Capital and Benefit in Social Networks

Louis Licamele, Mustafa Bilgic, Lise Getoor, and Nick Roussopoulos
Computer Science Dept.
University of Maryland
College Park, MD
{licamele,mbilgic,getoor,nick}@cs.umd.edu

ABSTRACT

Recently there has been a surge of interest in social networks. Email traffic, disease transmission, and criminal activity can all be modeled as social networks. In this paper, we introduce a particular form of social network which we call a *friendship-event network*. A friendship-event network describes two inter-related networks. One is a friendship network among a set of actors. The other is an event network that describes events, event organizers and event participants. Within these types of networks, we formulate the notion of *capital* based on the actor-organizer friendship relationship and the notion of benefit, based on event participation. 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 incorporate a temporal component by considering the notion of an event series. We explore the use of these measures on a data set describing three computer science conferences over the past ten years.

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. 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. Most of the research in social networks has focus on structural aspects of the networks.

In this paper we will look at networks that are a bit 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. Event networks include information about 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* (FEN).

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 the greater the overall social capital of 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 FEN.

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. We also have conference events along with their organizers (program committee (PC) members) and participants (authors) together which we will refer to as academic collaboration FEN. In this example dataset, a friend is defined as the people an author shares a co-authorship relation with, 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 publications that an author publishes in a conference.

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

LinkKDD'05 August 21, 2005 Chicago, IL, USA.

further extend our definitions with the important element of time. Finally, in Section 6 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 [11] 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, 7]). 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. [7] 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 [10]. Liben-Nowell and Kleinberg [8] attempt to predict future interactions between actors using the network topology. In addition, Palmer et al. [12] 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, 9, 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 [13]. 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; position refers to subsets of actors who have similar ties to other actors, and role refers to patterns of relationships between these actors or subsets [14].

3. THE FRIENDSHIP-EVENT NETWORK

We begin with a generic description of a family of social networks which we refer to as FEN. A FEN has the following sets of entities:

- **actors:** a set of actors $A = \{A_1, \ldots, A_n\}$
- **events:** a set of events $E = \{E_1, \ldots, E_m\}$

and the following sets of relationships:

- **friends:**

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

- **organizers:**

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

- **participants:**

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

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,

$$se(E_k) = \{e_{k1}, \ldots e_{kp}\},$$

and define a participant subevent relation:

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

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

$$P(E_k, A_i) = \exists\ e_{kj} \in se(E_k) \text{ s.t. } S(E_k, A_i, e_{kj})$$

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 the academic collaboration FEN is given in Figure 1.

4. EVENT-SPECIFIC CAPITAL AND BENEFIT

Next we introduce the notions of capital and benefit. *Personal social capital* is a measurement of the amount of "goodwill" 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 personal social capital of an actor A_i in an event E_k is the number of organizers with whom the actor is friends:

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

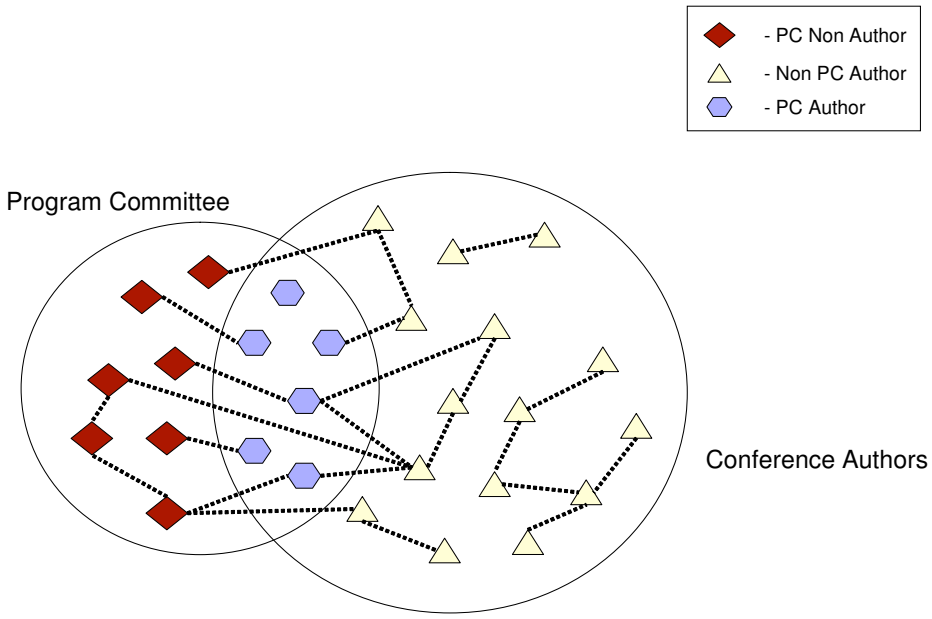


Figure 1: An event in the friendship-event network for academic collaboration. The actors in the network are PC members and authors. The edges in the network indicate co-authorship links (friendship). The organizers are the PC members (the set on the left), and the participants are the authors (the set 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.

where I is an indicator function which is 1 when the relation holds. [1]

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:

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

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

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

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

In the context of the academic collaboration FEN 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 (where a paper with 3 authors counts as 3 paper authorships):

Definition 4. 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.

