Mining dynamic social networks from public news articles for company value prediction

4 authors, including:

Ching-Yung Lin
Columbia University
141 PUBLICATIONS **6,285** CITATIONS

Mitsuru Ishizuka
Waseda University
529 PUBLICATIONS **10,900** CITATIONS

Mining dynamic social networks from public news articles for company value prediction

Yingzi Jin · Ching-Yung Lin · Yutaka Matsuo · Mitsuru Ishizuka

Abstract Dynamic networks are studied by sociologists to understand network evolution, belief formation, friendship formation, etc. Companies make and receive different impacts from other companies in different periods. If one can understand what types of network changes affect a company's value, then one would be able to predict the future value of the company, grasp industry innovations, and make business more successful. However, it is often difficult to collect continuous records of network changes, and the models of mining longitudinal network are complicated. In this study, we developed algorithms and a system to infer large-scale evolutionary company networks from public news during 1981–2009. Then, based on how networks change over time, and on the financial information of the companies, we predicted company profit and revenue growth. Herein, we propose a feature extraction and selection algorithm for longitudinal networks. This paper is the first to describe a study examining longitudinal network-mining-based company performance analysis. We measured how networks affect company performance and what network features are important.

1 Introduction

Social network analysis (SNA) examines relations and structures of actors in a network to measure an actor's behavior and performance (Wasserman et al. 1994). For example, some companies form numerous alliances with many small companies, whereas others make few collaborations, but with large companies only. The partners of some companies are mutually well connected, whereas those of other companies are connected only with the target company. These relations and structural embeddedness influence the behavior and performance growth of companies. Network structural features such as degree, betweenness, and closeness centralities have been used in studies reported in the literature to analyze actors' performance and value (Uzzi 1997). Several researchers have generated and selected useful network features automatically to diversify a different perspective of embeddedness of actors (Karamon et al. 2008; Backstrom et al. 2006).

Social networks are dynamic in nature. Relations between companies differ among periods. Companies continually establish and eliminate relations with others, and continually receive positive or negative impacts from other companies. Consequently, their performance changes with time. If one can understand useful effects of longitudinal networks of companies as well as the mechanism responsible for changing the company value, then a person would be able to infer a company's future value, decide strategic relational management for the company, and analyze the entire company's growth. Longitudinal networks are

Y. Jin (✉) · Y. Matsuo · M. Ishizuka
The University of Tokyo, Tokyo, Japan
e-mail: yzjin2006@gmail.com;
eiko-kin@biz-model.t.u-tokyo.ac.jp

Y. Matsuo
e-mail: matsuo@biz-model.t.u-tokyo.ac.jp

M. Ishizuka
e-mail: ishizuka@i.u-tokyo.ac.jp

C.-Y. Lin
IBM T. J. Watson Research Center,
Yorktown Heights, USA
e-mail: chingyung@us.ibm.com

studied by sociologists to understand network evolution, belief formation, friendship formation, and so on (Wasserman et al. 1994; Doreian et al. 1997; McCulloh and Carley 2009).

As described in this paper, we explore a new analytical paradigm that uses large-Keywordsscale representation of a longitudinal network of companies to predict a company's performance. Our goal is to find answers to the following research questions: Is it possible to predict a company's value (such as revenue and profit) based on dynamic (i.e. longitudinal) company networks? How can we infer evolutionary company networks? What features of a longitudinal network are useful for a company, and how can they be generated? Two challenging algorithms are proposed in this paper. They are used for extraction of longitudinal inter-company networks from public news and for mining longitudinal networks for future value prediction.

Regarding the first question, we develop a simple algorithm for temporal company network mining from public news. We specifically examine determination of the *impact* of relations that a company shares with other companies via news articles, and use document and sentence co-occurrence to extract impactors for each target company to construct valued directed intercompany networks over a period of years. With respect to the second question, we propose to investigate network effects related with company value from longitudinal impact relational networks. We generate network effects from local and global relations, historical relations, and delta-change in the relations for each target company. We also investigate effects of a network from each directed/undirected, valued/unvalued longitudinal network. Additionally, we discuss positive and negative structures of effects (as network features) and make feature selection based on the variance and correlation of features. Then, we propose the use of an existing machine learning algorithm such as linear regression and SVM regression to combine the features of the longitudinal network with a company's financial information to predict the company value. Experimentally obtained results show that our prediction model captures the trend of changes in the value of group companies or in an individual company's value over the years. Company profit prediction by joint network and financial analysis outperforms network-only by 150% and financial-only by 34%. Results also show that network features did not contribute to revenue prediction. The evolution of company networks over the years is apparent with different structural characteristics.

This paper describes the first study predicting company performance based on longitudinal inter-company networks. The proposed algorithms are applicable not only to company domains, but also to people, products, and web documents ranking (or value) prediction from their dynamic networks.

This paper is organized as follows. The following section presents a description of some related studies. Section 3 presents the problem statement and introduces the system outline. Section 4 proposes a method for extracting relations among companies from public news. Section 5 proposes a method for mining networks longitudinally to develop a value prediction model. Section 6 presents empirical results of the study and multivariate analysis. Section 7 concludes the study.

2 Related work

In the literature, most prior approaches to the prediction of company valuations have fallen into three categories. The first type of approach (designated as a financial approach) is based on a company's financial statement (e.g., return on assets, capital ratio, number of employees) to measure the company's future earnings and performance (Bengtsson and Kock 1999; Xiao et al. 2009). The second type of approach (a technical approach) is application of historical trends to identify price patterns and trends and to exploit those patterns to predict the direction of company valuations–stock prices (Wang and Chan 2007; Yang et al. 2002). The third type of approach (named social network analysis, SNA) examines relational and structural embeddedness of companies on intercompany networks from positional characteristics (Wasserman et al. 1994; Uzzi 1997). This study uses the third approach presented above, but both historical and financial information are combined.

Ranking entities based on network structure have often been used in the information retrieval (IR) field. For example, Pagerank (Page and Stanford 1998) and HITS (Kleinberg 1999) are well known algorithms for computing the importance of web pages. They rank web pages, respectively, based on the Markov chain model and authority-hub model, and both are unsupervised learning. Pagerank defines a voting mechanism for links: each link is regarded as a vote by a page for another page to which it links. Our algorithm is supervised learning. We rank companies by their network structure using longitudinal network features. Learning to ranking algorithms in the IR field is also popular. Some use relational information (Qin et al. 2008; Cao et al. 2007) for ranking objects, but relational definitions among web pages (e.g., similarity, parent–child, etc.) have different meanings with our social ties among companies.

Regarding automatic extraction of the intercompany networks, one stream of studies specifically examines specific relations among companies such as alliance, cooperation, and acquisition (Jin et al. 2007; Hu et al. 2009; Xiao et al. 2009; Ben-Zvi et al. 2009). However, specific relational networks lasting over time are usually sparse.

