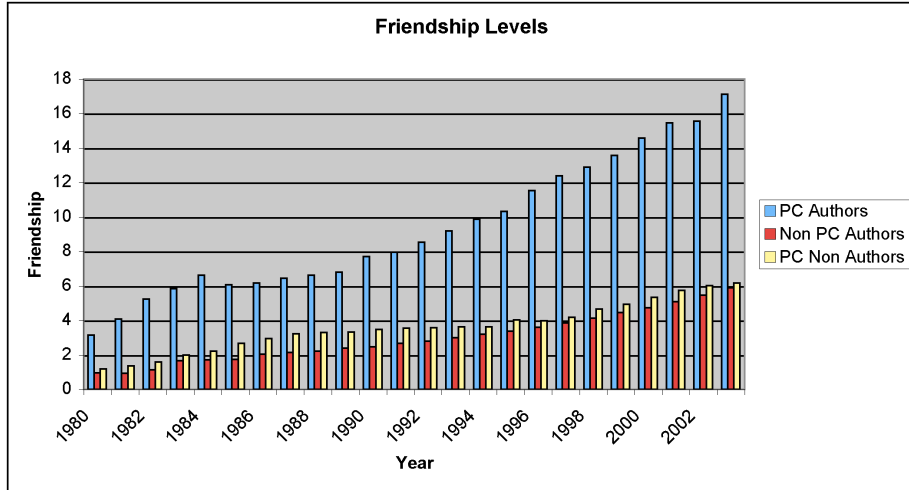
## 7.2 Predictive Analysis

The analysis in the previous section gives a descriptive characterization of an event series which may be useful for cross-event comparisons. In this section, we evaluate predictive models for event participation. We examine models which make use of participation history for the prediction and compare them with models based on social capital.

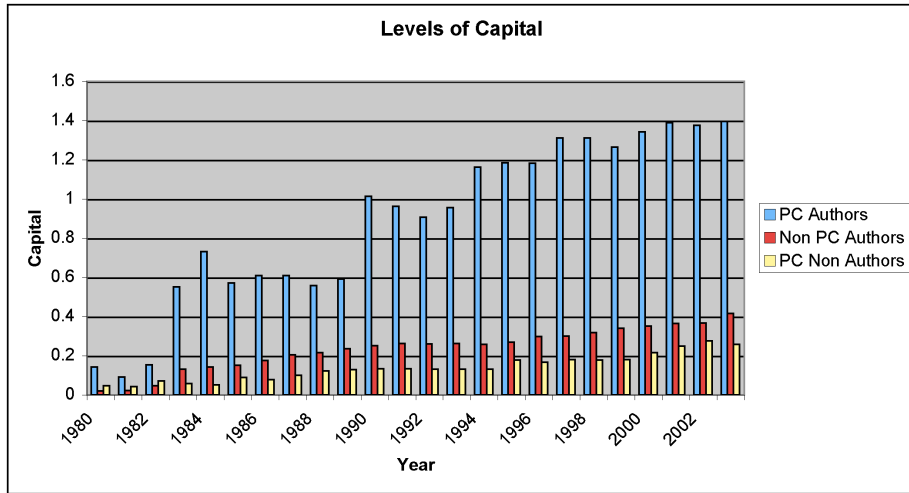
The first prediction task that we examined was participation prediction. For this domain, this translates into predicting whether or not an author will publish in a particular conference in a particular year. Unfortunately, this prediction task proved too difficult. Based solely on structural properties such as participation history and social capital, our models were not able to construct useful models that could be used with any confidence. This is perhaps not surprising, since authorship probability is so small.

The second prediction task that we investigated was event-participation prediction. We looked at event-participation for actors (Section 7.2.1) and the groups of actors (Section 7.2.2) in subevents. For this dataset, for an actor, this translates into predicting in which conference an author will publish, given that they have published once in some conference in the current year. For a group of authors that publish a paper together in the current year, this translates into predicting in which of the three conferences it appears, based on characteristics of the authors' publication history and social capital.