[1] 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.

Definition 5. 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 6. 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 7. 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 to our FEN.

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 FEN 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 $E_k(t_1), \ldots, E_k(t_q)$

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 $E_k(T)$.

Table 1: For each conference series, the average number of papers, average number of authors and average PC size for the past 10 years.

Conf.	Papers		Authors		PC	
	μ	σ	μ	σ	μ	σ
C1	78.90	9.45	223.20	25.24	32.60	5.87
C2	87.00	23.75	237.70	85.89	69.62	23.30
C3	29.20	2.94	66.30	9.87	9.30	2.87

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 FEN, 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.

Sadly enough, friendships may fade 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 FEN, 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 8. 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. EXPERIMENTAL RESULTS

We explored how these descriptive statistics apply to a real academic friendship-event network. 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 μ and σ 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 is has a more theoretical bent.

Our measures are *not* calculating bias in paper acceptance. There are many reasonable explanations for why there should be correlations in the measures we have defined, for example in certain communities PC members may be more likely

Table 2: Overall aggregate statics for friendship, capital, benefit given (BG) and benefit received (BR).

Conf.	Friendship		Capital		BR		BG	
	μ	σ	μ	σ	μ	σ	μ	σ
C1	8.29	2.50	0.55	0.20	1.16	0.46	4.64	10.28
C2	7.45	1.20	0.71	0.42	1.09	0.34	3.13	8.41
C3	8.37	2.31	0.57	0.29	1.10	0.33	3.15	4.64

to be younger, tenure-track academics under greater pressure to publish, while in other communities PC members may be more senior, with larger and more productive research groups. Further, because we do not have complete data, our results are at best approximations to the measures we have defined. In particular, we do not have information about rejected papers. Additionally, we do not know the reviewer assignments, so when an author submits a paper to a conference, we do not know who reviews her paper. Specifically, we do not know if an author has been assigned a reviewer whom is also a friend. We also do not have access to the reviews, so we do not have a measure of quality assigned to the paper. Even if the author's friends review the paper, we do not know if the paper was accepted because the paper was of good quality or as a result of a favor. Our notion of benefit therefore is *not* capturing unfairness in the reviewing process. Nonetheless, we believe that the notions that we have introduced are useful descriptive measures for friendship-event networks. And as far as we are aware, their quantitative definitions are novel.

Overall aggregate statistics for the conferences are shown in Table 2. Here we are using a friendship window size of 5 years (i.e. $n = 5$). 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. These statistics imply that the same phenomenon is present in all three conferences.

6.1 Role-based Comparison

In order to analyze the data in further detail, we broke the actors into three groups according to their roles in the network. For a particular conference and year, we have the following sets of actors:

- **PC-Authors:** Program committee members who have also published in that conference in that year.
- **PC-Non-Authors:** Program committee members who have not published in that conference in that year.
- **Non-PC-Authors:** Authors who are not in the program committee but published a paper in that conference in that year.

We analyzed the data to see if there are any apparent differences between these groups in terms of either capital or benefit. We began by looking at the different classes of authors, PC-Authors and Non-PC-Authors. Table 3 show the average benefit received for each of these groups. The average benefit received is not significantly different between these two groups. This reflects the fact that most people only have one paper in the conference. It could be interesting to further study how the capital of these two groups effects the number of publications of each author.

Table 3: Average Benefit Received for PC-Authors versus Non-PC-Authors.

Conf.	PC-Author		Non-PC-Author	
	μ	σ	μ	σ
C1	1.38	0.64	1.15	0.45
C2	1.26	0.58	1.08	0.31
C3	1.21	0.45	1.10	0.33

Table 4: Capital for PC-Authors and PC-Non-Authors

Conf.	PC-Author		PC-Non-Author	
	μ	σ	μ	σ
C1	1.29	0.09	0.20	0.05
C2	1.76	0.24	0.27	0.01
C3	0.72	0.10	0.11	0.05

We next shift our focus to determine what makes PC-Authors and PC-Non-Authors different. We start by examining how social capital differs between these groups. The average capital for each group is reported in Table 4. We see that the average capital is significantly higher for PC-Authors compared to Non-PC-Authors. So PC members are more likely to publish in their own conference if they have a lot of friends on the program committee with them. There are several potential explanations for this difference. Perhaps these PC members are benefiting from having many friends on the program committee. On the other hand, it may be that having a lot of fellow PC friends is an indication of how well-suited the conference is to the PC-Author's research. Alternatively, it may be that PC members with a lot of fellow PC friends, are more inclined to submit because they want to attend the conference with their friends. Of course we cannot draw concrete conclusions from this one insight, but it can help to further guide our understanding of this social network.

