

Ranking Companies Based on Multiple Social Networks Mined from the Web

Yingzi Jin, Yutaka Matsuo and Mitsuru Ishizuka
The University of Tokyo, Japan

1. Introduction

Many rankings existing for popularity, recommendation, evaluation, election, etc. can be found in the real world as well as on the Web. Many efforts are undertaken by people and companies to improve their popularity, growth, and power, the outcomes of which are all expressed as rankings (designated as target rankings). Are these rankings merely the results of its elements' own attributes? In the theory of social network analysis (SNA), the performance and power (i.e. ranking) of actors are usually interpreted as relations and the relational structures they embedded. For example, if we seek to rank companies by market value, we can extract the social network of the company from the Web and discern, and then subsequently learn, a ranking model based on the social network. Consequently, we can predict the ranking of a new company by mining its relations to other companies. We can learn from existing rankings to expect other target rankings. We can learn from existing rankings to expect other rankings. Furthermore, we can understand the kinds of relations which are important for the target rankings; we can determine the type of structural extension of companies that can improve the target rankings.

This study specifically examines the application of a social network that provides an example of advanced utilization of social networks mined from the Web. We present ranking learning approaches using a social network that is mined from the Web. The proposed model combines social network mining and ranking learning, which further uses multiple relations on the Web to explain arbitrary rankings in the real world. Experimental results for learning to rank companies based on multiple social networks mined from the Web confirm the effectiveness of our models for explaining target rankings as well as real world phenomena using multiple social networks. Several findings including social networks vary according to different relational indices or types even though they contain the same list of entities. Relations and networks of different types differently impact on target of ranking. Multiple networks have more information than single networks for explaining target ranking. Well-chosen attribute-based features have good performance for explaining the target ranking. However, by combining proposed network-based features, the prediction results are further improved. This study specifically examines the application of a social network that provides an example of advanced utilization of social networks mined from the Web.

The following section presents a description of an overview of the ranking learning model. Section 3 briefly introduces our previous work for extracting social networks from the Web. Section 4 describes proposed ranking learning models based on extracted social networks. Section 5 describes the experimental settings and results. Section 6 presents some related works before the paper concludes.

2. System Overview

Our study explores the integration of mining relations (and structures) among entities and the learning ranking of entities. For that reason, we first extract relations and then determine a model based on those relations. Our reasoning is that important relations can be recognized only when we define some tasks. These tasks include ranking or scores for entities, i.e., *target ranking* such as ranking of companies for job-seekers, CD sales, popular blogs, and sales of products. In short, our approach consists of two steps:

Step 1: Constructing Social Networks Given a list of entities with a target ranking, we extract a set of social networks among these entities from the Web.

Step 2: Ranking learning Learn a ranking model based on the relations and structural features generated from the networks.

Once we obtain a ranking model, we use it for prediction for unknown entities. Additionally, we can obtain the weights for each relation type as well as relation structure, which can be considered as important for target rankings. The social network can be visualized by specifically examining its relations if the important relations are identified. Alternatively, social network analysis can be executed based on the relations.

3. Constructing Social Networks from the Web

In this step our task is, given a list of entities (i.e., companies) $V=\{v_1, \dots, v_n\}$, we construct a set of social networks $G_i(V, E_i)$, $i \in \{1, \dots, m\}$, where m signifies the number of relations, and $E_i=\{e_i(v_x, v_y) \mid v_x \in V, v_y \in V, v_x \neq v_y\}$ denotes a set of edges with respect to the i -th relation, where $e_i(v_x, v_y)$ is equal to 1 if companies v_x and v_y have relation i , and 0 otherwise.

A social network is obtainable through various approaches; one is to use Semantic Web data. With developments in the Semantic Web, the Web includes growth of machine-readable descriptions of people: FOAF documents. The FOAF provides an RDF/XML vocabulary to describe personal information, including name, mailbox, homepage URI, interest, friends, and so on. Using FOAF documents, we can construct social networks among people. Given a list of persons V , we first use *foaf:Person* to mapping each name with FOAF instances, then connect persons with several meaning of relational properties such as *foaf:knows*, *foaf:interest*, *foaf:location*, *foaf:publications*, and *foaf:currentProject* properties. Consequently, we can construct social networks G_i of different kinds. When a person is described in more than one FOAF document, we must fuse information from multiple sources using identical properties such as *foaf:mbox*, *foaf:homepage* and *foaf:Weblog* and generate aggregated information about the person (Finin et al., 2005). Furthermore, by combining FOAF documents to DBLP data, we can construct more kinds of social networks such as *authorship* network, *citation* network (Aleman-Meza et al., 2006; Zhou et al., 2008).

Another is to extract social networks using Web mining. Several studies have particularly addressed the use of search engines as well as text mining for social network extraction. Through this study, we detail the co-occurrence approach and relation-identification approach used by Matsuo et al. (Matsuo et al., 2006) and Jin et al. (Jin et al., 2008), respectively, as a basis of our study. We are interested only in undirected networks.

3.1 Co-occurrence-based approach

The social network of the first kind is extracted using a co-occurrence-based approach. This approach was used originally by Kautz et al. (Kautz et al., 1997), and was recently applied and modelled by Mika (Mika, 2005) and Matsuo et al. (Matsuo et al., 2006) to extract researcher networks automatically from the Web. The fundamental idea underlying the co-occurrence approach is that *the strength of a relation between two entities can be estimated by co-occurrence of their names on the Web*. The strength of relevance of two persons, x and y , is estimated by putting a query x AND y to a search engine: If x and y share a strong relation, we can usually find various evidence on the Web such as links found on home pages, lists of co-authors of technical papers, organization charts, and so on. An edge will be invented when the relation strength by the co-occurrence measure is higher than a predefined threshold. Subsequently, we use the Overlap coefficient $n_{x \cap y} / \min(n_x, n_y)$ (used by (Matsuo et al., 2006)) as well as the Matching coefficient as relational indices and thereby construct co-occurrence-based networks of two kinds: an *overlap network* ($G_{overlap}$) and a *cooc network* (G_{cooc}). Many advanced algorithms are described in (Matsuo et al., 2006).

3.2 Relation-identification approach

We proposed the *relation-identification* approach to extract target relational social networks in (Jin et al. 2008). This approach emphasizes real-world relations such as a mutual stock holding relation, capital combination, trade relation, personal relation (i.e., mutual dispatch of officials), rivalry, and a competitive relation. These relations are published in news articles or by news releases that might be obtained easily from the Web.

Given a list of companies and target relations as input, the method extracts a social network of entities. To collect target relational information from the tops of Web pages, it makes elaborate queries to emphasize a specific relation, and applies text processing to those pages to form an inference of whether or not the relation actually exists.

First, queries are produced by adding *relation keywords* (such as "*alliance AND corporate*") to each pair of companies. Relation keywords are in advance for each target relation by measuring the Jaccard relevance from given seed words. Then, to extract target relations from Web documents, a simple pattern-based heuristic is useful: First pick all sentences that include the two company names (x and y), and assign each sentence the sum of relation keyword scores in the sentence. The score of companies x and y is the maximum of the sentence scores. An edge is invented between the two companies if that score is greater than a certain threshold. Subsequently, we extract two kinds of relational networks: a *business-alliance network* ($G_{business}$) and a *capital-alliance network* ($G_{capital}$).

Extracted networks for 312 companies related to the electrical products industry from Japan are portrayed in Fig. 1. It is apparent that the social networks vary with different relational indices or types even though they contain the same list of entities.

