Social Capital in Friendship-Event Networks

Louis Licamele and Lise Getoor
Computer Science Dept., University of Maryland
College Park, MD 20742 USA
{licamele,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 examine 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. 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 researchers 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 author's 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 co-sponsored 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 the academic collaboration domain is a domain with which we are all familiar, we will continue to focus on it for the rest of this paper.

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, 12, 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 [17]. In our analysis, the "sources of the capital" are the organizers of the events.

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]). Other work, such as Liben-Nowell and Kleinberg [9], attempts to predict future interactions between actors using the network topology.

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 [16] examine citation prediction. Here we look at conference publication prediction, both for a single author and for a collection of authors.

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. The work presented here expands on an earlier, unpublished workshop paper [10].

We begin by giving 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 2. In Section 3, we define capital and benefit and we further extend our definitions with the important element of time. In Section 4 we explain the participation prediction task. Finally, in Section 5 we describe results applying these measures to three different computer science conferences over a 10 year time period.

2 The Friendship-Event Network

We start 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 $A = \{A_1, \ldots, A_n\}$

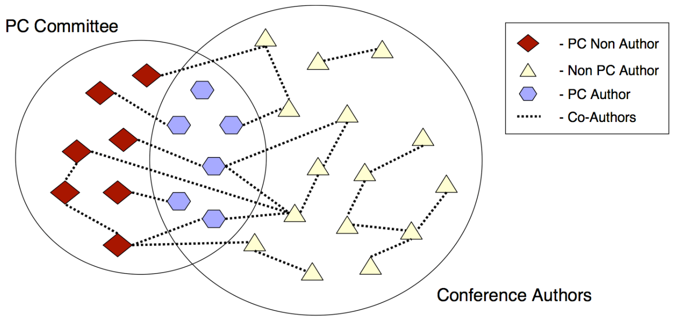

Figure 1. An event in the FEN for academic collaboration. The edges in the network indicate co-authorship links (friendship). The organizers are the PC members (left), and the participants are the authors (right).

events: a set of events $E = \{E_1, \ldots, E_m\}$
and the following sets of relationships:
friends: $F(A_i, A_j) = A_i$ is friends with A_j
organizers: $O(A_i, E_k) = A_i$ is an organizer of event E_k
participants: $P(A_i, E_k) = A_i$ 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 relationship:
$p(A_i, E_k, E_{kj}) = A_i$ is a participant in subevent E_{kj} 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 in a conference. An illustration of an academic collaboration network is given in Figure 1.

3 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. 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. 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. [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 *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.

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

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

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).

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 4 *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.

The preceding definitions consider friendship as a non-temporal relationship. We can 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. Friendships evolve over 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.

4 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

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

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:

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

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 $t-1, t-2, \ldots, t-n$. 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 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.

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.

5 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 to 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.

We calculated simple structural statistics for the three conferences. It turns out that two of the conferences are very similar to each other. while the third conference has a more theoretical bent. We performed a role-based comparison using these descriptive measures and we found several interesting trends across the conferences (see [11] for more details).

5.1 Predictive Analysis

In this section, we evaluate predictive models for event participation. We examine models which make use of par-

Table 1. The accuracy for group event-participation prediction using different measures of the histories is shown.

	Min	Max	Mean	Total	All
Pub_{hist}	52.5	72.3	64.0	75.6	82.7
SC_{hist}	52.2	77.2	70.8	79.2	83.4

ticipation 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 5.1.1) and the groups of actors (Section 5.1.2) in subevents. In 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.

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.

5.1.1 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.

5.1.2 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.

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.

Table 1 shows the results for different measures for the group publication history. We examined not only the average 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, max-

Table 2. Group event-participation prediction accuracy is shown.

SC_{cur}	SC_{hist}	SC	PUB	$PUB+SC$
64.88	83.01	83.39	82.67	88.70

imum social capital and total social capital. Table 1 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%. 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 $p < 0.05$).

Next we explored the importance of using the current social capital as compared to using the social capital history. Table 2 shows that using only the current social capital alone (SC_{cur}), is not a very good predictor. The social capital history, SC_{hist}, (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, and we also show the result of publication history together with the combined social capital measures, $PUB+SC$. Combining both publication history and social capital gives us the best performance, 88.69%.

An issue that comes up in performing this analysis is the question of how to correctly select the appropriate window for defining friendship as well as for determining social capital and publication histories. For a more in depth analysis of friendship windows as well as evaluations of more complex definitions of friendship the interested reader is directed to [11].

6 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 conferences 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.

Acknowledgements: This work was supported by the National Science Foundation under NSF #0423845 and NSF #0438866.

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 ACM SIGKDD*, 2001.

[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 ACM SIGKDD*, 2003.

[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 *Proceedings of ACM SIGKDD workshop on Link Discovery*, 2005.

[11] L. Licamele and L. Getoor. Social capital in friendship-event networks. *University of Maryland Technical Report CS-TR-4828*, 2006.

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

[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 *Proceedings of ACM SIGKDD workshop on Link Discovery*, 2005.

[16] A. Popescul and L. H. Ungar. Statistical relational learning for link prediction. In *IJCAI03 Workshop on Learning Statistical Models from Relational Data*, 2003.

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