6.2 Event Series Analysis

As has been shown, the notion of friendship and capital can allow for insights to be made when comparing different conferences. We have been aggregating the values over a

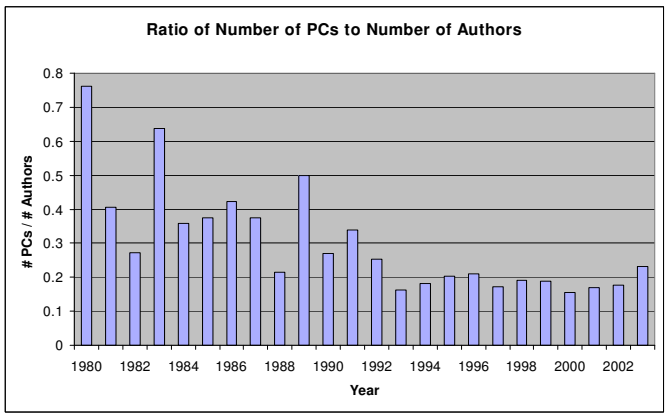


Figure 3: PC Author Ratio

ten year period to allow for a comparison of conferences as a whole, but we can also look at trends in a specific conference. The patterns of a conference over time can be shown by inspecting 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 help for this exploratory data mining process.

One of the first things that someone might be interested in finding out is how the levels of friendship compare between the different categories of actors. This information was calculated for the last 23 years of conference C1 and is shown in Figure 2(a). The temporal trends of the levels of friendship among all of the categories is shown in this figure. 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 is why they are chosen to be on the program committee. In that case, we would expect for the friendship values of all PC members, not just the ones who are authors but also the PC-Non-Authors, to be higher than the friendship of the Non-PC-Authors. The PC-Non-Authors have a slightly higher friendship value than the Non-PC-Authors but it is still a much smaller friendship value than what the PC-Authors have. So we can see that, on average, PC members have more friends than Non-PC members.

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 had an equal chance of being on the program committee, we would expect to find similar patterns in the capital values between the groups as was shown in Figure 2(a). 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. Of course, in many ways this is not surprising because they had the most 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.

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. Maybe this is why they are much less likely to publish in the conference that they are the PC on.

It is hard to draw conclusions from just the changes in friendship and capital alone. It could be possible for outside variables to affect these two values. One possible scenario that would lead to an increase 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. These results are presented in Figure 3. 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.

7. FUTURE WORK

We would like to examine richer notions of friendship. The friendship relationship is currently a boolean feature. If two actors are related, or in the scenario presented here if they have co-authored a paper together, within the last n years, we consider them friends. Alternatively, we can formulate friendship as a function that maps to a real number, which monotonically decreases as the relationship ages without reinforcement. That is, we can consider two authors' friendship to be stronger than another pair's friendship if the former pair had published a paper together more recently. Moreover, we can also take the number of times two authors published together as an indicator of the strength of the friendship as well; the more they publish together, the stronger the relationship they have. It might be interesting to explore the effects of these new formulations of friendship on the statistics.

8. CONCLUSION

We have formulated a general family of friendship-event networks. We have given quantitative definition for social capital, benefit recieved, and benefit given. At this point, our analysis is purely descriptive; we are interested in measures that help us understand friendship-event networks and that allow us to compare different event series. Ideally, these definitions could be used as part of a design process, that would, depending on the context allow us to construct friendship-event networks that would either maximize or minimize benefit. This could be of use for a variety of tasks such as

constructing program committees, assigning reviewers and author networking.

9. 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] 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.

[8] D. Liben-Nowell and J. Kleinberg. The link prediction problem for social networks. In *Intl. Conf. on Information and Knowledge Management*, 2003.

[9] G. C. Loury. Intergenerational transfers and the distribution of earnings. *Econometrica*, 49(4):843–867, July 1981.