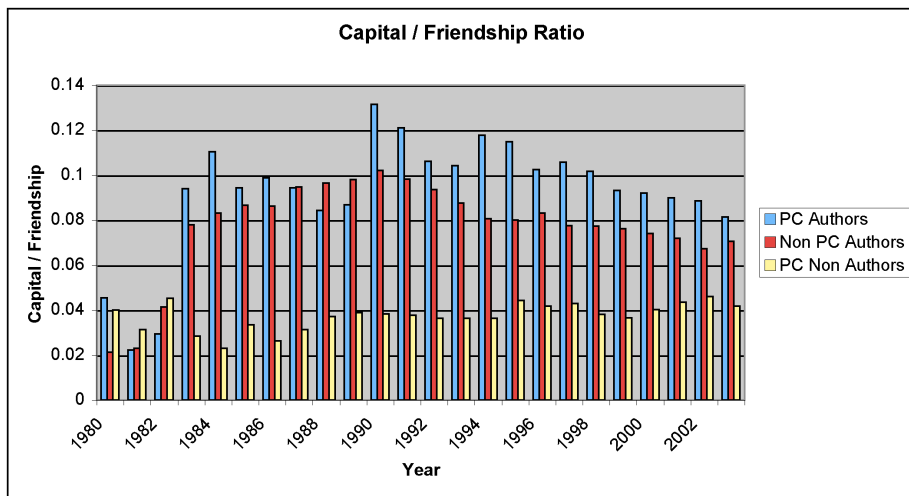
We explored a variety of classifiers; here we present our results for an SVM using a radial basis kernel [7]. All of the experiments presented here were done using ten-fold cross validation. The folds were created by random sampling from the dataset. There are 2,574 distinct authors in the dataset that have published in the ten year window that we are interested. In this time period there are 1,529 papers total in the three conferences.



(a)



(b)



(c)

**Figure 2. Detailed analysis of C1 over 10 years (1994-2003) for (a) friendship (b) capital and (c) capital-friendship ratio**



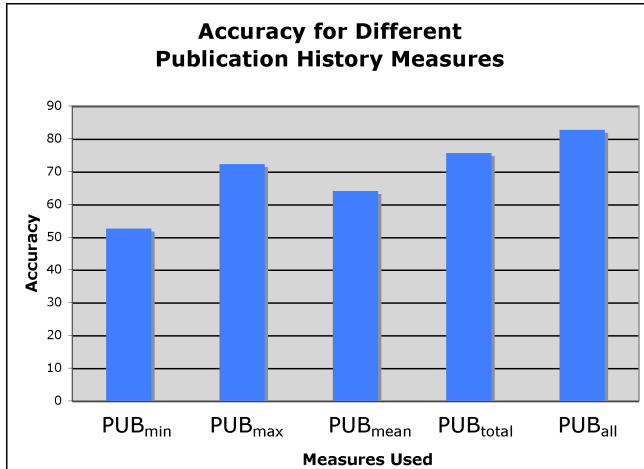


Figure 3. The accuracy for group event-participation prediction using several different publication history measures is shown. These results use a five year window. For each year in the window, the min, max, mean or total publication count of the group of authors is computed. *All* includes the four measures, for each year in the window.

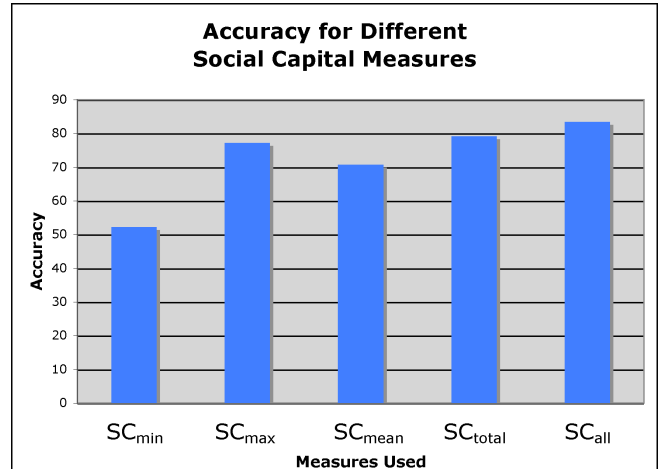


Figure 4. The accuracy for group event-participation prediction using several different social capital measures is shown. These results use a five year window. For each year in the window, the min, max, mean or total social capital of the group of authors is computed.

We use the following features for this prediction task:

**publication history:** The number of publications for each of the authors in each of the three conference per year over the past five years.