Therefore, Hu et al. (Hu et al. 2009) sum specific relations of six types between companies to construct a comprehensive temporal relational network. Another avenue of research has explored the use of co-occurrence approaches. They use co-occurrence of names on the web (Katz and Proctor 1959; Mika 2005; Matsuo et al. 2006) or in public news (Bernstein et al. 2002; Bao et al. 2008) to measure the relational strength between actors. Bao et al. observed that a company is more likely to co-occur with its competitors on web pages (i.e. document-level co-occurrence) than with noncompetitors. In this study, we do not extract specific relations separately; we are more interested in impact relations, i.e., how many impacts a company receives from others. We do not discuss positive/negative relations, but we consider positive/negative structural impacts from networks.

Analysis of over time network data has been presented in the social science literature (Kautz et al. 1997; Doreian et al. 1997; Snijders et al. 1997). Longitudinal network analysis is used to elucidate network evolution, belief formation, friendship formation, etc. (Leenders 1995a; Feld 1997; Snijders et al. 1997; Xiao et al. 2009; Ben-Zvi et al. 2009). Markov chain models, multi-agent models, and statistical models have been applied for mining network evolution and group dynamics (McCulloh and Carley 2009). However, few studies of longitudinal network analysis have addressed intercompany networks because the relations among companies are complex and unspecific; moreover, it is difficult to track companies' network changes over time. Some studies have focused only on a specific relation (e.g. alliance), or, using self-report data, simulation data with time. A common complaint is that scalability suffers and incomplete information problems arise (Xiao et al. 2009, Ben-Zvi et al. 2009). This study of longitudinal network-mining-based company performance analysis is the first reported in the literature.

Several researchers have used network-based features for analyses (Backstrom et al. 2006; Liben-Nowell and Kleinberg 2007). Backstrom et al. (2006) describe analyses of community evolution and show some structural features characterizing individuals' positions in a network. Liben-Nowell et al. (2007) elucidate features using network structures for link prediction in the link prediction problem. Our generated features include those described in reports by Backstrom and Liben-Nowell. We specifically examine relations and structural features for individuals and systematically address various structural features from longitudinal networks.

3 Problem statement

Given a set of companies V and a period time T with a data source D, the system will first extract longitudinal inter-company networks in each period $\mathcal{G}^T = \{G^{t_1}, G^{t_2}, \ldots, G^{t_k}\}$, where $t_1 < t_2 \ldots < t_k$ (detail in Sect. 3). Then, for each focal company $x \in V$, we generate and select a structural feature vector $\mathbf{F}_x^T$ from its embeddedness in the longitudinal networks $\mathcal{G}^T$, and use these feature vectors to learn and to predict company valuations y_x (in Sect. 4).

We collected news articles from the *New York Times* (*NYT*)[1] during 1981–2009. Indexable company names are listed in the *New York Times*.[2] The company valuations are obtained from the Fortune 500 list published by Fortune magazine[3] from 1955 to the current year, which lists the revenues and profits of companies every year. Because we need continuous records of company valuations over several years, we select as target companies only those companies that have appeared on the Fortune list at least three times.

4 Extraction of longitudinal network from public news

Longitudinal social network data have been collected using questionnaires, interviews, observations, and so on (Wasserman et al. 1994). With extensive public data available on the web, in public news, and through electronic media, many researchers are interested in extracting social networks automatically from large scalable data. Public news sites (e.g., *Wall Street Journal*, *New York Times*) broadcast company news daily. News articles include a title, content, and publishing time, and therefore constitute a good resource for extracting longitudinal inter-company networks. For example, *IBM* appeared in about 300 news articles in the *New York Times* in 2009 (277 articles as *IBM* and 84 articles as *International Business Machines*). If one could read and remember all the news stories, general knowledge about the company would be clarified—which companies made an *impact* on *IBM*, and by how much?

Several studies have specifically examined large volumes of public news articles to extract valuable information such as risk statements and future earnings of companies (Tetlock et al. 2008). Bao et al. reported that a company is more likely to co-occur with its competitors on web pages (i.e. document-level co-occurrence) than with non-competitors. In this study, we extract longitudinal impact networks among companies by mining *New York Times* articles published during 1981–2009. Companies receive different degrees of impact from different companies

---

[1] http://www.nytimes.com/.

[2] http://money.cnn.com/magazines/fortune/fortune500/2010/full_list/ (7,594 company names are listed).

[3] http://money.cnn.com/magazines/fortune/fortune500/2010/.

during the year. Therefore, the network is a directed valued network. This makes our task unique.

As described in this paper, we propose to use document-level and sentence-level co-occurrence to construct longitudinal *impact networks* from public news (i.e. *New York Times*). Our assumption for the impact relation is that if a company frequently has co-appeared in the target company' important news articles and has been described frequently together with the target company in important sentences over a period of time, the company will make a large impact on the target company in that period. Therefore, they are regarded as having strong mutual relations. We propose the use of document-level and sentence-level co-occurrence to measure the frequency, and to assign weight to each document and sentence to measure the importance.

We use *New York Times* articles from 1981–2009, and we set 1 year (Jan. 1–Dec. 31) as the period during which we can extract longitudinal networks year by year.[4] We have a company name list from the New York Times (7,594 companies can be indexed).[5] Because we require continuous records of company activities during years, we select only those large companies that have appeared on the Fortune list[6] at least three times and indexable form NYT articles as target companies. Therefore, we can match the network period with the obtainable company valuations.

Details of our algorithm are the following. For each target company x, we score candidate companies[7] Y by their impact to x in a period t. First, for each candidate company $y \in Y$, we collect a document set $D_{x,y}^t$ and a sentence set $S_{x,y}^t$, in which it has co-occurred with the target company x during period t. Then, we sum up each of those document-weight $w_d(i)$ and sentence-weight $w_s(j)$ to calculate the final relational score for each y related to x as follows:

$$\text{score}_x(y) = a \cdot \sum_{i \in D_{x,y}^t} w_d(i) + b \cdot \sum_{j \in S_{x,y}^t} w_s(j). \tag{1}$$

As described in this paper, we use the following equations to assign importance in terms of weight to each co-occurring document and sentence.

$$w_d(i) = \log\left(1 + \frac{1}{|Y'(i)|} + \frac{tf_x(i)}{\sum_{y \in \{x,Y\}} tf_y(i)}\right) \tag{2}$$

$$w_s(j) = \log\left(1 + \frac{1}{|Y''(j)|}\right) \tag{3}$$

In those equations, $Y'(i)$ and $Y''(j)$ respectively denote the company names from document i and sentence j. $|Y'(i)|$ and $|Y''(j)|$ are counts of those names, and $tf_y(i)$ is the frequency of name y appearing in a document. Intuitively, if a document includes many company names, then it will be less important for those two companies than a document that mentions only a few companies. In addition, the sentence weight is high for companies x and y if it mentions only two companies, and low if it lists many companies. Constants a and b represent a tradeoff between the document weight and sentence weight. Heuristically, we set $a = 1$, $b = 5$.[8]

Finally, we obtain longitudinal inter-company networks year by year. Figure 1 compares the evolution of ego networks in different years. From *IBM* evolution networks between 2003 and 2009, some companies such as *Motorola*, *Novell*, and *NEC* were listed on the top in 2003, but disappeared from the network in 2009. Instead of those, *Google*, *SPSS*, and *Xerox* newly appeared on the network of *IBM*. *Microsoft*, *HP*, and *Sun* remained on the network, but they slightly changed their relational strength with *IBM*. From *Microsoft* evolution networks in 1995, 2003, and 2009, we found that their top-related companies changed from *Intuit Inc.*, *IBM* to *Google*, and the relational strengths for top companies also changed slightly during these years.