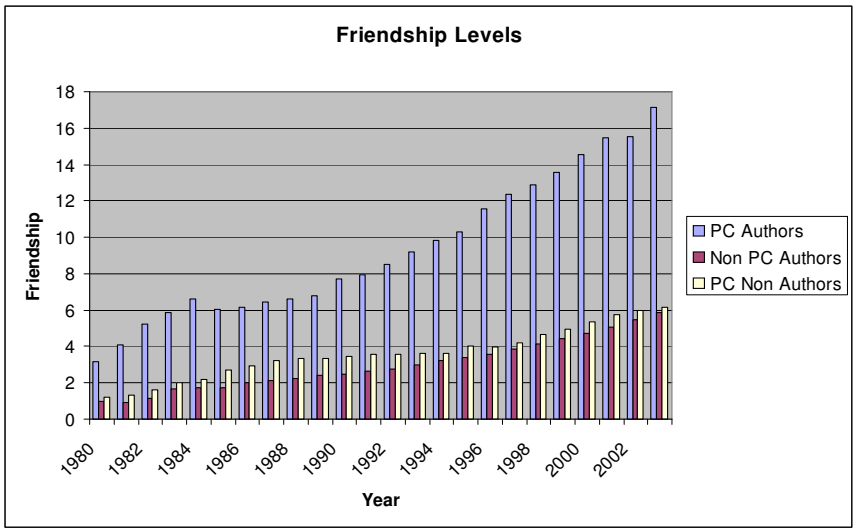
[10] A. McCallum, A. Corrada-Emmanuel, and X. Wang. Topic and role discovery in social networks. In *IJCAI '05: Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence*, 2005.

[11] M. Newman. The structure and function of complex networks. *IAM Review*, 45(2):167–256, 2003.

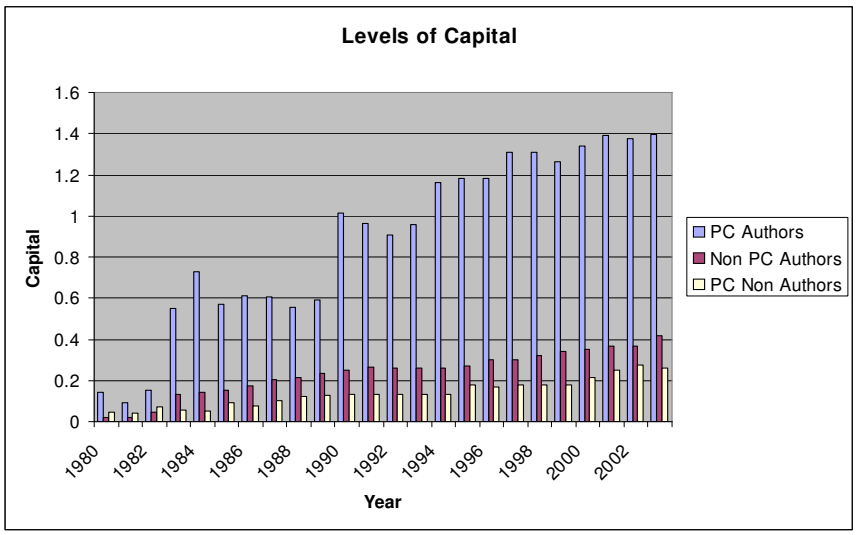
[12] C. Palmer, P. Gibbons, and C. Faloutsos. ANF: A fast and scalable tool for data mining in massive graphs. In *ACM Intl. Conf. on Knowlege Discovery and Data Mining*, 2002.

[13] A. Portes. Social capital: Its origins and applications in modern sociology. *Annual Review of Sociology*, 24:1–24, 1998.

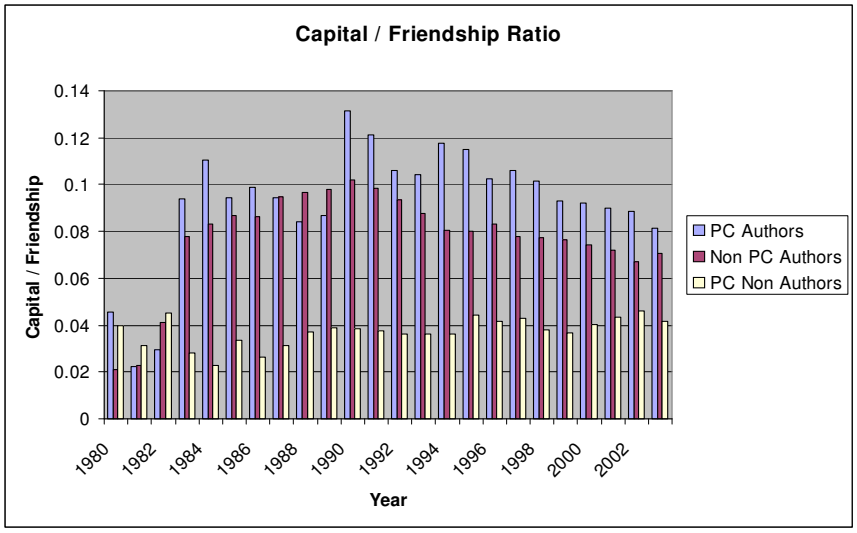
[14] H. C. White, S. A. Boorman, and R. L. Breiger. Social structure from multiple networks: I. blockmodels of roles and positions. *American Journal of Sociology*, 81:730–779, 1976.



(a)



(b)



(c)

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