**social capital history:** The social capital of each author in each conference per year over the past five years.

**current social capital:** The social capital of each author in each conference in the current year.

### 7.2.1 Author Event-Participation Task

We now discuss the author event-participation prediction task. The goal of this task is to be able to predict which conference an author will publish in for a given year, given that they have published in the current year. We evaluated the classification accuracy for this prediction based on publication history alone, social capital alone, and publication history combined with social capital. Social capital alone gives us an accuracy of 42.5%. Based solely on the publication history we were able to achieve an accuracy of 45.2%. Adding in social capital raises this accuracy to 45.9% (not statistically significant). While our predictions are better than random, since there are three possible conferences, this

prediction task is still quite difficult. While an author has a strong relationship with a conference, there could be several reasons why they might not publish in that conference for this particular year. It is important to note that we are not considering any attributes of the paper, such as the content or the coauthors, the only measures that we are using are the publication history and the social capital.

### 7.2.2 Group Event-Participation Prediction

The next prediction task that we explored is group-event participation. In this domain, this corresponds to predicting where a paper will be published given the group of authors of the paper. Intuitively, this task can be seen as augmenting the author event-participation task with co-author information.

**Capturing Group Histories** In the case where we have groups of authors, the best ways of measuring the group participation and social capital histories is not clear. We began by comparing a variety of different methods for aggregating the measures, including taking the minimum, maximum, mean and total. All of these aggregates were computed for each year in the history window.

Figure 3 show the results for different measures for the group publication history. We examined not only the av-

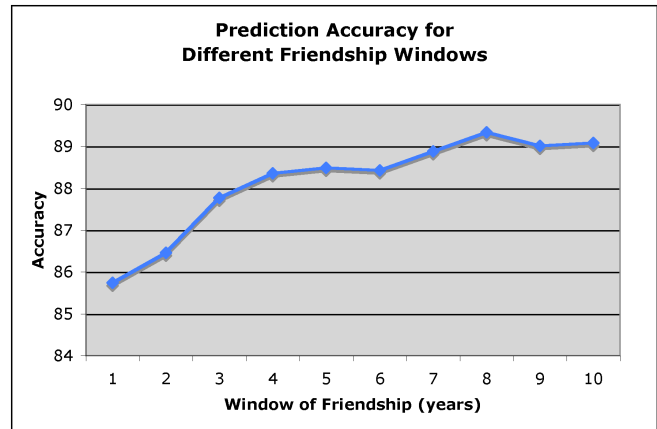
erage publication history, but also the minimum publication history, the maximum publication history, and the total publication history. Separate evaluations were done for each measure. In terms of predictive power, the minimum publication history gives the lowest accuracy at 52.52%. The maximum publication history, on the other hand, has an accuracy of 72.27%. This is better than the mean publication history which is 64.03%. The total publication history is the most informative of the single measures for this prediction task and leads to an accuracy of 75.61%. We can do even better by using a combination of all four of these measures. This combination of all of the measures for publication history achieves 82.67% accuracy. Because the combination of all the measures achieves the best results, in later reported results which use publication history, all measures are used.

Next we examined different ways of measuring a group's social capital. In addition to the average social capital for a group, we also measured the minimum social capital, maximum social capital and total social capital. Figure 4 shows the results using these different measures. Evaluating each measure in isolation, we see that minimum social capital gives the lowest result which is slightly over 50%. Minimum social capital helps the prediction task but the other measures do better. Maximum social capital has a much higher accuracy of 77.17%, and does better than the average social capital at 70.77%. Total social capital is the best predictor of the individual representations of social capital with 79.20%. By utilizing all four of these metrics as features in the classifier, we are able to obtain an accuracy of 83.39%. Note that this is better than we are able to achieve using publication history, and the difference is statistically significant (with  $p < 0.05$ ).

### Combining Social Capital and Publication History

Next we explored the importance of using the current social capital as compared to using the social capital history. Figure 7 shows that using only the current social capital alone, is not a very good predictor. The social capital history (using all of the measures over the past 5 years) is significantly better. Combining both, denoted in this figure as  $SC$ , is a bit better. We show the best results for publications,  $PUB_{all}$ , and we also show the result of publication history together with the combined social capital measures,  $PUB_{all} + SC$ . Combining both publication history and social capital gives us the best performance, 88.69%.