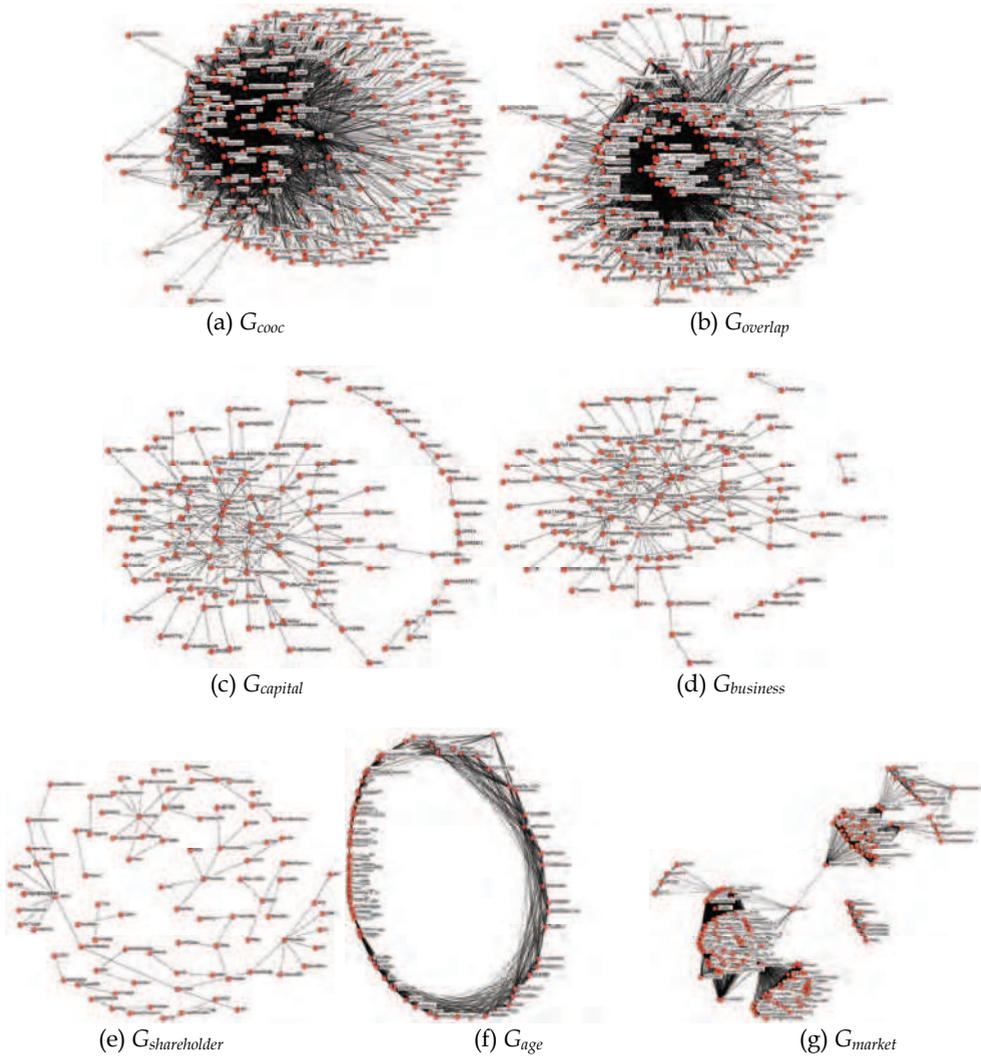


Fig. 1. Social networks for companies in electrical industrial with different relational indices or types.

4. Ranking Learning Model

For the list of nodes $V = \{v_1, \dots, v_n\}$, given a set of networks $G_i(V, E_i)$, $i \in \{1, \dots, m\}$ (constructed by section 3) with a target ranking \mathbf{r}^* ($\in R^t$) (where $t \leq n$, and r_k^* denotes k -th element of the vector \mathbf{r}^* and means the target ranking score of entity v_k), the goal is to learn a ranking model based on these networks.

First, as a baseline approach, we follow the intuitive idea of simply using approach from SNAs (i.e. centrality) to learn ranking. As the second approach, multiple relations are

combined into one to consider a combination model for ranking. Finally, to learn ranking, we propose a more useful algorithm that generates various network features for individuals from social networks.

4.1 Baseline Model

Based on the intuitive approach, we first overview commonly used indices in social network analysis and complex network studies. Given a set of social networks, we rank entities on these networks using different network centrality indices. We designate these rankings as *network rankings* because they are calculated directly from relational networks. We use \mathbf{r}_i ($\in R^n$) to denote network ranking that is directly attributable to the i -th relational network G_i . Our task is to find a ranking model based on network rankings that maximally explain the target ranking.

4.1.1 Choosing the most predictive type of relation

To address the question of what kind of relation is most important for companies, we intuitively compare rankings caused by relations of various types. Although simple, it can be considered as an implicit step of social network analysis given a set of relational networks. We merely choose the type of relation that maximally explains the given ranking. We rank each type of relational network; then we compare the *network ranking* with the *target ranking*. Intuitively, if the correlation to the network ranking \mathbf{r}_i is high, then the relation \hat{i} represents the important influences among entities for the given target ranking. Therefore, this model is designed to find an optimal relation \hat{i} from a set of relations:

$$\hat{i} = \arg \max_{i \in \{1, \dots, m\}} \text{Cor}(\mathbf{r}_i, \mathbf{r}^*) \quad (1)$$

We define a ranking function $h(G)$ that returns a vector of network ranking ($\in R^n$) for given network $G(V, E)$. Therefore, the i -th network ranking \mathbf{r}_i is obtained from $h(G_i)$. Here are the other questions for what kind of ranking indices are most appropriate to explain the target ranking. In the next section, we treat several *centrality* measures from SNAs as our different network ranking function $h(G)$.

4.1.2 Choosing the most predictive type of centrality indices

Different meanings of prominence and importance can be generated from a network, such as “having a powerful position”, and having “more opportunities” and “fewer constraints”. Several *centrality* measures are useful to rank network entities with these different meanings: degree centrality, betweenness centrality, and closeness centrality and other centralities. Below, we introduce these different meanings of centrality.

- *Degree centrality* is an assessment of the number of relations that any given actor is engaged in. Actors with more ties to other actors might be in advantaged positions, which can be defined as

$$C_d(v_i) = \frac{d(v_i)}{(n-1)} \quad (2)$$

Therein, $d(v_i)$ is the degree of node v_i , and n is the number of nodes.

- *Betweenness centrality* measures an actor as central if it lies between other actors on their geodesics. More actors depend on one actor v_l to make connections with other actors (geodesics passing through).

$$C_b(v_l) = \frac{\sum_{(v_p, v_q) \in (V \times V), v_p \in V, v_q \in V} g_{v_p, v_q}(v_l) / g_{v_p, v_q}}{(n-1)(n-2)} \quad (3)$$

where g_{v_p, v_q} is the number of shortest geodesic paths from node v_p to v_q , and $g_{v_p, v_q}(v_l)$ is the number of shortest paths from v_p to v_q that pass through node v_l

- *Closeness centrality* is a sophisticated measure that is defined as the mean shortest path between an actor i and all other actors that are reachable from that actor. Closeness can be regarded as a measure of how long it will take information to spread from a given actor v_l to other reachable actors in the network.

$$C_c(v_l) = \frac{\sum_{v_p \in V, v_p \neq v_l} g_G(v_l, v_p)}{(n-1)} \quad (4)$$

In that equation, $g_G(v_l, v_p)$ is the shortest geodesic paths from v_l to reachable node v_p .

These measures characterize some aspects of the local (i.e. degree) or global (i.e., closeness, betweenness) network structure, as indicated by a given actor's embeddedness in the network (Wasserman & Faust, 1994). Intuitively, given a target ranking, the most predictive type of centrality measure is finding optimal centrality measure h_j for target ranking \mathbf{r}^* from a set of ranking functions.

$$\hat{j} = \arg \max_{h_j \in \{h_1, \dots, h_s\}} \mathbf{Cor}(\mathbf{r}_j, \mathbf{r}^*) \quad (5)$$

For different relational networks, the network ranking from i -th network with j -th ranking can be presented as $\mathbf{r}_{i,j} (\in R^n)$, which is obtainable from $h_j (G_i)$, where $h_j \in \{h_1, \dots, h_s\}$, $i \in \{1, \dots, m\}$. Therefore, the first method can be extended simply to find a pair of optimal parameters $\langle \hat{i}, \hat{j} \rangle$ (i.e., i -th network by j -th ranking indices) that maximizes the coefficient between network rankings with a target ranking.

$$\langle \hat{i}, \hat{j} \rangle = \arg \max_{i \in \{1, \dots, m\}, h_j \in \{h_1, \dots, h_s\}} \mathbf{Cor}(\mathbf{r}_{i,j}, \mathbf{r}^*) \quad (6)$$

4.2 Network Combination Model

Many centrality approaches related to ranking network entities specifically examine graphs with a single link type. However, multiple social networks exist in the real world, each representing a particular relation type, and each of which might be integrated to play a distinct role in a particular task. We combine several extracted multiple social networks into one network and designate such a social network as a *combined-relational network* (denoted as $G_c(V, E_c)$). Our target is using combined-relational network, which is integrated with multiple networks extracted from the Web, to learn and predict the ranking. The important

questions that must be resolved here is *how to combine relations to describe the given ranking best*.

For $G_c (V, E_c)$, the set of edges is $E_c = \{e_c (v_x, v_y) \mid v_x \in V, v_y \in V, v_x \neq v_y\}$. Using a linear combination, each edge $e_c (v_x, v_y)$ can be generated from $\sum_{i \in \{1, \dots, m\}} w_i e_i (v_x, v_y)$, where w_i is i -th element of \mathbf{w} (i.e., $\mathbf{w} = [w_1, \dots, w_m]^T$). Therefore, the purpose is to learn optimal combination weights $\hat{\mathbf{w}}$ to combine relations as well as optimal ranking method h_j on G_c :

$$\langle \hat{\mathbf{w}}, \hat{j} \rangle = \arg \max_{\mathbf{w}, h_j \in \{h_1, \dots, h_m\}} \text{Cor}(\mathbf{r}_{c,j}, \mathbf{r}^*) \quad (7)$$

Cai et al. (D.Cai et al. 2005) regard a similar idea with this approach: They attempt to identify the best combination of relations (i.e., relations as features) which makes the relation between the intra-community examples as tight as possible. Simultaneously, the relation between the inter-community examples is as loose as possible when a user provides multiple community examples (e.g. two groups of researchers). However, our purpose is learning a ranking model (e.g. ranking of companies) based on social networks, which has a different optimization task. Moreover, we propose innovative features for entities based on combination or integration of structural importance generated from social networks.