Here we give examples for *IBM* in 2009. From Table 1, it is apparent that *Microsoft* had the greatest impact on *IBM* in 2009. They co-occurred in 55 articles and were described together in 264 sentences. From these sentences, we can infer that they are direct competitors. Many top-ranked companies are competitors of *IBM*, which made a big impact on *IBM* in 2009. As described in this paper, we do not categorize relations as negative or positive ones because a company might have many large competitors because it has a high demonstrated performance. Sometimes impact

---

[4] We can also set the extraction period as six or three months to construct longitudinal networks of various types. To match the network period with the obtainable company valuations (i.e., the annual revenues and profits from Fortune list 500), we set 1 year as the period for analyses described in this paper.

[5] For the name alias (e.g., *IBM* and *International Business Machines*), we use a unified name.

[6] http://money.cnn.com/magazines/fortune/fortune500/2010/.

[7] Candidate companies are those companies which appear in news articles about x for that year.

[8] (Actually, we use different sets of a and b, and compare the results by user questionnaire. A user study was conducted on IBM employees to verify the accuracy of extracted networks. Consequently, $a = 1$ and $b = 5$ receive the best votes. We asked, "*Please compare the three different ranked lists of IBM's 'closest companies' for each year.*)" We list 5 years' (1991, 1996, 2001, 2006, and 2009) ranking results for each participant. We have 33 responses from 159 votes (from 33×5 votes) during one week (in fact, from 58 participants, we chose 33 responses who completed the user study. Most of them are senior employees in IBM who have considerable knowledge about IBM's historical relations). The proposed ranking b yielded best results from the response, which obtained 68 votes (i.e. 42.8%) as the best ranking, 49 votes (i.e. 30.8%) as the second ranking, and the remainder as the third ranking.
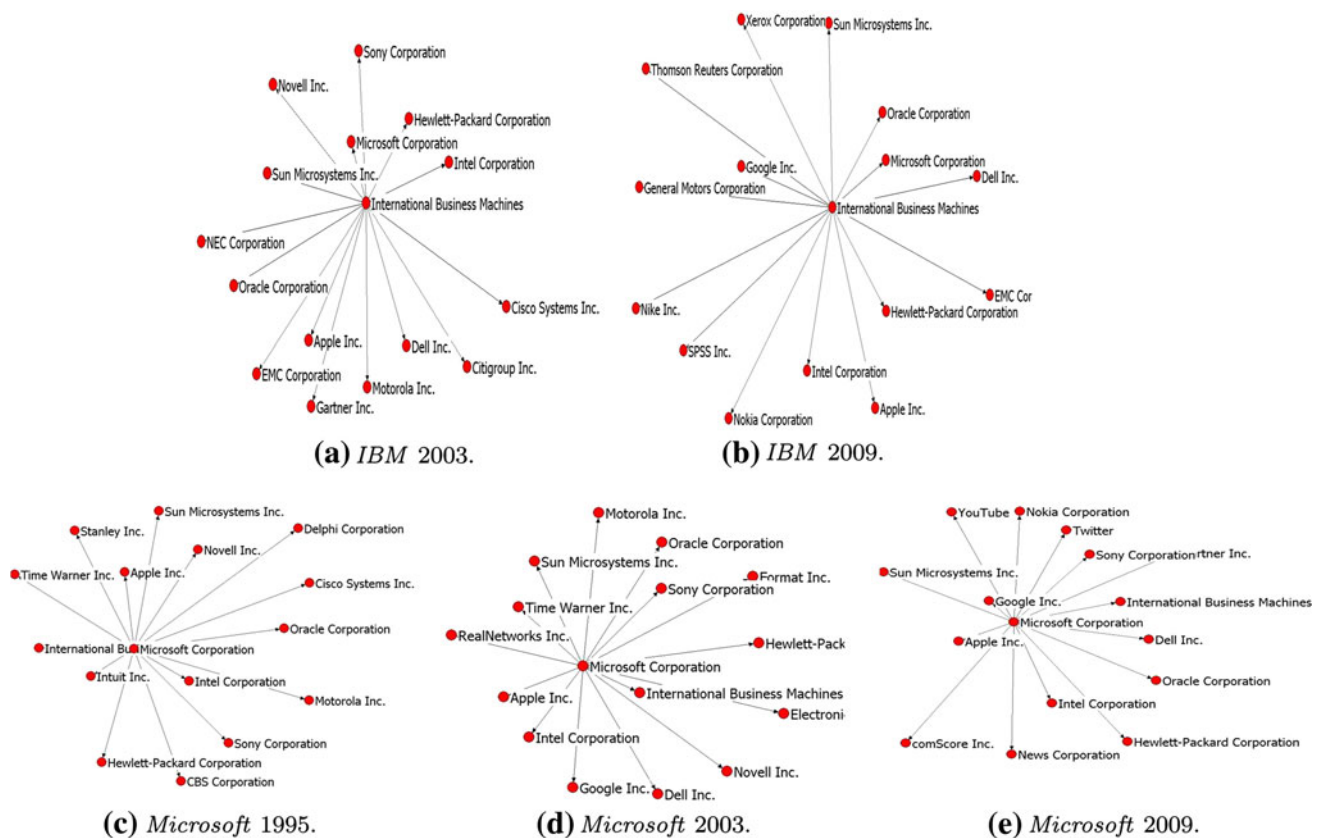
Fig. 1 Evolution of networks in different years

relations are not described in many articles. For example, *SPSS* and *IBM* are not competitors and co-occurred in only one article. They were described together in three sentences, but their relation is important because *SPSS* and *IBM* co-appeared in an article in a high-weight document (which describes only *SPSS* and *IBM*'s acquisition relation in the entire article), and that they are described together in high-weight sentences (in which they are closely described, and in which no other companies appear). *Nike* and *IBM* are not competitors, and have no specific relations, but they were described together because they took similar action or came together for one product. Therefore, they might also exert impact on each other. Consequently, our algorithm can extract competitors, specific relations, and other relations that might have some effect on the focal company, i.e. impact relations.

As described herein, we do not categorize relations as negative or positive because a company might have many competitors for the reason that it has high demonstrated performance. Our intercompany networks are extracted based on a statistical count obtained from news articles about companies over a long period. Therefore, they are

suitable for predicting a company's long-term value change. Prediction of short-term changes (e.g. market price) might require the use of an additional algorithm. We will work to develop such an algorithm in the future.

5 Mining longitudinal network predicting value

After constructing a dynamic impact relational inter-company network, we measure how networks change over time as well as the value-changing mechanism, to predict the company's future value. We first extract net-work effects for each focal node in each period, make feature selection for choosing useful network effects, and then use those features to learn and predict the company value.

5.1 Step 1: network effect generation

First, we calculate network effects for each target node x by its embeddedness in longitudinal *impact* networks $\mathcal{G}^T$.