**Determining the Proper Window Size** An issue that comes up in performing this analysis is the question of how to correctly select the appropriate window for friendship. For the results presented thus far a default window size of 5 years has been used. In order to show how varying the friendship window changes the prediction accuracy, we

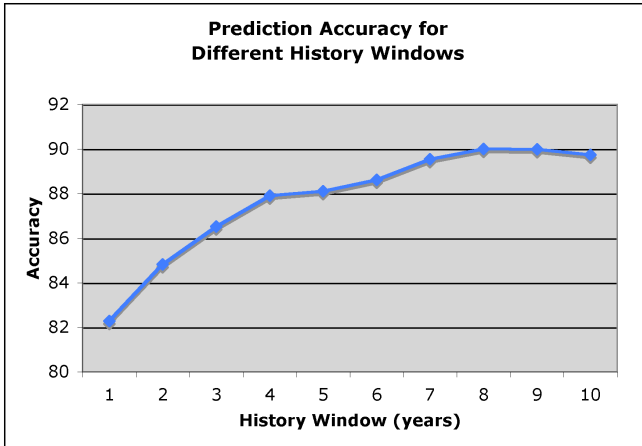


**Figure 5. The accuracy for group event-participation prediction as the friendship window is varied.**

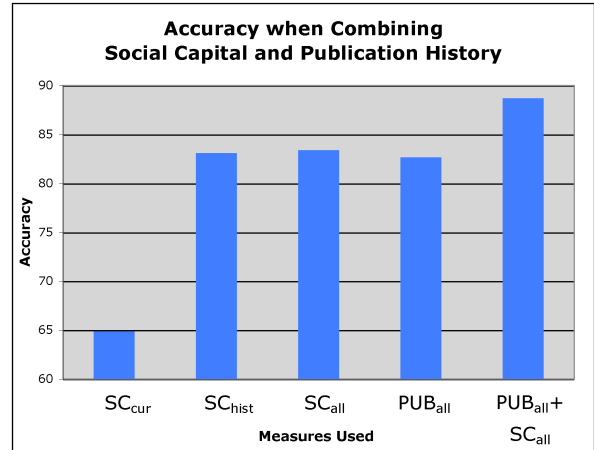
trained classifiers for friendship windows ranging from one to ten years. For each of these we included the publication history and the social capital. The results for different friendship window durations are shown in Figure 5. The accuracy steadily increased for windows of up to 8 years and then begins to decline. This shows the importance of choosing the friendship window appropriately; we need enough of a friendship history to give a good picture of the social capital, but not so much that the measure is out of date.

The friendship window is not the only window size that we must select. The calculations of the publication and social capital history rely on the fact that we have a meaningful history window. The predictive ability of the classifier given different history windows for both publication and social capital history from one to ten years is shown in Figure 6. Just like the friendship window, the accuracy rapidly increases in the beginning and then appears to peak at 8 years.

We also considered more complex definitions of friendships. A simple approach is to weight the friendships by the number of shared publications. In our preliminary experiments, for a five year friendship window this leads to slightly lower accuracy (not statistically significant) with 88.48% compared to 88.69% when not using the weighted measure. Other alternative schemes that combine the weights with a more sophisticated temporal model are possible, although again, in our preliminary experiments we did not find this to be beneficial.



**Figure 6.** The accuracy for group event-participation prediction as the publication and social capital history window is varied.



**Figure 7.** Group event-participation prediction accuracy is shown. Based on social capital for the current year ( $SC_{cur}$ ), social capital history ( $SC_{all}$ ), combined social capital ( $SC$ ), publication history ( $PUB_{all}$ ) and the combination of the publication history and combined social capital ( $PUB_{all} + SC$ )

## 8 Future Work

Given the research and results acquired, there are a number of avenues for future work. We would like to examine richer models for defining friendship. The friendship relationship is currently treated as a measure that, within the friendship window, does not diminish over time. For example, it would be interesting to examine models in which friendship decays over time. A methodology for determining optimal friendship and history windows also warrants further research. It would also be informative to study how the strength of capital as a predictor changes across different social networks, such as the political and corporate examples presented in Section 1. These examples suggest further directions of research. In the political domain there are several levels of organizers, each with potentially different amounts of influence. It would be interesting to explore models which can capture this hierarchical structure. Lastly, while our current dataset is probably too small to explore distributional issues with confidence, given a larger dataset, it would be worthwhile to examine some of the theoretical properties of social capital (e.g. heavy-tail, lognormal distribution, etc.) and how they are affected by changes in the network.