In this study, we simply use Boolean type ($w_i \in \{1, 0\}$) to combine relations. Using relations of m types to combine a network, we can create $2^m - 1$ types of combination-relational networks (in which at least one type of relation exists in the G_c). We obtain network rankings in these combined networks to learn and predict the target rankings. Future work on how to choose parameter values will be helpful to practitioners.

4.3 Network-based Feature Integration Model

The most advanced method in our research is to integrate multiple indices that are obtained from multiple social networks. A feature by itself (e.g. a centrality value) may have little correlation with the target ranking, but when it is combined with some other features, they may be strongly correlated with the target rankings (Zhao & Liu, 2007). The idea in this model is the integration of all network features for individuals from networks as a context of the actors to learn the target ranking. Those features are expected to be useful to interpret a given target ranking accurately.

We integrate multiple indices from social networks, thereby combining several perspectives of importance for individuals from different relational structures. Simply, we can integrate various centrality values (described in the Baseline model) for each actor, thereby combining different meanings of importance to learn the ranking model. Furthermore, we can generate more relational and structural features from a network for each, such as how many nodes are reachable, how many connections one's friends have, and the connection status in one's friends. We might understand some about the behaviour and power about the individual as well as we predict their ranking if we could know the structural position of individuals. Herein, we designate these features generated from relations and networks as *network-based features*. The interesting question is *how to generate network-based features from networks for each, and how to integrate these features to learn and predict rankings*. Below we will describe the approach of generating and integrating network-based features.

4.1.1 Generating Network-based Features for nodes

For each x , we first define node sets with relations that might effect x . Then we apply some operators to the set of nodes to produce a list of values. Subsequently, the values are aggregated into a single feature value. Therefore, we can generate several structural features for each node. For example, when calculating the closeness centrality (i.e., average distance from node x to all others) of node x , we discern its value fundamentally in three steps: we first select reachable nodes from x ; secondly, we calculate the distance between node x and each node; finally, we take the average of these distances. Additionally, we can discern the value of the closeness centrality of node x . For that reason, we can construct indices used in SNAs through these steps. Below, we explain each step in detail.

- Step 1: *Defining a node set*

First, we define a node set. Most straightforwardly, we can choose the nodes that are adjacent to node x . The nodes are those of distance one from node x . The nodes with distances of two, three, and so on are definable as well. We define a set of nodes $C_x^{(k)}$ as a set of nodes within distance k from x . For example, we can denote the node set adjacent to node x as C_x^1 . In addition, we use $C_y^{(k)}$ to express a set of nodes within distance k from y (where $y \neq x$).

- Step 2: *Operation on a Node Set*

Given a node set, we can conduct several calculations for the node set. Below, we define operators with respect to two nodes; then we expand it to a node set with an arbitrary number of nodes. The simple operation for two nodes is to check whether the two nodes are adjacent or not. We denote these operators as $s^1(x,y)$, which returns 1 if nodes x and y are mutually connected, and 0 otherwise. We also define operator $t(x,y) = \operatorname{argmin}_k \{s^{(k)}(x,y) = 1\}$ to measure the geodesic distance between the two nodes on the graph. These two operations are applied to each pair of nodes in N if given a set of more than two nodes (denoted as N). This calculation can be defined as follows.

$$\text{Operator} \circ N = \{\text{Operator}(x, y) \mid x \in N, y \in N, x \neq y\} \quad (8)$$

For example, if we are given a node set $\{n_1, n_2, n_3\}$, we can calculate $s^{(1)}(\{n_1, n_2\})$, $s^{(1)}(\{n_1, n_3\})$, and $s^{(1)}(\{n_2, n_3\})$ and return a list of three values, e.g., (1, 0, 1). We denote this operation as $s^{(1)} \circ N$.

In addition, to s and t operations, we define two other operations. One operation is to measure the distance from node x to each node, denoted as t_x . Instead of measuring the distance between two nodes, $t_x \circ N$ measures the distance of each node in N from node x . Another operation is to check the shortest path between two nodes. Operator $u_x(y,z)$ returns 1 if the shortest path between y and z includes node x . Consequently, $u_x \circ N$ returns a set of values for each pair of $y \in N$ and $z \in N$. The other is to calculate the structural equivalence between node x and y . This is denoted as $e_x(y)$.

- Step 3: *Aggregation of Values*

Once we obtain a list of values, several standard operations can be added to the list. Given a list of values, we can take the summation (*Sum*), average (*Avg*), maximum (*Max*), and minimum (*Min*). For example, if we apply *Sum* aggregation to a value list (1,0,1), we obtain a value of 2. We can write the aggregation as e.g., $\text{Sum} \circ s^{(1)} \circ$

N. Although other operations can be performed, such as taking the variance or taking the mean, we limit the operations to the four described above. The value obtained here results in the network-based feature for node x . Additionally, we can take the difference or the ratio of two obtained values. For example, if we obtain 2 by $Sum \circ s^{(1)} \circ C_x^{(1)}$ and 1 by $Sum \circ s^{(1)} \circ C_x^{(k)}$, the ratio is $2/1 = 2.0$.

We can thereby generate a feature by subsequently defining a nodeset, applying an operator, and aggregating the values. The number of possible combinations is enormous. Therefore, we apply some constraints on the combinations. First, when defining a nodeset, k is an arbitrary integer theoretically; however, we limit k to be 1 for a nodeset of neighbors, k to be 3 for a nodeset of reachable nodes simplicity. Operator $s^{(k)}$ is used only as $s^{(1)}$. We also limit taking the ratio only to those two values with neighbor nodeset $C_x^{(1)}$ and reachable nodeset $C_x^{(\infty)}$. The nodesets, operators, and aggregations are presented in Table 1. We have $2(\text{nodesets}) \times 5(\text{operators}) \times 4(\text{aggregations}) = 40$ combinations. There are ratios for $C_x^{(1)}$ to $C_x^{(k)}$ if we consider the ratio. In all, there are 4×5 more combinations: there are 60 in all. Each combination corresponds to a feature of node x . Some combinations produce the same value. One example is that $Sum \circ t_x \circ C_x^1$ is the same as $Sum \circ s \circ C_x^{(\infty)}$, representing the degree of node x .

Notation	Input	Output	Description
$C_x^{(1)}$	node x	a nodeset	adjacent nodes to x
$C_x^{(k)}$	node x	a nodeset	nodes within distance k from x
$s^{(1)}$	a nodeset	a list of values	1 if connected, 0 otherwise
t	a nodeset	values	distance between a pair of nodes
t_x	a nodeset	a list of values	distance between node x and other nodes
γ	a nodeset	values	number of links in each node
u_x	a nodeset	a list of values	1 if the shortest path includes node x , 0 otherwise
<i>Avg</i>	a list of values	a value	average of values
<i>Sum</i>	a list of values	a value	summation of values
<i>Min</i>	a list of values	a value	minimum of values
<i>Max</i>	a list of values	a value	maximum of values
<i>Ratio</i>	Two values	value	ratio of value on neighbor nodeset $C_x^{(1)}$ by reachable nodeset $C_x^{(\infty)}$

Table 1. Operator list

The resultant value sometimes corresponds to a well-known index, as we intended in the design of the operators. For example, the network density can be denoted as $Avg \circ s^{(1)} \circ N$. It represents the average of edge existence among all nodes; it therefore corresponds to the network density. These features represent some possible combinations. Some lesser-known features might actually be effective.

4.1.2 Network-based features with SNAs indices

It is readily apparent that centralities described in baseline approach are also a particular case of this model because our network-base feature include those centrality measures and other SNAs indices for each node. Below, we describe other examples that are used in the social network analysis literature.

- diameter of the network: $Min \circ t \circ N$
- characteristic path length: $Avg \circ t \circ N$
- degree centrality: $Sum \circ s_x^{(1)} \circ C_x^{(1)}$
- node clustering: $Avg \circ s^{(1)} \circ C_x^{(1)}$
- closeness centrality: $Avg \circ t_x \circ C_x^{(\infty)}$
- betweenness centrality: $Sum \circ u_x \circ C_x^{(\infty)}$,
- structural holes: $Avg \circ t \circ C_x^{(1)}$

When we set the element $Sum \circ s_x^{(1)} \circ N_x^{(1)}$ in a feature vector equal to 1, and all others to 0, we can elucidate the effect of degree centrality for predicting target ranking.