Table 1 Example of generic relation extraction for IBM in 2009

r	Name	Score	Examples of documents and sentences
1	Microsoft	85.85	IBM, Microsoft (55 articles, 264 sentences, score = 85.85455)
2	Oracle	65.49	http://www.nytimes.com/2009/03/06/business/06layoffs.html
3	Google	57.75	Two days after IBM's report, Microsoft said that its quarterly profits were disappointing
4	HP	50.70	http://www.nytimes.com/2009/01/31/business/31nocera.html
5	Intel	48.52	Caterpillar, Kodak, Home Depot, I.B.M., even mighty Microsoft: they are all cutting jobs
6	Dell	32.75	IBM SPSS (1 articles, 9 sentences, score = 13.675)
7	Sun	29.45	http://www.nytimes.com/2009/07/29/technology/companies/29ibm.html
8	EMC	15.16	IBM to buy SPSS, a maker of business software
9	Apple	14.65	IBM's $50-a-share cash offer is a premium of more than 40 percent over SPSS's closing stock price…
10	SPSS	13.67	IBM, Nike (4 articles, 9 sentences, score = 8.212)
11	GM	13.18	http://www.nytimes.com/2009/01/22/business/22pepsi.html
12	Xerox	12.13	A list of companies that have taken steps to reduce carbon emissions includes IBM, Nike, Coca-Cola and BP, the oil giant
13	Nokia	8.95	
14	Nike	8.21	http://www.nytimes.com/2009/11/01/business/01proto.html
15	Thomson Reuters	7.78	Others are water-based shoe adhesives from Nike and a packing insert from IBM

We use a vector $\mathbf{F}_x^T$ to indicate multi-dimensional network effects for x, which includes current network effects, historical network effects, and the delta-change of effects.

The current network effect (denoted as $\mathbf{F}_x^t$) for the target node x is generated based on the idea from (Karamon et al. 2008) as follows. First, we define a node set N_x for x that might exert impact on x directly or indirectly. Then we define node pairs of three types among N_x: $\langle x, i \rangle$ (in which $i \in N_x$), $\langle i, j \rangle$ (in which $i \in N_x, j \in N_x, i \neq j$), and $i\langle i, k \rangle$ (in which $i \in N_x, k \in V$). We conduct basic operations[9]— *connectivity* $\beta(i, j)$ (returns 1 if i and j are reachable; 0, otherwise), *distance* $\mu(i, j)$ (returns distance between i and j.), and *betweenness* $\zeta^x(i, j)$ (returns 1 if the shortest path between i and j includes x; 0, otherwise)—for these node pairs, and take the sum and standardize those valuations by the network size $|V|$, to compare effect valuations across networks. Finally, we obtain basic network effects of x six types as the following list.

- $\sum_{i \in N_x} \beta(x, i)/(|V| - 1)$, which means the number of connections that x has.
- $\sum_{i \in N_x} \mu(x, i)/(|V| - 1)$, which signifies the distance between x and its related nodes.
- $\sum_{k \in V} \beta(i, k)/(|V| - 1)$, which denotes the number of connections that nodes related to x have.

- $\sum_{i,j \in N_x} \beta(i, j)/(|V| - 1)(|V| - 2)$, which represents the number of connections among x's related nodes.
- $\sum_{i,j \in N_x} \mu(i, j)/(|V| - 1)(|V| - 2)$, which means the distance between x's related nodes.
- $\sum_{i,j \in N_x} \zeta(i, j)/(|V| - 1)(|V| - 2)$, which is the number of node pairs having x on the shortest path.

We consider node set N_x, which might exert an impact on x by a *neighboring node set* L_x (i.e. directly connected) and a *reachable node set* G_x (i.e. indirectly connected) of x. In addition, the difference of impact from local and global node sets is important. For example, the ratio of connections with x between L_x and G_x sets indicates the degree to which companies are related directly with x rather than indirectly. Furthermore, from the constructed valued directed network, we reduce the direction and weight information to generate networks of different types. For example, if we retain only the direction and ignore the weights, then the binary-directed network will show who exerts an impact on whom, but it will not show how much impact is exerted. This arrangement resembles that of a friendship network (e.g., *Facebook* or *linkedIn*), we only consider who treats whom as friends; we do not know how strong the friendship is. Therefore, by considering local and global impacts and their ratio, as well as networks of four different types (i.e., valued/unvalued, directed/undirected), we can generate a $72 = (3 \times 4 \times 6)$-dimensional network effect vector for each target node x in the current network, i.e. $\mathbf{F}_x^t = F(N_x, d, v, t)$, where $N_x \in \{L_x, G_x, L_x/G_x\}$, $(v, d) \in \{(0, 1)\} \times \{0, 1\}$, and $t \in T$.

---

[9] basic operators can be extended to consider triad relations (Wasserman et al. 1994) and more to expand the basic effect vector in the future.

After we have generated network effects from the current network, we further generate historical network effects $\mathbf{F}_x^H$ by considering temporal information, i.e., $\mathbf{F}_x^H = \{\mathbf{F}_x^{t-1}, \mathbf{F}_x^{t-2}, \ldots, \mathbf{F}_x^{t-w}\}$, where $\mathbf{F}_x^{t-w}$ indicates the network effects from the network that existed w years prior, which implies a historical network impact exerted by other companies. In addition, the amount of change over time is considered: $\Delta\mathbf{F}_x^H = \{\Delta\mathbf{F}_x^{t-1}, \Delta\mathbf{F}_x^{t-2}, \ldots, \Delta\mathbf{F}_x^{t-w}\}$. For example, we can examine the delta-change in neighboring nodes from last year to this year, or delta-changes from 3 years prior.

Consequently, for each company for each year, we generate 72-dimensional current-year network effects $\mathbf{F}_x^t$, plus $72\times$ *window size*-dimensional historical network effects $\mathbf{F}_x^H$, and plus $72\times$ delta size-dimensional network effects $\Delta\mathbf{F}_x^H$.

$$\mathbf{F}_x^T = \{\{\mathbf{F}_x^t\}, \{\mathbf{F}_x^H\}, \{\Delta\mathbf{F}_x^H\}\}. \tag{4}$$

In addition, our prediction model can combine effects of historical financial statements of companies, such as the previous year's profit and the profit earned 3 years prior. We indicate those effects as $\mathbf{P}_x^H$ and $\Delta\mathbf{P}_x^H$.

5.2 Step 2: network feature selection

Some automatically generated features (i.e. network effects in this paper) have dependency and redundancy problems. Feature selection can help enhance accuracy in many machine learning problems and improve the efficiency of training (Blum and Langley 1997; Geng et al. 2007). As described in this paper, we consider three processes of feature selection: *individual feature selection*, *feature variance*, and *feature set selection*.