## 9 Conclusion

We have formulated a general family of friendship-event networks, and given a quantitative definition for social capital, benefit received, and benefit given. We have presented results on the author collaboration network describing con-

ferences as event series, event organizers as PC members and event participants as conference authors. We have examined the prediction of participation, and shown that social capital is a useful predictor. Social capital in fact performs better than past participation as a predictor for group event-participation. Ideally, these definitions could be used as part of a design process, which could, depending on the context, allow us to construct friendship-event networks that would optimize benefit. This could be of use for a variety of tasks such as constructing program committees, assigning reviewers and author networking.

## References

- [1] P. Bourdieu. *Handbook of Theory and Research for the Sociology Education*, chapter The forms of capital, pages 241–258. Greenwood, New York, 1985.
- [2] J. S. Coleman. Social capital in the creation of human capital. *The American Journal of Sociology*, 94:S95–S120, 1988.
- [3] A. DeGenne and M. Forse. *Introducing Social Networks*. SAGE Publications, London, 1999.
- [4] P. Domingos and M. Richardson. Mining the network value of customers. In *Proceedings of the seventh ACM SIGKDD*, pages 57–66, New York, NY, USA, 2001. ACM Press.
- [5] T. Hobbes. *Leviathan*. Collier, New York, 1962.
- [6] D. Jensen and J. Neville. Data mining in social networks. In *National Academy of Sciences Symposium on Dynamic Social Network Analysis*, 2002.

- [7] T. Joachims. Making large-scale support vector machine learning practical. In A. S. B. Scholkopf, C. Burges, editor, *Advances in Kernel Methods: Support Vector Machines*. MIT Press, Cambridge, MA, 1998.
- [8] D. Kempe, J. Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD*, pages 137–146, New York, NY, USA, 2003. ACM Press.
- [9] D. Liben-Nowell and J. Kleinberg. The link prediction problem for social networks. In *Intl. Conf. on Information and Knowledge Management*, 2003.
- [10] L. Licamele, M. Bilgic, L. Getoor, and N. Roussopoulos. Capital and benefit in social networks. In *LinkKDD '05: Proceedings of the 11th ACM SIGKDD workshop on Link Discovery: Issues, Approaches and Applications*, New York, NY, USA, 2005. ACM Press.
- [11] G. C. Loury. Intergenerational transfers and the distribution of earnings. *Econometrica*, 49(4):843–867, July 1981.
- [12] A. McCallum, A. Corrada-Emmanuel, and X. Wang. Topic and role discovery in social networks. In *IJCAI '05: Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence*, 2005.
- [13] M. Newman. The structure and function of complex networks. *IAM Review*, 45(2):167–256, 2003.
- [14] J. O'Madadhain, J. Hutchins, and P. Smyth. Prediction and ranking algorithms for event-based network data. *SIGKDD Explorations*, 7(2):23–30, 2005.
- [15] J. O'Madadhain and P. Smyth. Eventrank: A framework for ranking time-varying networks. In *LinkKDD '05: Proceedings of the 11th ACM SIGKDD workshop on Link Discovery: Issues, Approaches and Applications*, New York, NY, USA, 2005. ACM Press.
- [16] C. Palmer, P. Gibbons, and C. Faloutsos. ANF: A fast and scalable tool for data mining in massive graphs. In *ACM Intl. Conf. on Knowledge Discovery and Data Mining*, 2002.
- [17] A. Popescul and L. H. Ungar. Statistical relational learning for link prediction. In *IJCAI03 Workshop on Learning Statistical Models from Relational Data*, 2003.
- [18] A. Portes. Social capital: Its origins and applications in modern sociology. *Annual Review of Sociology*, 24:1–24, 1998.
- [19] H. C. White, S. A. Boorman, and R. L. Breiger. Social structure from multiple networks: I. blockmodels of roles and positions. *American Journal of Sociology*, 81:730–779, 1976.