4.1.3 Network-based feature Integration

Next, generated network-based features to learn rankings are used for entities. The goal of learning is to integrate all features from networks into a single ranking of individuals. Combined, they are expected to be useful to interpret a given target ranking most accurately. After we generate various network-based features for individual nodes, we integrate them to learn ranking. This integration is accomplished through regression of features. We introduce an f -dimensional feature vector F , in which each element represents a network-based feature for each node. We identify the f -dimensional combination vector $\mathbf{u} = [u_1, \dots, u_f]^T$ to combine network-based features for each node. The inter-product $\mathbf{u}^T \mathbf{F}$ for each node produces n -dimensional ranking. For relational networks of m kinds, the feature vector can be expanded to $m \times 56$ -dimensions. In this case, the purpose is finding out whether optimal combination weight $\hat{\mathbf{u}}$ to $\mathbf{u}^T \mathbf{F}$ maximally explains the target ranking:

$$\hat{\mathbf{u}} = \arg \max_{\mathbf{u}} \text{Cor}(\hat{\mathbf{u}}^T \cdot \mathbf{A}, \mathbf{r}^*) \quad (9)$$

This model can be extended easily to add attributes (or profiles) of entities as features such as Sales, Assets, or number of employees of a company. We can use any technique, such as SVM, boosting and neural network, to implement the optimization problem. In this study, we consider using the Ranking SVM technique. Ranking SVM utilizes instance pairs and

their preference labels in training. The optimization formulation of Ranking SVM is the following:

$$\min \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i,j,q} \zeta_{i,j,q} \quad (10)$$

$$s.t. \forall (d_i, d_j) \in r_q^* : \mathbf{w} \phi(q, d_i) \geq \mathbf{w} \phi(q, d_j) + 1 - \zeta_{i,j,q}$$

where \mathbf{w} is a weight vector that is adjusted by learning to minimize the upper bound $\sum \zeta_{i,j,q}$. In addition, C is a parameter that enables trading-off of the margin size against training error. The result is a ranking function that has few discordant pairs with respect to the observed of the target ranking. For multi-relational networks, we can generate features for each single-relational network. Subsequently, we can compare the performance among them to understand which relational network produces more reasonable features. Thereby, we can see which relation(s) is important for the target ranking. Generating Network-based Features for nodes

5. Experimental Results

5.1 Datasets

We extract social networks for companies from 312 electrical product-related industry companies that are listed on the Tokyo Stock Exchange. All financial information about these companies is published in Yahoo! Finance (<http://profile.yahoo.co.jp/industry/electrical/electrical1.html>). For these companies, we extract social networks of seven kinds (Fig. 1) from the Web using a search engine Yahoo! Search Boss (<http://developer.yahoo.com/search/boss/>) and information from Toyo Keizai Inc. (<http://www.toyokeizai.co.jp>) i.e. a Japanese book and magazine publisher: the cooc network (G_{cooc}) and overlap network ($G_{overlap}$) network are extracted using the co-occurrence-based approach described in Section 3.1; the business-alliance network ($G_{business}$) and capital-alliance network ($G_{capital}$) are extracted using the relation-identification approach described in Section 3.3; same-market network (G_{market}) includes links that connect two companies listed on the same stock market; shareholding network ($G_{shareholder}$) connects shareholding relations among companies; similar-age network (G_{age}) connects two companies if their average age is similar (age-gap is less than two years); Each extraction method and corresponding figure of networks is listed in Table 2.

G_i	Network name	Extraction Method	Fig.
G_{cooc}	cooc network	Section 3.1	Fig. 1(a)
$G_{overlap}$	overlap network	Section 3.1	Fig. 1(b)
$G_{business}$	business-alliance network	Section 3.2	Fig. 1(c)
$G_{capital}$	capital-alliance network}	Section 3.2	Fig. 1(d)
G_{market}	same-market network	connect companies listed on the same stock market	Fig. 1(e)

$G_{shareholder}$	shareholding network	connect shareholding relations	Fig. 1(f)
G_{age}	similar-age network	connect similar average-age companies	Fig. 1(g)

Table 2. Constructed networks of electrical industry companies.

For our experiments, we set the target ranking of the companies by market capitalization (designated as *Market-Cap*), ranking of average annual income (designated as *Avg-In*), and the ranking of excellent accounts (designated as *Excellent*). The target ranking of *Avg-In* is collected from quarterly corporate reports from Toyo Keizai Inc. *Market-Cap* represents the market's valuation of all the equity in a corporation. From Yahoo! Finance we can obtain all *Market-Cap* information for listed companies in Japan. The ranking of *Excellent* is published by Nihon Keizai Shimbun Inc. (<http://www.nikkei.co.jp/>) every year in March. They rank companies based on evaluating factors of flexibility & sociality, earning & growth ability, development & research, age of employees, etc. The top 300 excellent companies include 22 electrical industry companies used in our experiments. Table 3 shows the top 20 companies ranked by *Avg-In*, *Market-Cap*, and *Excellent* in the electrical industry.

In our experiments, we conducted three-fold cross-validation. In each trial, two folds of actors are used for training, and one fold for prediction. The results we report in this section are those averaged over three trials. We use Spearman's rank correlation coefficient (ρ) (Spearman, 1904) to measure the pairwise ranking correlation.

5.2 Ranking Results

First, we rank companies on different networks according to their network rankings. Table 4 and Table 5 show the top 20 companies ranked by degree centrality and betweenness centrality, respectively, on different types of networks in the electrical industry field. Results show that *Hitachi*, *NIEC*, and *Fujitsu* have good degree centrality in different networks. In addition, *Hitachi* has good betweenness centrality in the networks: we can implicitly understand that *Hitachi* has good network embeddedness in the electrical industry.

r^*	Avg-In	Market-Cap	Excellent
1:	Keyence	Canon	Canon
2:	Advantest	Sony	Fanuc Ltd.
3:	AXELL	Panasonic	TDK
4:	Lasertec	Toshiba	Omron
5:	Fanuc Ltd.	Hitachi	Kyocera
6:	TEL	Mitsubishi	Systemex
7:	Sony	Fanuc Ltd.	Ricoh
8:	Screen	Sharp	Toshiba
9:	Yokogawa	Kyocera	Ibiden
10:	Elpida	Fujitsu	Rohm
11:	Canon	Ricoh	Sharp
12:	Nihon Kohden	Murata	Sony
13:	Panasonic	Keyence	Eizo Nanao
14:	Megachips	Ibiden	Fujitsu
15:	Ricoh	TEL	Optex

16:	Nippon Signal	Nidec	Cosel
17:	Ulvac	Rohm	Daihen
18:	Hirose Elec.	Konica	SMK
19:	SK Elec.	Minolta	
20:	Panasonic Elec.	TDK	Yamatake
		NEC	Ulvac

Table 3. Top 25 companies ranked by target rankings i.e. Avg-In, Market-Cap, and Excellent in an electrical industry field.

Additionally, these results reflect that companies have different centrality rankings even if they are in the same type of relational network. For instance, *Phoenix Elec.* and *SanRex* have good degree rankings in G_{market} and G_{age} networks respectively, but do not have good betweenness rankings in those networks. We also use seven carefully chosen fundamental indices as attributes of companies for comparison of our proposed network indices: **Capital**, **Employee Number**, **Sales**, return on equity (ROE), return on assets (ROA), the price earnings ratio (PER), and the price to book value ratio (PBR). Each of them has been used traditionally for company valuation. Additionally, we use the number of hits of names (HitNum) on the Web as another attribute (i.e. popularity on the Web) of a company. Table 6 shows the top 20 companies ranked by each attribute in the electrical industry field.

I_i, Cd	I_{cooc}, Cd	$I_{overlap}, Cd$	$I_{business}, Cd$	$I_{capital}, Cd$	I_{market}, Cd	$I_{shareholder}, Cd$	I_{age}, Cd
1:	NIEC	Keyence	Hitachi	Hitachi	Phoenix Elec.	Hitachi	SanRex
2:	JEM	Shindengen	Fujitsu	Suzuki	NIEC	Fujitsu	ALOKA
3:	Toshiba	HDK	Suzuki	Fujitsu	Shibaura	Mitsubishi	Koito
4:	JAE	Casio	Panasonic	Toshiba	Hamamatsu	Panasonic	TOA
5:	Pioneer	JAE	NEC	Mitsubishi	Nihon Kohden	Toshiba	Hitachi Medical
6:	JDL	Murata	Mitsubishi	Panasonic	Kenwood	Panasonic Elec.	Maxell
7:	Sony	Pulstec	Sharp	Nidec	Pixela	Sharp	Ichikoh
8:	Nippon Antenna	Clarion	Toshiba	Sony	ALOKA	ALOKA	Noble
9:	Chuo Seisakusho	Real Vision	Kenwood	NEC	Iwasaki	Nidec	Lasertec
10:	Panasonic	Kenwood	Okii	Sharp	JRC	Japan Radio	Soshin Elec.
11:	Shindengen	Hitachi Medical	Pioneer	Canons	JAE	Hitachi Medical	HitachiKokusaiElec
12:	Leader Kikusui Elec.		Sony	Sanyo	Mutoh	NIEC	Minebea
13:	Fujitsu	Ikegami	Sanyo	Kenwood	Ikegami	Okii	Twinbird
14:	Canon Elec.	Toshiba	Omron	Yokogawa	Shinko	Fuji Elec.	Daido Signal
15:	Nagoya Component	Fujitsu	Canon	Takaoka Elec.	Optex	Fanuc Ltd.	Omron
16:	ADTEC	Sony	Nidec	Victor	ENPLAS	Elpida	Toyo Denki

17:	Murata	Noda	Screen	Victor	Ricoh	D&M	Canon	Shindengen
18:	HDK	JRC	Tietech	Okai	Casio	Coito	SPC	
19:	NEC	SPC	Kyocera	D&M	Shindengen	JEOL	Meidensha	
20:	Victor	Epson	Casio	Keyence	FB	Clarion	ETA Elec.	
		Toyocom						

Table 4. Top 20 companies ranked by degree centrality on different social networks in the electrical industry.