5.2.1 Individual feature selection

Because not all the network features that we have generated contribute positively to prediction of company valuations, we first score each feature by its relation with the target ranking, and remove unimportant features. We rank companies by each network feature f_i and represent the rank vector as $\mathbf{X}_i$. We also rank companies by their valuations (e.g. profit), represented as $\mathbf{Y}$. If the feature f_i is important to a company value, then the correlation between $\mathbf{X}_i$ and $\mathbf{Y}$ will be high. We obtain each feature's relation with a company value by measuring Spearman's correlation between X_i and Y, $\rho_i = 1 - \frac{6\sum d_k^2}{n(n^2-1)}$, where n denotes the number of companies in each observed dataset, and d_k represents the difference between the ranks of corresponding valuations from X_i and Y.
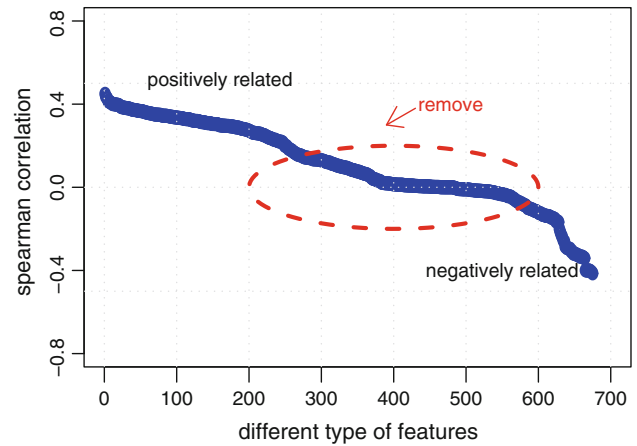
Fig. 2 Positive, negative, and unrelated network features

Results show that positive, negative, and unrelated features exist (Fig. 2). For example, *delta-change in the ratio of x's neighbors and reachable nodes in binary undirected network with 3 years prior* (i.e. one feature in $\Delta\mathbf{F}^H$) has a positive correlation with the company profit. The salient implication is that if there is an increase in the ratio of the number of connections that a company has with the numbers of connections that its neighbors have, then the value of its profits will increase. Another example of a negative feature is number of connections in the previous year (one of the features in $\mathbf{F}^H$), which implies that if x's partners were mutual partners last year, company x is not in a good position. These structures as "influences" affect the target company' value through a network. Our mechanism reveals these network effects (feature or feature set) to predict a company's future value. Table 2 lists some examples of highly positive and negative features.
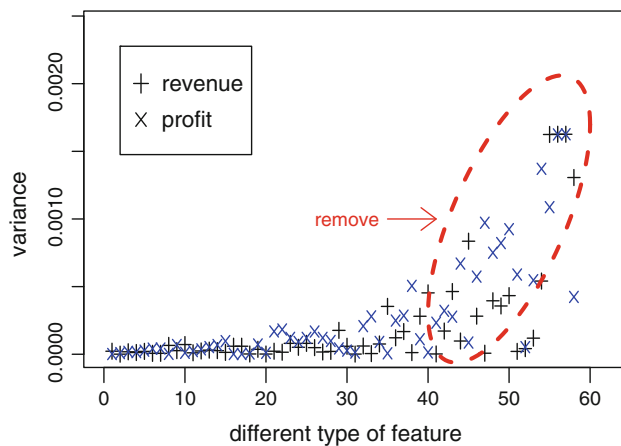
5.2.2 Network feature variance

We tune a network with a threshold, but some features will depend very sensitively on the existence of a particular edge. We measure feature variance in networks constructed with different thresholds, as $\sigma_i^2 = \frac{\sum_{j\in K}(x_{ij}-\mu)}{K}$, where K denotes the number of networks with different thresholds. In addition, we set x_{ij} as ρ_i in the j-th network, which measures the relevance of a feature with the target ranking in the j-th network, and μ represents the mean of relevance values of the feature in various networks. For example, *sum of the number of connections of x's neighbors* is a high-variance feature because it will vary remarkably depending on whether or not an edge with a highly connected neighbor is retained. Figure 3 shows feature variances in networks with a different threshold. We remove those highly sensitive features present on the right side.

Table 2 Example of longitudinal network features related positively or negatively with company value

Correlation	Positively related features
0.421	Delta-change in the ratio of x's neighbors and reachable nodes in binary undirected network with 3 years prior
0.421	Delta-change in x's degree in binary undirected network with respect to its value 3 years prior
0.420	x's degree of 3 years prior in binary undirected network
0.413	x's current year's degree in binary-undirected network
0.413	Ratio of the quantities of x's neighbors and reachable nodes in a binary undirected network
0.353	x's 2 years prior in-degree in weighted-undirected network

Correlation	Negatively related features
−0.487	Prior year's connections among x's neighbors in binary undirected network
−0.477	Delta-change in the sum of degree of x's neighbors in binary undirected network with 2 years prior
−0.462	Prior year's connections among x's neighbors in valued undirected network
−0.381	Ratio of the number of connections among x's neighbors and reachable nodes in valued undirected network
−0.379	Prior year's ratio of the number of connections between x' neighbors and reachable nodes in a valued undirected network

Fig. 3 Feature variance in networks having a different threshold

5.2.3 Feature set selection

When we use only one feature to learn and predict the company value, we should select individual features that have a high correlation with the target ranking. The correlation is calculated between lists of observed data scored by the feature value and the company value (e.g., profit). A feature by itself (e.g. a centrality) might have little correlation with the target ranking, but when it is combined with other features, they may be strongly correlated with the target ranking (Zhao and Liu 2007). Therefore, it is better to combine several features in the prediction model. This technique is similar to the classification or ranking used in the field of data mining (e.g. document classification), wherein entities are not classified or ranked solely by one feature or by all features (one word, or all words) because the former classification would cause a sparsity problem (for) and the latter would cause a redundancy problem. We apply feature selection models from Geng et al. (Geng

et al. 2007) to select a feature set, which is based on the concept that selects features having the largest total importance and smallest total similarity scores.

$$\max \sum_i w_i f_i - c \sum_i \sum_{j \neq i} s_{i,j} f_i f_j$$
$$s.t. \quad f_i \in \{0, 1\} \; i = 1, \ldots, m, \text{ and } \sum_i f_i = t \tag{5}$$

Therein, t denotes the number of selected features, $f_i = 1$ (or 0) shows that the feature f_i is selected (or not), w_i is the importance score of the feature f_i, and $s_{i,j}$ represents the similarity between the features f_i and f_j. Here we let $s_{i,j}$ as Spearman's correlation coefficient between the lists scored by f_i, and f_j, and $s_{i,j} = s_{j,i}$. The objective function is to maximize the sum of the importance scores of individual features, and to minimize the sum of similarity scores between any two features. After generating and selecting feature sets, we use them to predict the company value.

5.3 Step 3: prediction model

After we have generated longitudinal network effects for each target node x, we integrate those valuations as features to learn and predict the future value of the company.

$$y_x^{t'} = f(\mathbf{F}_x^t, \beta) = \sum_k f_k \, \beta_k \tag{6}$$

Therein, $t' > t$, and f_k is the k-th effect from the historical network, and β_k is the importance of f_k. The prediction model is designed to learn the unknown parameter β from observed data, and it can use any up-to-date regression model. For this study, we use a linear regression (LR) model and a support vector regression (SVR) model. We fit the predictive model to the observed dataset of company value and effect variables. When an

additional company's effect set is given, the model can predict the company value. Additionally, we can predict future valuations of a list of companies and understand the future trends of the industry.