$r_{i,Cb}$	$r_{cooc,Cb}$	$r_{overlap,Cb}$	$r_{business,Cb}$	$r_{capital,Cb}$	$r_{market,Cb}$	$r_{shareholder,Cb}$	$r_{age,Cb}$
1:	NIEC	Shindengen	Hitachi	Hitachi	Shinko	Hitachi	Konica Minolta
2:	JEOL	Keyence	Fujitsu	Suzuki	Eneserve	Fujitsu	Brother
3:	Toshiba	Casio	Mitsubishi	Mitsubishi	Konica Minolta	Mitsubishi	NetIndex
4:	Sony	HDK	Omron	Fujitsu	Hitachi	Panasonic	Sanyo Denki
5:	Fujitsu	JAE	Sharp	Nidec	Ibiden	Panasonic Elec.	Minebea
6:	ADTEC	JEOL	Panasonic	Toshiba	Nishishiba Elec.	Elpida	Hitachi
7:	Mitsubishi	Murata	Suzuki	NEC	Brother	Clarion	Mitsubishi
8:	Chuo Seisakusho	Toshiba	Okai	Takaoka Elec.	Noda Screen	ALOKA	Daiichi Seiko
9:	JEM	Sony	Nidec	Canon	Energy Support	Japan Radio	Shinko
10:	NEC	Mitsubishi	NEC	Sanyo	Showa KDE	JEOL	Ibiden
11:	Panasonic	Pulstec	Sony	Panasonic Elec.	Toyo Elec.	Toshiba	Morio Denki
12:	Pioneer	Kenwood	Toshiba	Kenwood	Tabuchi Elec.	Sharp	Nidec
13:	Panasonic Elec.	Real Vision	Sanyo	Okai	Sophia	Koito	Showa KDE
14:	JDL	NEC	Kenwood	Sharp	NIEC	Hitachi Medical	Syswave
15:	Nagoya	Hitachi Medical	Pioneer	Yokogawa	Ferrotec	Nidec	MCJ
16:	Sharp	Fujitsu	Keyence	Ricoh	Shibaura	Tabuchi Elec.	Origin Elec.
17:	Japan Radio	Suzuki	Panasonic Elec.	Panasonic	Santec	Canon	Sophia
18:	Canon Elec.	SEIWA	Toko.	Casio	Hamamatsu	NIEC	NIEC
19:	I-O Data	ADTEC	Maxell	Brother	Mimaki	Okai	Ferrotec
20:	Canon	JEM	Japan Radio	Omron	Enomoto	TDK	Shibaura

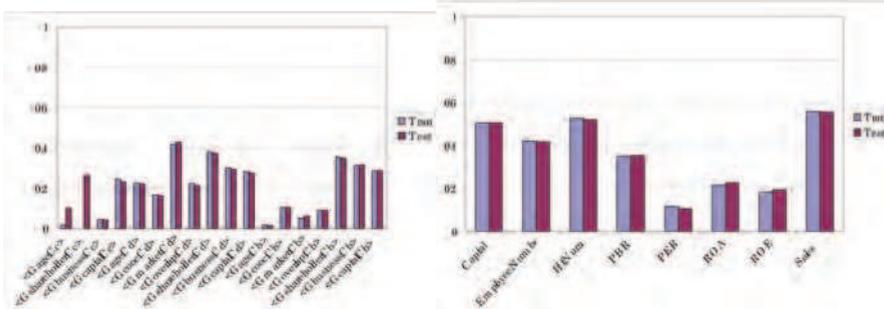
Table 5. Top 20 companies ranked by betweenness centrality on different social networks in the electrical industry.

r_A	Capital	Employee Number	Sales	PER	PBR	ROA	ROE	HitNum
1:	Sony	Hitachi	Panasonic	ENPLAS	Sanyo	AXELL	Nagano JRC	NEC
2:	NEC	Panasonic	Sony	Santec	Meisei	Keyence	KKDI-Nikko Eng.	Sony
3:	Fujitsu	Toshiba	Toshiba	FDK	Tokki	OptexFA	TEAC	Suzuki
4:	Sanyo	Sony	Fujitsu	SK Elec.	TEAC	Canon Elec.	AXELL	Toshiba
5:	Hitachi	Fujitsu	Canon	NEC	Wacom	TEL	Yaskawa	Sharp
6:	Toshiba	NEC	Sharp	Sanko	Ibiden	Lasertec	Tabuchi Elec.	Fujitsu
7:	Panasonic	Canon	Hitachi	Fujitsu General	SPC	Roland DG	Fujitsu Component	Pioneer
8:	Sharp	Mitsubishi	Mitsubishi	Seiko Giken	Japan Servo	Hioiki E.E.	Canon Elec.	Canon
9:	Mitsubishi	Sanyo	NEC	ALPS	AXELL	Wacom	Shinko	Mitsubishi
10:	Canon	Nidec	Sanyo	Maxell	Roland DG	Cosel	Konica Minolta	Ricoh
11:	Elpida	Seiko Epson	Ricoh	Sony	Yaskawa	Ibiden	TEL	Hitachi
12:	Panasonic Elec.	Ricoh	Panasonic Elec.	Anritsu	Nidec	Nidec-READ	Epson Toyocom	Panasonic
13:	Ricoh	Kyocera	TEL	Hosiden	Nagano JRC	Fanuc Ltd.	NEC	Kyocera
14:	Kyocera	TDK	Seiko Epson	UNIPULSE	Tabuchi Elec.	Optex	Lasertec	Sanyo
15:	Rohm	Panasonic Elec.	NEC Elec.	Raytex	Sysmex	Syswave	Ibiden	Panasonic Elec.
16:	NEC Elec.	Minebea	Pioneer	Iwasaki	Mimaki	JEM	Wacom	Omron
17:	Okai	Sharp	Kyocera	Miyakoshi	Keyence	Canon	Elpida	AXELL
18:	Murata	Mabuchi Motor	Murata	Wacom	Foster Elec.	CCS	Nidec-READ	Yamatake
19:	Fanuc Ltd.	Mitsumi Elec.	Casio	ETA Elec.	Micronics Japan	Noda Screen	Terasaki	TDK
20:	Minebea	Pioneer	Elpida	Nishishiba Elec.	Hamamatsu	Techno Medica	Mimaki	KEL

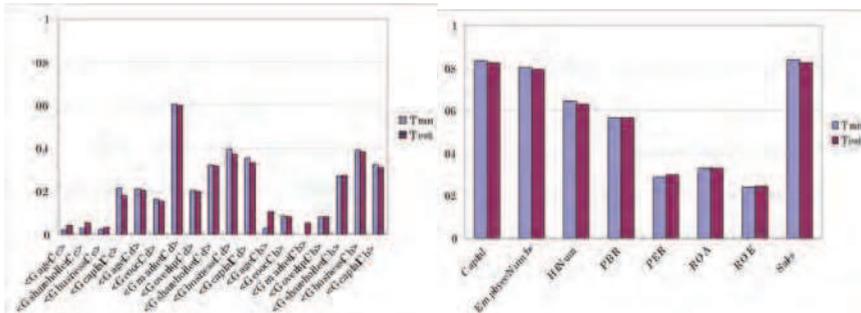
Table 6. Top 20 companies ranked by attributes of companies in the electrical industry.

As a baseline model, we use three centrality indices (i.e., degree centrality C_d , closeness centrality C_c , and betweenness centrality C_b) on different networks (G_{cooc} , $G_{overlap}$, $G_{capital}$, $G_{business}$, $G_{shareholder}$, G_{age} , G_{market}) as network rankings, and calculate the correlation between network rankings with each target ranking: Avg-In, Excellent, and Market-Cap. For comparison, we also rank companies according to previously described attributes (i.e., seven fundamental indices and hit number of names on the Web), and calculate the correlation with target rankings. Fig. 2 presents correlations (mean of three tries) of each network ranking as well as each attribute-based ranking with different target rankings on training

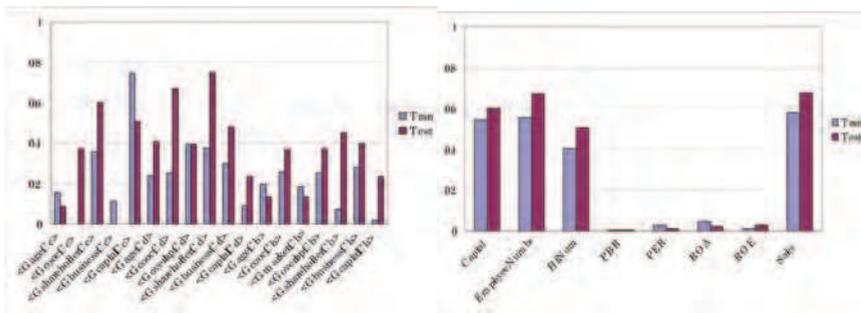
and testing data in the electrical industry. These results demonstrate that rankings of betweenness centrality in same-market network (r_{G_{market}, C_t}) and in shareholding relational network ($r_{G_{shareholder}, C_t}$) have good correlation with the target ranking of Avg-In. Betweenness centralities in the cooc network (r_{G_{cooc}, C_t}), betweenness centralities and degree centralities in the business-alliance network as well as the capital-alliance network ($r_{G_{business}, C_t}$, $r_{G_{capital}, C_t}$, $r_{G_{business}, C_d}$, $r_{G_{capital}, C_d}$) all show good correlation with the target ranking of Market-Cap. Betweenness centralities in the capital-alliance network and shareholding relational network correlate well with the target ranking of Excellent.



(a) Centrality-based rankings with Avg-In (b) Attribution-based rankings with Avg-In



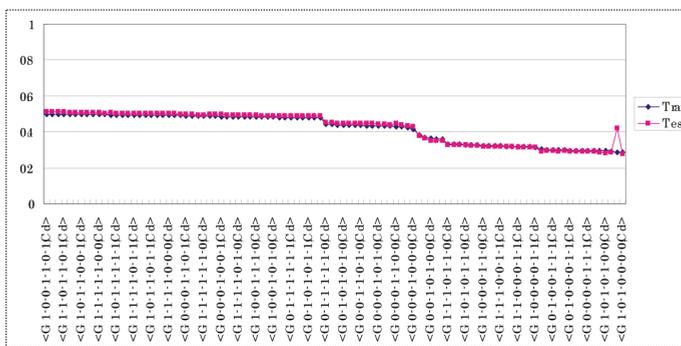
(c) Centrality-based rankings with Market-Cap (d) Attribution-based rankings with Market-Cap



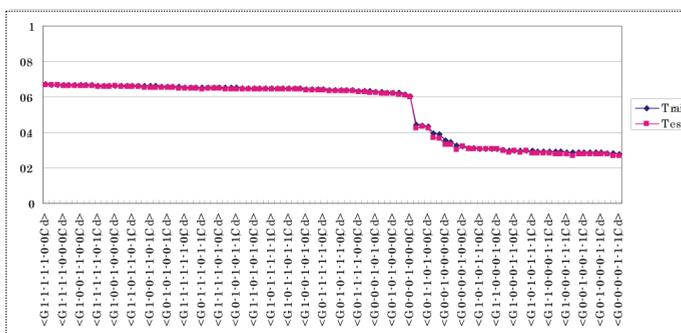
(e) Centrality-based rankings with Excellent (f) Attribution-based rankings with Excellent

Fig. 2. Evaluation for each centrality-based ranking, along with a attribute-based ranking with different target rankings in the electrical industry.

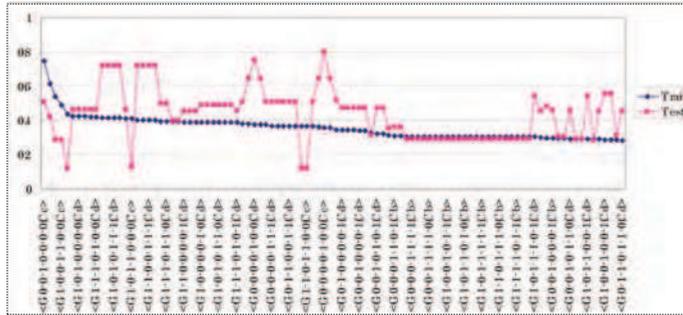
In the combination model, we simply use Boolean type ($w_i \in \{1, 0\}$) to combine relations. Using relations of seven types to combine a network $G_{overlap-business-capital-market-shareholder-age-cooc}$ we can create 2^7-1 (=127) types of combination-relational networks (in which at least one type of relation exists). We obtain network rankings in these combined networks to learn and predict the target rankings. The top 50 correlations between network rankings in combined-relational network and target rankings are presented in Fig. 3. Results demonstrate that degree centralities on combined-relational network produce good correlation with target rankings. For the target ranking of Avg-In, a network ($G_{1-0-0-1-1-0-1}$) comprising overlap relations, same-market relations, shareholding relations, and cooc relations shows good correlation. They outperform the baseline approach. For the target ranking of Market-Cap, the combined-relational networks which combined by overlap relation, capital-alliance relation, same-market relation, and shareholding relation ($G_{1-1-1-1-1-0-0}$, $G_{1-0-1-1-1-0-0}$) show good correlation. For the target ranking of Excellent, closeness centralities in the capital-alliance network outperform other combinations. Future work on how to choose parameter values will yield results that will be especially helpful to practitioners.



(a) Centrality-based rankings with Avg-In



(b) Centrality-based rankings with Market-Cap



(c) Centrality-based rankings with Excellent

Fig. 3. Evaluation for network rankings in a combined-relational network with different target rankings in the electrical industry.

We execute our feature integration ranking model (with several varies) to single and multi-relational social networks to train and predict three different targets rankings: Avg-In, Excellent, and Market-Cap. We use Ranking SVM to learn the ranking model which minimize pairwise training error in the training data; then we apply the model to predict rankings on training data (again) and on testing data. Comparable results for several varieties of model are presented in Table 7. Below we will explain a trial of each and interpret the results. First, we integrate the attributes of companies (i.e., several fundamental indices plus hit number of names on the Web) as features, and treat it as a baseline of feature-integration models to learn and predict the rankings. We can obtain 0.389 correlation for Avg-In, 0.571 correlation for Excellent, and 0.718 correlation for Market-Cap using these attribute-based features. This means that fundamental indices are quite good features for explaining target rankings, and are especially good for Market-Cap. Then, we integrate proposed network-based features obtained from each type of single network as well as multi-relational networks to train and predict the rankings. These results show that integrating the features in the network of G_{market} , G_{age} , $G_{capital}$ yields good performance for explaining the ranking of Avg-In, features in the G_{cooc} , $G_{shareholder}$ explain ranking of Excellent, and features in the G_{market} , $G_{business}$ and $G_{capital}$ have good performance for explaining the ranking of Market-Cap. These results reflect that relations and networks of different types produce different impacts on different target of rankings. Some examples are the following. Listing on the same stock market and connection with similar average-age companies are related to higher average incomes of companies. Co-occurrence with many other companies on the Web, shareholding relations with big companies are associated with a company being more well-known; consequently, the company has an excellent ranking. Active collaboration with other companies through business and capital alliances are associated with higher market value company. Using the features from multi-relational networks G_{ALL} , the prediction results are higher than those of any other single-relational network. This conforms to the intuition that multi-relational networks have more information than single networks to explain real-world phenomena. Furthermore, we combine network-based features with attribute-based features to train the model. The prediction results for any target ranking outperform each of the use of attribute-based features alone or network-based features alone. The correlation with target ranking of Market-Cap improved little from 0.718 (attribute only), 0.645 (network only) to 0.756 (both); the correlation with Avg-In shows

remarkable changes from 0.389 (attribute only), to 0.584 (network only) and 0.601 (both), which means that market values are explained more by fundamental attributes than relations among companies, although average incomes for companies are more understandable according to relations among companies than fundamental indices. The overall results demonstrate that, even though the attribute-based features have good performance for explaining Market-Cap than network-based features, by combining network-based features with attribute-based features, the prediction results are improved. The target rankings of Avg-In and Excellent are more explainable by integrating network-based features than attribute-based features. Demonstrably combining both network and attribute-based features yields further improved prediction results.

Electrical	Feature	Avg-In		Excellent		Market-Cap	
		Train	Test	Train	Test	Train	Test
Network	G_{age}	0.357	0.341	0.443	-0.107	0.361	0.233
	G_{cooc}	0.247	0.120	0.364	0.619	0.346	0.197
	G_{market}	0.535	0.475	0.425	0.357	0.761	0.651
	$G_{overlap}$	0.409	0.284	0.423	0.381	0.519	0.295
	$G_{shareholder}$	0.397	0.190	0.771	0.400	0.514	0.117
	$G_{business}$	0.501	0.182	0.699	-0.500	0.590	0.421
	$G_{capital}$	0.641	0.329	0.818	0.300	0.643	0.350
	G_{ALL}	0.758	0.584	0.912	0.574	0.685	0.645
Attributes	ALL	0.559	0.389	0.811	0.571	0.735	0.718
Network + Attributes	$G_{age}+A$	0.681	0.573	0.762	0.429	0.791	0.710
	$G_{cooc}+A$	0.572	0.396	0.804	0.429	0.725	0.700
	$G_{market}+A$	0.643	0.555	0.746	0.595	0.808	0.754
	$G_{overlap}+A$	0.604	0.418	0.631	0.452	0.745	0.655
	$G_{shareholder}+A$	0.580	0.438	0.739	0.456	0.764	0.625
	$G_{business}+A$	0.596	0.396	0.873	0.619	0.747	0.692
	$G_{capital}+A$	0.592	0.470	0.811	0.524	0.752	0.705
	$G_{ALL}+A$	0.812	0.601	0.947	0.580	0.811	0.756

Table 7. Results of feature integration in the electrical industry.