6 Experimental results

This section presents our evaluation of the prediction results. The company valuations we try to learn and predict were obtained from the Fortune 500. The Fortune 500 list, published by Fortune magazine,[10] ranks top (gross revenue) American public corporations during 1955 to the current year. Therefore, we use longitudinal network effects generated during 1981–2009 to learn and predict the company's value in terms of profit and revenue.

First, we learn and predict 20 Fortune companies' profits (and calculate their mean value). Then we calculate the mean profit with real profits (and also calculate the mean value). Second, we train the model for individual company, i.e., *IBM* and *Intel*, and predict their profits. Finally, we compare different feature sets and parameters.

For performance evaluation, we use the squared correlation coefficient (r^2) and the mean squared error (MSE) to quantify the correlation and error between the predicted valuations and the true valuations, respectively.

$$r^2 = \frac{\left(l\sum_{i=1}^{l} f(\mathbf{x})y_i - \sum_{i=1}^{l} f(\mathbf{x})\sum_{i=1}^{l} y_i \right)^2}{\left(l\sum_{i=1}^{l} f(\mathbf{x})^2 - \left(\sum_{i=1}^{l} f(\mathbf{x}) \right)^2 \right)\left(l\sum_{i=1}^{l} y_i^2 - \left(\sum_{i=1}^{l} y_i \right)^2 \right)}$$

$$(7)$$

$$\text{MSE} = \frac{1}{l}\sum_{i=1}^{l}(f(\mathbf{x}) - y_i)^2 \qquad (8)$$

Therein, l indicates the observed data, and y_i, f_i respectively denote the real data valuations and predicted valuations.

6.1 Company value prediction for Fortune companies

We select 20 large Fortune companies from different industries: *IBM, Intel Corp., Microsoft Corp., General Motors Co., Hewlett-Packard Co., Honda Motor Co., Ltd., Nissan Motor Co., Ltd., AT&T Communications Inc., Wal-Mart Stores Inc., Yahoo Inc., Nike Inc., Dell Inc., Starbucks Corp., JPMorgan Chase and Co., PepsiCo, Inc., Cisco Systems, Inc., FedEx Corp., The Gap Inc., American Electric Power Inc.,* and *Sun Microsystems Inc.* Because these companies continually appeared on the Fortune list

over several years, we have continuous records of both their company valuations and networks. We learn the profit model from each 5 years' networks, and we predict the next year's profits; then we compare with the real value of the profit earned in that year. Figure 4 displays a plot of the mean value of the predicted profits of these 20 companies, as learned from each of the prior 5 year periods, and the mean value of the real profit earned for that year. We use SVR (with an RBF kernel) model to learn parameters and to make predictions. We use the r^2 and MSE to quantify the correlation and error between the predicted valuations and true valuations, respectively. It is apparent that the output of the predicted valuations can capture the profit trend of these companies over the years, where $r^2 = 0.440$ and $\text{MSE} = 0.437$. Only in 1995 was the prediction much lower than the actual value, perhaps because these companies created profits but the intercompany relations still suffered an impact from the previous years' networks.

We also predict profits for two individual companies, *IBM* and *Intel*, based on their embeddedness in longitudinal networks. We learn from the prior 10 years' networks and predict the subsequent year's profit. Then we compare it with the real value of the profit. It is apparent that our prediction results can capture the trends of company profits moving with $r^2 = 0.888$ and $MSE = 0.108$ for *IBM*, and $r^2 = 0.243$ and $MSE = 0.251$ for *Intel* (Fig. 5).

6.2 Effectiveness of network features and parameters

To evaluate the effectiveness of network features, we use different feature sets for predicting 20 companies' mean profits during several years, and take the average over years to compare the prediction performance by each feature set. Notations "s", "t", "p", and "d" respectively indicate that results are obtained only using current network structural
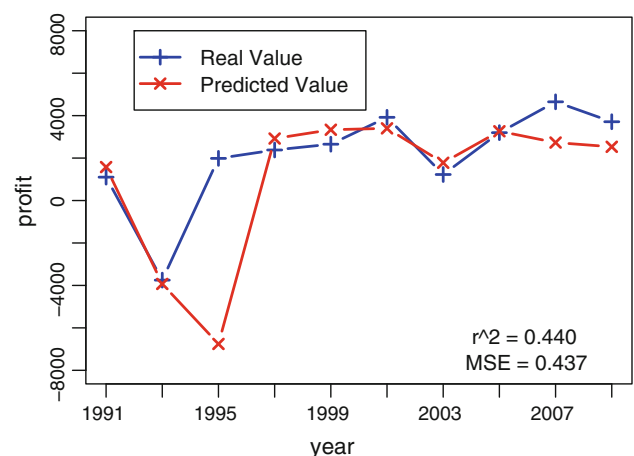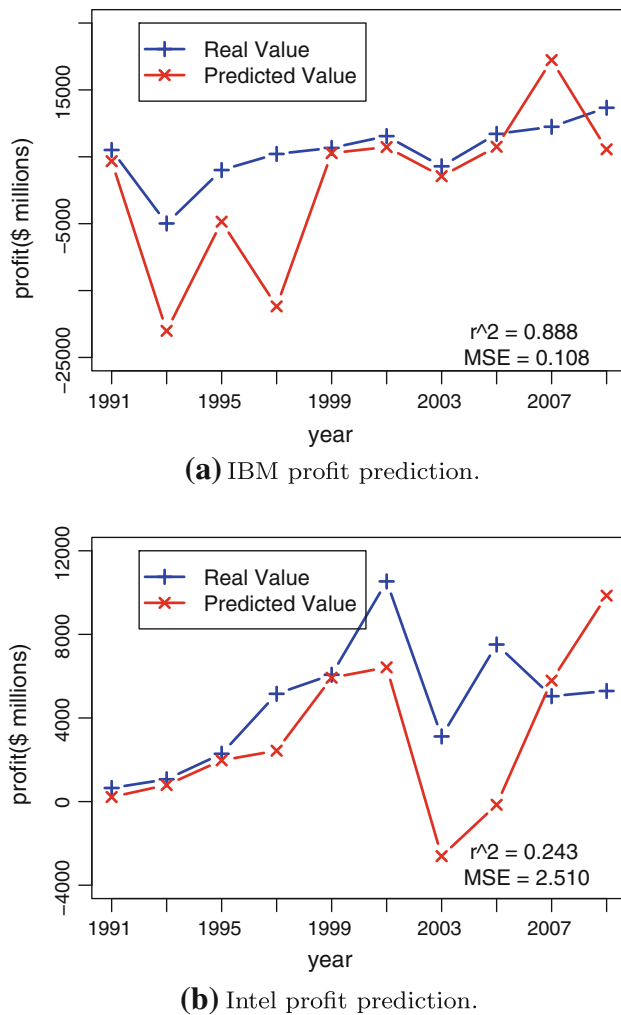

Fig. 4 Prediction of the mean profits of 20 Fortune companies

(a) IBM profit prediction.

(b) Intel profit prediction.

Fig. 5 Profit prediction for *IBM* and *Intel*

features (i.e. $\mathbf{F}_x^t$), historical network structural features (i.e. $\mathbf{F}_x^H$), delta-changes in the network features (i.e. $\Delta\mathbf{F}_x^H$), and financial features only ($\mathbf{P}_x^H$). We also combine these features for prediction: "sp" implies combining current network features with financial features, and "stdp" signifies a combination of all features, both network and financial features. We use the SVR (with the RBF kernel) model to learn parameters from the prior 5 years' networks and predict the subsequent year's profit. As the results presented in Fig. 6 reveal, for profit prediction, using the feature set "p" i.e., financial profile features only (e.g., prior year's profit, revenue.) has better performance ($r^2 = 0.383$, MSE $= 0.287$) than that realized using "s", "t", and "d" features only. However, by combining structural and temporal features with financial features as in "sp", "tp", "dp", and "stdp", the prediction results will improve. Particularly, the prediction performance realized using "stdp" ($r^2 = 0.512$, MSE $= 0.363$), i.e., joint network and financial features, will outperform that realized

using only the network features and only the financial features by 150% and 34%, respectively.[11]
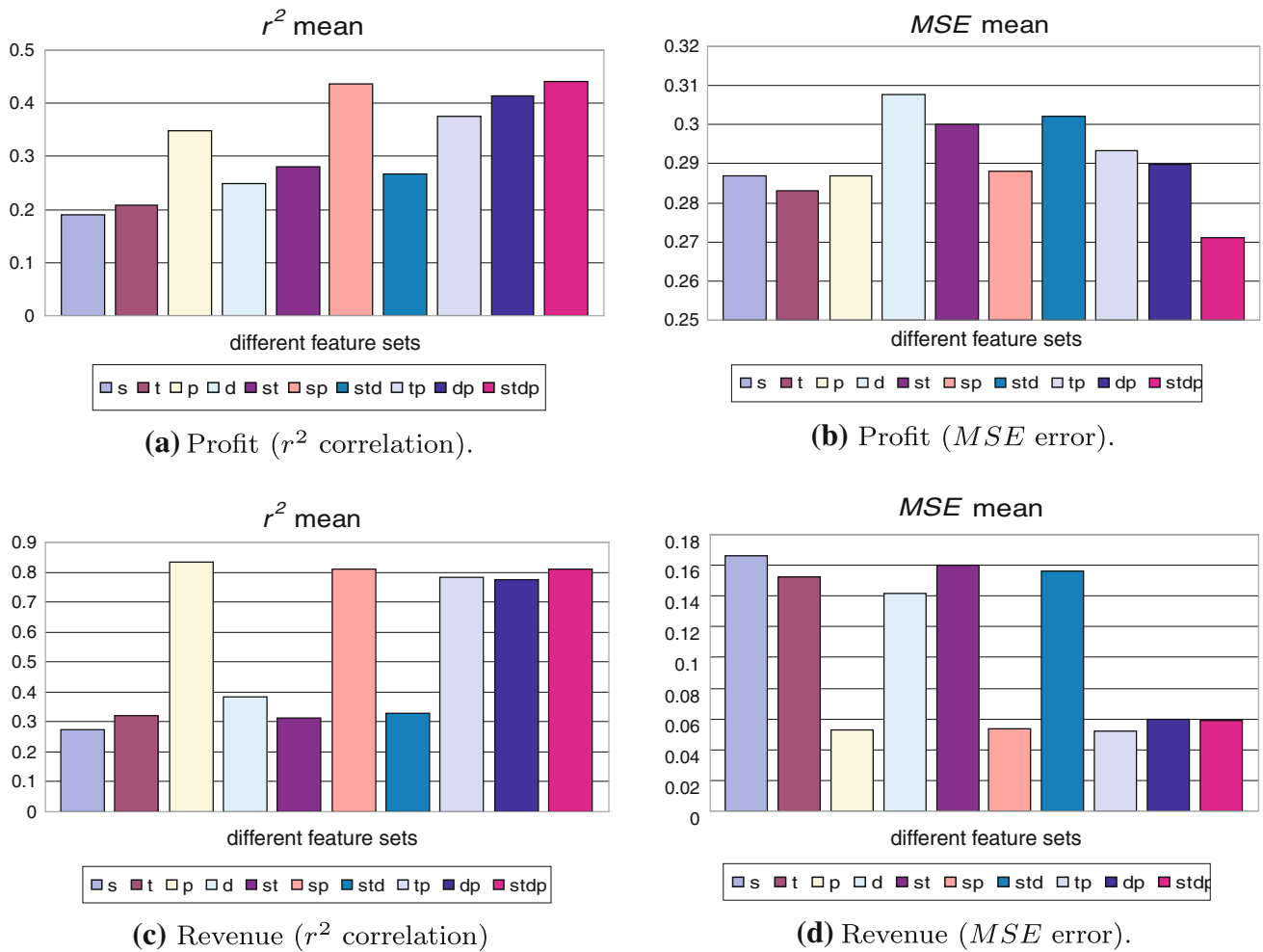
However, from the results of revenue prediction, we found network features do not seem to contribute to prediction. When we use a different prediction model, i.e., a linear regression model and obtained similar results. We can say that (longitudinal) network features contribute to predicting companies' profit, but not to revenue prediction. It is interesting to understand what types of company valuations or performance are sensitive to network features and receive impact from the network embeddedness of companies.

To tune the best parameters of historical *window size* and *delta size*, we compare their valuations that existed 1 and 3 years prior. Figure 7 shows the prediction performance of 20 companies' mean profits, the prediction being made using different window and delta sizes. We take the average of r^2 of different years r^2. Results showed that, of both the *window* and *delta* sizes, one is sufficient for profit prediction, which implies that using last year's network effects and the delta-change in them from the prior year to the present year provides better results than using the networks existing 3 years prior. Networks therefore apparently show a 1-year lagged impact on changes in a company's value.

7 Discussion and conclusions

Many different definitions of the company value (or performance) exist, such as financial performance, market performance, and employee satisfaction and responsibility (Xiao et al. 2009). Our experimentally obtained results show that longitudinal network features contribute to companies' profit prediction, but do not contribute to revenue prediction. It is interesting to note what types of company valuations or performance are sensitive to intercompany relations and which receive impact from network embeddedness of companies. In addition, company value is affected by influences both inside of the network (through ties) and those outside of the network (e.g., customer behavior, political decision, economic crisis, etc.). In the investigation described in this paper, we are interested in the impact from inside of the network, and we find valuable

---

[11] As comparison approaches, predictions using financial features alone can be regarded as a financial approach, using network features alone can be regarded as a network approach (but the network features are proposed/generated by ourselves), and using both financial and network features can be regarded as the proposed approach. Results show that a financial approach provides better performance than network approaches. However, by combining those financial and network features, the proposed approach outperforms both financial and network approaches.

(a) Profit (r^2 correlation).

(b) Profit (MSE error).

(c) Revenue (r^2 correlation)

(d) Revenue (MSE error).