5.3 Detailed Analysis of Useful Features

We use network-based features separately to train and expect the target rankings to clarify their usefulness. Leaving out one feature, the others are used to train and predict the rankings to evaluate network-based features. In fact, k -th feature is a useful feature for explaining the target ranking if the result worsens much when leaving out the feature k from the feature set. Table 8 presents the effective features for the different target rankings of Market-Cap, and Excellent, respectively, in company networks. For example, the maximum number of links in the neighbor nodeset of x from overlap network $Max \circ \gamma \circ C_x^{(1)} \circ G_{overlap}$ is effective for the target ranking of Avg-In, which means that if a famous company is reachable from a company, the company's income can be more high. The ratio of the sum of paths through x among neighbors to the sum of paths through x among reachable nodes from overlap network $Ratio \circ (Sum \circ u_x \circ C_x^{(1)}, Sum \circ u_x \circ C_x^{(\infty)}) \circ G_{overlap}$ is effective for the

target ranking of Market-Cap, which means that maintaining high betweenness among neighbors from all of reachable nodes in the Web makes the company' market value higher. The minimum number of edges among reachable companies from the business-alliance network $Min \circ s^{(1)} \circ C_x^{(\infty)} \circ G_{business}$ is effective for the target ranking of Excellent, which means that x will be more excellent when the reachable companies have little business-alliance among them.

Top	Features for Avg-In	Features for Market-Cap	Features for Excellent
1	$Max \circ \gamma \circ C_x^{(1)} \circ G_{overlap}$	$Ratio \circ (Sum \circ u_x \circ C_x^{(1)}, Sum \circ u_x \circ C_x^{(\infty)}) \circ G_{overlap}$	$Min \circ s^{(1)} \circ C_x^{(\infty)} \circ G_{business}$
2	$Min \circ s^{(1)} \circ C_x^{(\infty)} \circ G_{business}$	$Min \circ s^{(1)} \circ C_x^{(\infty)} \circ G_{shareholder}$	$Max \circ s^{(1)} \circ C_x^{(1)} \circ G_{business}$
3	$Avg \circ u_x \circ C_x^{(1)} \circ G_{capital}$	$Avg \circ \gamma \circ C_x^{(\infty)} \circ G_{business}$	$Ratio \circ (Max \circ s^{(1)} \circ C_x^{(1)}, Max \circ s^{(1)} \circ C_x^{(\infty)}) \circ G_{business}$
4	$Max \circ \gamma \circ C_x^{(\infty)} \circ G_{age}$	$Max \circ \gamma \circ C_x^{(\infty)} \circ G_{business}$	$Avg \circ t_x \circ C_x^{(1)} \circ G_{capital}$
5	$Avg \circ \gamma \circ C_x^{(\infty)} \circ G_{capital}$	$Min \circ \gamma \circ C_x^{(\infty)} \circ G_{business}$	$Max \circ t_x \circ C_x^{(1)} \circ G_{capital}$
6	$Ratio \circ (Sum \circ u_x \circ C_x^{(1)}, Sum \circ u_x \circ C_x^{(\infty)}) \circ G_{age}$	$Ave \circ s^{(1)} \circ C_x^{(1)} \circ G_{cooc}$	$Min \circ t_x \circ C_x^{(1)} \circ G_{capital}$
7	$Max \circ \gamma \circ C_x^{(\infty)} \circ G_{market}$	$Max \circ t \circ C_x^{(1)} \circ G_{market}$	$Min \circ t_x \circ C_x^{(\infty)} \circ G_{capital}$
8	$Ave \circ s^{(1)} \circ C_x^{(1)} \circ G_{business}$	$Max \circ \gamma \circ C_x^{(\infty)} \circ G_{market}$	$Ratio \circ (Min \circ t_x \circ C_x^{(1)}, Min \circ t_x \circ C_x^{(\infty)}) \circ G_{capital}$
9	$Avg \circ u_x \circ C_x^{(\infty)} \circ G_{capital}$	$Avg \circ \gamma \circ C_x^{(\infty)} \circ G_{shareholder}$	$Max \circ t_x \circ C_x^{(\infty)} \circ G_{capital}$
10	$Avg \circ u_x \circ C_x^{(\infty)} \circ G_{age}$	$Sum \circ t \circ C_x^{(\infty)} \circ G_{shareholder}$	$Max \circ s^{(1)} \circ C_x^{(\infty)} \circ G_{capital}$
11	$Min \circ \gamma \circ C_x^{(1)} \circ G_{cooc}$	$Avg \circ \gamma \circ C_x^{(\infty)} \circ G_{capital}$	$Avg \circ t_x \circ C_x^{(1)} \circ G_{cooc}$
12	$Sum \circ \gamma \circ C_x^{(1)} \circ G_{cooc}$	$Max \circ t \circ C_x^{(\infty)} \circ G_{capital}$	$Max \circ t_x \circ C_x^{(1)} \circ G_{cooc}$
13	$Min \circ \gamma \circ C_x^{(\infty)} \circ G_{business}$	$Avg \circ u_x \circ C_x^{(\infty)} \circ G_{age}$	$Max \circ s^{(1)} \circ C_x^{(\infty)} \circ G_{cooc}$
14	$Ratio \circ (Avg \circ s^{(1)} \circ C_x^{(1)}, Avg \circ s^{(1)} \circ C_x^{(\infty)}) \circ G_{business}$	$Min \circ s^{(1)} \circ C_x^{(1)} \circ G_{age}$	$Max \circ \gamma \circ C_x^{(1)} \circ G_{cooc}$
15	$Max \circ t \circ C_x^{(1)} \circ G_{age}$	$Min \circ \gamma \circ C_x^{(\infty)} \circ G_{age}$	$Max \circ \gamma \circ C_x^{(\infty)} \circ G_{cooc}$

Table 8. Effective features in various networks for Avg-In, Market-Cap, and Excellent, respectively, among companies.

We understand that various features have been shown to be important for real-world rankings (i.e. target ranking). Some of them correspond to well-known indices in social network analysis. Some indices seem new, but their meanings resemble those of the existing indices. The results support the usefulness of the indices that are commonly used in the social network literature, and underscore the potential for additional composition of useful features.

Summary:

Several conclusions are suggested by the experimental results presented above: Social networks vary according to different relational indices or types even though they contain the same list of companies; Companies have different centrality rankings even though they are in the same type of relational networks: Relations and networks of different types differently impact on different targets of rankings: Multi-relational networks have more information than single networks to explain target rankings. Well-chosen attribute-based features have good performance for explaining target rankings. However, by combining proposed network-based features, the prediction results are further improved: various network-based features have been shown to be important for real-world rankings (i.e., target ranking), some of which correspond to well-known indices in social network analysis such as degree centrality, closeness centrality, and betweenness centrality. Some indices seem new, but their meanings resemble those of the existing indices.

6. Related Works

Recently, many studies deal with social networks among various online resources such as social network services (SNSs) (Zhou et al. 2008), online Instance Messengers (IM) (Singla & Richardson, 2008), as well as Friend-of-a-Friend (FOAF) instances (Ding et al., 2005; Finin et al., 2005). Unfortunately, these resources are not specifically applicable to relations among companies or other organization structures. However, many relations among companies are published on the Web in news articles or news releases. Our work emphasizes the investigation of such published relations on the Web. A news site might deal little with information related to small companies and foreign corporations. Therefore, we use a search engine to extract more coverable relations among any given set of companies.

The location of actors in multi-relational networks and the structure of networks composed of multiple relations are interesting areas of SNAs. Recent efforts to address this problem adopt consideration of multi-modal networks---a network composed of a set of different kind of nodes---and mainly consider the relations among these nodes (Nie et al., 2005; Rodriguez, 2007; Zhou et al., 2008). They usually use papers, authors, and conferences (or journals) as different types of nodes, and considering the relational impact from different models (or layers) paper-paper, paper-author, as well as paper-conference (or journal) relations to calculate document similarity for document recommendation as well as support the scholarly communication process. This paper presents different views of multi-relational networks comprising multiple different kinds of relations (ties) among the same set of social actors (nodes) to elucidate what kinds of relations are important, as well as what kinds of relational combinations are important.

In the context of information retrieval, PageRank (Page et al., 1998) and HITS (Kleinberg, 1998) algorithms can be considered as well known examples for ranking Web pages based on the link structure. Recently, more advanced algorithms have been proposed for ranking

entities. Several studies have examined learning certain relational weights as conductance of Markovian walks on a network, given preference orders over nodes using gradient descent (Chang et al., 2000), error back-propagation (Diligenti et al., 2005), and approximate Newton method (Chakrabarti & Agarwal, 2006). Our networks are social networks with connections among nodes according to relations. Therefore, we neither give assumptions that the network must be a Markov network nor that the weight is positive only (because negative relations such as a lawsuit relation might damage the company ranking). Furthermore, our model is target-dependent: the important features of relations and structural embeddedness vary among different tasks.