Fig. 6 Mean profit and revenue prediction for 20 Fortune companies using different feature sets. *s* structural features, *t* temporal features, *d* delta-change in temporal features, *p* financial profiles

(a) *window size.*
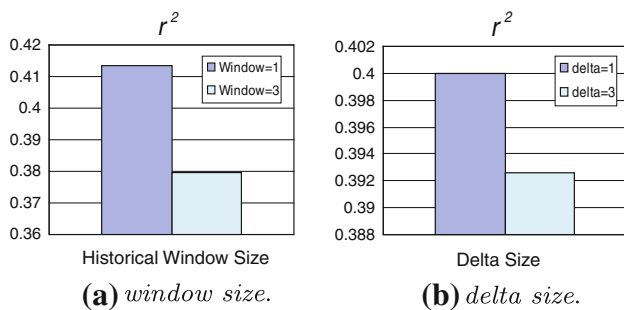
(b) *delta size.*

Fig. 7 *Window size* and *delta size* for profit prediction

structural embeddeness of companies that are related with company valuations. For making predictions, we are interested in finding valuable structural embeddeness of companies that are related with company valuations. Our intercompany networks are extracted based on a statistical count obtained from news articles about companies over a period. Therefore, they are suitable for predicting a

company' long-term value change. Prediction of short-term changes, might necessitate the use of another algorithm, which we intend to develop in future studies. However, our algorithm, which generates longitudinal network features and our prediction model are applicable for both long-term and short-term predictions of company value.

In this study, we explored a new analytical paradigm of using social networks of companies to predict a company's value. We developed an algorithm and system for inferring longitudinal intercompany networks from public news. We described impact relations among companies and developed an extraction method based on document-level and sentence-level co-occurrence and importance. Using the system, we can elucidate the evolution of company networks over the years with different structural characteristics. After constructing valued directed longitudinal company networks over several years, we defined and extracted network effects for each target company from the networks. We investigated network characteristics from

local and global relations, and combined historical structural effects as well as delta changes in structures to generate network effects. We applied an SVM regression model to learn and predict company valuation in terms such as profits and revenues.

Our prediction model can capture the trends of changes in the valuations of a group of companies or those in an individual company over several years. Results show that generic relational networks are useful for predicting company value, particularly company profit. Profit prediction based on joint networks and financial analysis outperforms predictive methods based solely on network effects by 150% and those based solely on financial effects by 34%. By tuning the *window size* and the *delta size* of longitudinal network effects, we found that the last year's network embeddedness is good for predicting this year's company value. Networks therefore apparently show a 1-year lagged impact on changes in a company's value. In this study, our networks are found to be suitable for predicting long-term changes in company value. Future studies will specifically examine detection of short-term changes in the company value based on intercompany networks and specifically examine real-time network-effect extraction and processing infrastructure.

References

Backstrom L, Huttenlocher D, Lan X, Kleinberg J (2006) Group formation in large social networks: Membership, growth, and evolution. In: The twelfth ACM SIGKDD international conference on knowledge discovery and data mining (SIGKDD'06)

Bao S, Li R, Yu Y, Cao Y (2008) Competitor mining with the web. IEEE Trans Knowl Data Eng 20(10):1297–1310

Ben-Zvi T (2009) Network structure and centrality: a simulation experiment. In: 8th IEEE/ACIS international conference on computer and information science (ICIS 2009)

Bengtsson M, Kock S (1999) Cooperation and competition in relationships between competitors in business networks. J Bus Ind Mark 14:178–194

Bernstein A, Clearwater S, Hill S, Perlich C, Provost F (2002) Discovering knowledge from relational dataextracted from business news. In: SIGKDD-2002 workshop on multi-relational data mining

Blum A, Langley P (1997) Selection of relevant features and examples in machine learning. Artif Intell 97(1–2):245–271

Cao Z, Liu T (2007) Learning to rank: from pairwise approach to listwise approach. In: Proceedings of the 24th international conference on machine learning, pp 129–136

Doreian P, Stokman F (1997) Evolution of social networks. Gordon and Breach, Amsterdam

Feld SL (1997) Structural embeddedness and stability of interpersonalrelations. Social Netw 19:91–95

Geng X, Liu TY, Qin T, Li H (2007) Feature selection for ranking. In: SIGIR'07, pp 407–414

Hu C, Xu L, Shen G, Fukushima T (2009) Temporal company relation mining from the web. In: APWeb/WAIM'09, pp 392–403

Jin Y, Matsuo Y, Ishizuka M (2007) Extracting interbusiness relationship from world wide web. J Jpn Soc Artif Intell 22(1): 48–57

Karamon J, Matsuo Y, Ishizuka M (2008) Generating useful network-based features for analyzing social networks. In: Twenty-third conference on artificial intelligence (AAAI-08), pp 1162–1168

Katz L, Proctor CH (1959) The concept of conguration of interpersonal relations in a group as a time-dependent stochastic process. Psychometrika 24(4):317–327

Kautz H, Selman B, Shah M (1997) Referral web: combining social networks and collaborative filtering. Commun ACM 40(3): 63–65

Kleinberg JM (1999) Hubs, authorities, and communities. J ACM 46(5):604–632

Leenders R (1995a) Models for network dynamics: a markovian framework. J Math Sociol 20:1–21

Liben-Nowell D, Kleinberg J (2007) The link-prediction problem for social networks. J Am Soc Inf Sci Technol 58(7):1019–1031

Matsuo Y, Mori J, Hamasaki M, Ishida K, Nishimura T, Takeda H, Hasida K, Ishizuka M (2006) POLYPHONET: an advanced social network extraction system. In: Proceedings of WWW2006

McCulloh IA, Carley KM (2009) Longitudinal dynamic network analysis. Casos technical report. Center for theComputational Analysis of Social and Organizational Systems

Mika P (2005) Flink: semantic web technology for the extraction and analysis of social networks. J Web Semant 3(2):211–223

Page L, Brin S, Motwani R, Winograd T (1998) ThePageRank citation ranking: bringing order to the web. Technical report, Stanford University

Qin T, Liu T, Zhang X, Wang D, Xiong W, Li H (2008) Learning to rank relational objects and its application to web search. In: Proceedings of WWW2008

Snijders TA (1997) Models for longitudinal network data. In: Carrington PJ, Scott J, Wasserman S (eds) Models and methods in social network analysis.Cambridge University Press, New York, pp 148–161

Tetlock PC, Saar-Tsechansky M, Macskassy S (2008) More than words: quantifying language to measure firms' fundamentals. J Fin 63(3):1437–1467

Uzzi B (1997) Social structure and competition in interfirm networks: the paradox of embeddedness. Admin Sci Quart 42:35–67

Wang J.L, Chan S.H (2007) Stock market trading rule discovery using pattern recognition and technical analysis. Expert Syst Appl 33: 304–315

Wasserman S, Faust K (1994) Social network analysis. Methods and applications. Cambridge University Press, Cambridge

Xiao L, Dasgupta S (2009) The effects of dynamic it capability and organizational culture on firm performance: an empirical study. In: 8th IEEE/ACIS international conferenceon computer and information science (ICIS'2009)

Yang H, Chan L, King I (2002) Support vector machine regression for volatile stock market prediction. In: Proceedings of the third international conference on intelligent data engineering and automated learning (IDEAL'02). Springer, London, pp 391–396

Zhao Z, Liu H (2007) Searching for interacting features. In: Proceedings of IJCAI2007