Relations and structural embeddedness influence behavior of individuals and growth and change of the group (Singla & Richardson, 2008). Several researchers use network-based features for analyses. L. Backstrom et al. (Backstrom et al., 2006) describe analyses of community evolution, and show some structural features characterizing individuals' positions in the network. D. Liben-Nowell et al. (Liben-Nowell & Kleinberg, 2003) elucidate features using network structures for link prediction in the link prediction problem. We specifically examine relations and structural features for individuals and deal with various structural features from multi-relational networks systematically. Our generated features include those described in works from Backstrom and Liben-Nowell. Our approaches are similar to text classification given the document features and correct categories. Features are designed beforehand. Similarly, the relation is defined beforehand. The classifier learns the model to predict the given categories. Similarly, the ranking is given and is used for learning. Specifically regarding feature weights, we can understand which features are important for categorization, thereby yielding a better classification model. Furthermore, examining the weights of each relation, we can understand which relations are important for ranking. Cai et al. (Cai et al., 2005) regarded a similar idea with this approach: They try to identify the best combination of relations (i.e., relations as features) which makes the relation between the intra-community examples as tight as possible. Simultaneously, the relation between the inter-community examples is as loose as possible when a user provides multiple community examples (e.g., two groups of researchers). However, our purpose is learning of a ranking function (e.g., ranking of companies) based on social networks, which has a different optimization task. Moreover, we propose innovative features for entities based on the combination or integration of structural importance generated from social networks. However, our purpose is learn the ranking function (e.g. ranking of companies) based on social networks, which has different optimization task. Moreover, we propose innovative features for entities based on combination or integration of structural importance generated from social networks.

7. Conclusion

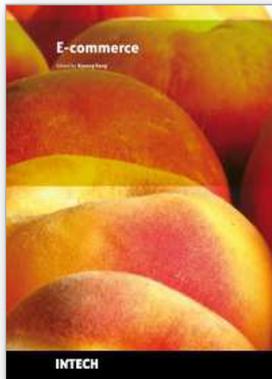
This chapter described methods of learning the ranking of entities from multiple social networks mined from the Web. Various relations and relational embeddedness pertain to our lives: their combinations and their aggregate impacts are influential to predict features of entities. Based on that intuition, we constructed our ranking learning model from social networks to predict the ranking of other actors. We first extracted social networks of different kinds from the Web. Subsequently, we used these networks and a given target ranking to learn the model. We proposed three approaches to obtaining the ranking model.

Results of experiments using two domains (i.e., companies in the electrical industry in Japan and researchers in The University of Tokyo) reveal that effectiveness of our models for explaining target rankings of actors using multiple social networks mined from the Web. Our models provide an example of advanced utilization of a social network mined from the Web. The results underscore the usefulness of our approach, by which we can understand the important relations as well as important structural embeddedness to predict the rankings. We use multiple social networks extracted from the Web, which are more realistic than a single relational network. In addition, the model can be combined with a conventional attribute-based approach. Our model provides an example of advanced utilization of a social network mined from the Web. More kinds of networks and attributes for various target rankings in different domains can be designated for improving the usefulness of our models in the future.

8. References

- Aleman-Meza, B.; Nagarajan, M.; Ramakrishnan, C.; Sheth, A.; Arpinar, I.; Ding, L.; Kolari, P.; Joshi, A. & Finin, T. (2006). Semantic Analytics on Social Networks: Experiences in Addressing the Problem of Conflict of Interest Detection, *15th International World Wide Web Conference (WWW2006)*, Edinburgh, Scotland, May 2006
- Backstrom, L.; Huttenlocher, D.; Lan, X. & Kleinberg, J. (2006) Group formation in large social networks: Membership, Growth, and Evolution, *The 12th Annual SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD'06)*, Philadelphia, USA, August 2006
- Cai, D.; Shao, Z.; He, X.; Yan, X. & Han, J. (2005). Mining Hidden Community in Heterogeneous Social Networks, *Workshop on Link Discovery: Issues, Approaches and Applications (LinkKDD-2005)*, Chicago, IL, USA, August 2005
- Chakrabarti, S. & Agarwal, A., (2006). Learning Parameters in Entity Relationship Graphs from Ranking Preferences, *European Conference on Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD 2006)*, Vol. 4213, pp. 91-102, Berlin, Germany, September 2006
- Chang, H.; Cohn, D. & McCallum, A. (2000). Creating customized authority lists, *17th International Conference on Machine Learning (ICML2000)*, Stanford, CA, USA June 2000
- Diligenti, M.; Gori, M. & Maggini, M. (2005). Learning Web page scores by error back-propagation, *19th International Joint Conference on Artificial Intelligence (IJCAI-05)*, Edinburgh, Scotland, July 2005
- Ding, L.; Zhou, L.; Finin, T. & Joshi, A. (2005). How the Semantic Web is Being Used: An Analysis of FOAF, *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*, Waikoloa, Hawaii, January 2005
- Finin, T.; Ding, L.; Zhou, L. & Joshi, A. (2005). Social Networking on the Semantic Web, *The Learning Organization*, Vol. 5, No. 12, pp. 418-435
- Jin, Y.; Matsuo, Y. & Ishizuka, M. (2008). Extracting Inter-Firm Networks from World Wide Web Using General-Purpose Search Engine, *Journal of Online Information Review*, Vol. 32, No. 22, pp. 196-210, ISSN: 1468-4527
- Kautz, H.; Selman, B. & Shah, M. (1997). The Hidden Web, *AI magazine*, Vol. 18, No. 2, pp. 27-35, 1997

- Kleinberg, J.M. (1998). Authoritative Sources in a Hyperlinked Environment, *Proc. ACM-SIAM Symposium on Discrete Algorithms*, pp. 668-677, 1998
- Liben-Nowell, D. & Kleinberg, J. (2003). The link prediction problem for social networks, *Proceedings of the 2003 ACM CIKM International Conference on Information and Knowledge Management*, New Orleans, Louisiana, USA, November 2003
- Matsuo, Y.; Mori, J.; Hamasaki, M.; Ishida, K. Nishimura, T.; Takeda, H.; Hasida, K. & Ishizuka, M.(2006). POLYPHONET: an advanced social network extraction system, *15th International World Wide Web Conference (WWW2006)*, Edinburgh, Scotland, May 2006
- Mika, P. (2005). Flink: semantic web technology for the extraction and analysis of social networks, *Journal of Web Semantics*, Vol. 3, No. 2, pp. 211-223, 2005
- Nie, Z.; Zhang, Y.; Wen, J.-R. & Ma, W.-Y. (2005). Object-level ranking: Bringing order to Web objects, *14th International World Wide Web Conference (WWW2008)*, Tokyo, Japan, May 2005
- Page, L.; Brin, S.; Motwani, R. & Winograd, T. (1998). The PageRank Citation Ranking: Bringing order to the Web, *Technical report*, Stanford Digital Libraries, 1998
- Rodriguez, M.A. (2007). A Multi-relational Network to Support the Scholarly Communication Process, *International Journal of Public Information Systems*, Vol. 2007, Issue 1, ISSN: 1653-4360, 2007
- Singla, P. & Richardson, M. (2008). Yes, there is a correlation - from social networks to personal behavior on the web., *17th International World Wide Web Conference (WWW2008)*, Beijing, China, April 2008
- Spearman, C. (1904). The proof and measurement of association between two things, *American Journal of Psychology*, Issue 15, pp. 72-101, 1904
- Wasserman, S.& Faust, K. (1994). Social network analysis: methods and applications, Cambridge University Press, ISBN: 9780521387071, Cambridge
- Zhao, Z. & Liu, H. (2007). Searching for Interacting Features, *20th International Joint Conference on Artificial Intelligence (IJCAI-07)*, Hyderabad, India, January 2007
- Zhou, D.; Zhu, S.; Yu, K.; Song, X.; Tseng, B.L. Zha, H. & Giles, C.L. (2008). Learning Multiple Graphs for Document Recommendations, *17th International World Wide Web Conference (WWW2008)*, Beijing, China, April 2008



E-commerce

Edited by Kyeong Kang

ISBN 978-953-7619-98-5

Hard cover, 284 pages

Publisher InTech

Published online 01, February, 2010

Published in print edition February, 2010

E-commerce provides immense capability for connectivity through buying and selling activities all over the world. During the last two decades new concepts of business have evolved due to popularity of the Internet, providing new business opportunities for commercial organisations and they are being further influenced by user activities of newer applications of the Internet. Business transactions are made possible through a combination of secure data processing, networking technologies and interactivity functions. Business models are also subjected to continuous external forces of technological evolution, innovative solutions derived through competition, creation of legal boundaries through legislation and social change. The main purpose of this book is to provide the reader with a familiarity of the web based e-commerce environment and position them to deal confidently with a competitive global business environment. The book contains a numbers of case studies providing the reader with different perspectives in interface design, technology usage, quality measurement and performance aspects of developing web-based e-commerce.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Yingzi Jin, Yutaka Matsuo and Mitsuru Ishizuka (2010). Ranking Companies Based on Multiple Social Networks Mined from the Web, E-commerce, Kyeong Kang (Ed.), ISBN: 978-953-7619-98-5, InTech, Available from: <http://www.intechopen.com/books/e-commerce/ranking-companies-based-on-multiple-social-networks-mined-from-the-web>

INTECH

open science | open minds

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2010 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License](#), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.