Journal of Organizational and End User Computing
Volume 35 • Issue 1

A Text-Based Competition Network:
The Perspective of Information Disclosure

Wei Wang, Chongqing University, China
https://orcid.org/0000-0001-6895-4466

Fengzhang Chen, Chongqing University, China

Zewei Long, Chongqing University, China

Fengwen Chen, Chongqing University, China*
https://orcid.org/0000-0002-7852-4504

Fu-Sheng Tsai, North China University of Water Resources and Electric Power, China & Department of Business Administration, Center for Environmental Toxin and Emerging-Contaminant Research, China & Super Micro Mass Research and Technology Center, Cheng Shiu University (CSU), Taiwan

ABSTRACT

This paper utilizes nonfinancial information disclosure to develop a measure of text-based competition network. Using the data of China's listed firms, the authors adopt the textual analysis method to identify a unique group of competitors for the focal firm and construct the text-based competition network. In the whole network, leading firms receive increasing attention from competitors, and they play a vital role for the dynamic changes in the whole market. Moreover, the interactions between the focal firm and competitors in the text-based competition network are shown by some financial indicators. The characteristics of the text-based competition network have a significant impact on the future performance of the focal firm. Finally, economic links in the competition network are discussed by varying the number of competitors, which shows the impact of various competitors on economic similarities. The text-based competition network shows the relative importance of competitors for the focal firm and explains firms' decision-making from the perspective of dynamic competition.

KEYWORDS

Competitors, Economic Links, Information Disclosure, Text-Based Competition Network, Textual Analysis

INTRODUCTION

As a special kind of interaction between firms, competition is the determinant of business strategy and represents contested goals (Galvin et al., 2020). The emergence of competition may be determined by some factors. In terms of the product market, one firm will face similar suppliers and customers with competitors, and the competitive environment faced by this firm is built by its products or services (Hoberg & Phillips, 2016). Some competitive relationships may come from other sources,

DOI: 10.4018/JOEUC.317138 *Corresponding Author

such as labor, patents, and prices (Galvin et al., 2020). According to the resource-based view, unique and rare resources will encourage firms to design competitive actions and reactions, which may determine their survival. During this process, competition emphasizes limiting imitation from competitors, and the close relationship between firms and their competitors will be built by utilizing the same or similar resources (Barney, 1986). The reactions of a firm to its competitors are directly influenced by the decision-making and performance of competitors, which may produce imitation among such firms (Porter, 1989). The core of exploring competition is to explain how firms engage in some competitive actions and reactions, which are represented by the structure of the whole market (Andrevski et al., 2016).

The identification of competitors is a key step in exploring the competitive relationships between firms. In Porter's (1989) theory, the factors related to competition mainly come from the product market where firms compete for customers or suppliers, and these factors will determine the future development goals of such firms. Consistent with this view, the process of identifying competitors depends on the identification of the product market; that is, firms in the same product market are engaged in competition for economic surplus, produced by their goods or services (Salop, 1979). Many studies have used industry classification to determine the boundary of the product market and identify the competitors of one firm by industry tags (Chen et al., 2017; Hrazdil et al., 2014; Katselas et al., 2019). The common industry classification systems include the Guidance for Industry Classification of Listed Companies released by the China Securities Regulatory Commission (CSRC), the Standard Industrial Classification (SIC), and the Global Industry Classification Standard (GICS). The decision-making of a firm is regarded as its reactions to the actions of competitors in the same industry, which explain some behaviors of this firm, such as capital structure (Phillips & Mackay, 2005), earnings management (Kedia et al., 2015), information disclosure (Lin et al., 2018), and corporate social responsibility (Cao et al., 2019). From the nature of industry classification, the measure of competition is mainly based on industry boundaries.

According to the theory of industrial organization, the industry's degree of competition can be displayed through the concentration or differentiation of products (Li et al., 2013). The Herfindahl-Hirschman Index (HHI) is often used to describe the intensity of competition for firms in the same industry. Many existing studies have discussed the effectiveness of identifying competitors based on industry classification, but there are some obvious limitations in this method (Engelberg et al., 2018; Kang et al., 2022). First, the update cycle of industry classification systems is long, and the fixed industry boundary may make it difficult to match the relationship between industry tags and economic activities. Second, the unique competitive environment of individual firms cannot be described directly by industry classification, so it is difficult to measure the relative importance of competitors for one firm. Third, some firms may participate in one or more product markets, and the traditional industry classification cannot measure the competition between firms belonging to different product markets. For example, Apple Inc. competes with Lenovo Group in the laptop computer market but also competes with Huawei Technologies Co. in the mobile phone market. Similarly, the various products of Amazon will also create different competitive relationships in the e-commerce market or e-book market. It is worth noting that an increasing number of listed firms in China's retail industry have entered the real estate and medical industries, and firms in different markets may have vague relationships with their competitors based on industry classification. In this situation, the industry classification system has difficulty identifying competitors of individual firms and is unable to measure their relative importance.

Considering the nature of competition, when two or more firms have a similar development goal, these firms are engaged in competition (Medlin & Ellegaard, 2015). This view has been demonstrated in Text-based Network Industry Classifications (TNIC) proposed by Hoberg & Phillips (2016); that is, there is a competitive relationship between firms with similar products. From the idea of TNIC, the similarities of products create the competitive environment faced by focal firms from a micro perspective, which also shows the social interactions between firms (Lee et al., 2015). Many studies have described the relationship between a focal firm and its competitors from various dimensions, including product words (Hoberg & Phillips, 2016), competition words (Li et al., 2013), search behaviors (Lee et al., 2015), and technological links (Lee et al., 2019). In these methods, identifying competitors for each focal firm should capture their close relationships in the first step, and then build the social networks for these firms. These methods intend to obtain the competitive environment of individual firms, which identifies competitors subject to managers' constant attention (Kaustia & Rantala, 2015).

In this paper, we utilized the text of nonfinancial information disclosure to develop a measure of a text-based competition network. First, the descriptive text in annual reports was used to represent the operation and development goals of firms, and the degree of competition between firms was measured by the semantic similarity of such text. Second, considering managers' constant attention, the top *k* similar firms of the focal firm were defined as the focal firm's competitors according to the semantic similarity of descriptive text. Finally, based on the social network theory, we used the links between the focal firm and its competitors to build the text-based competition network. Compared with the existing methods of identifying competitors, our method pays more attention to the impact of text semantics on the degree of competition, which is different from the TNIC method based on product words. Based on the text of nonfinancial information disclosure in annual reports, we wanted to test whether firms with similar text semantics seek similar development goals and provide similar products. A neural network language model was used to capture the text semantics of annual reports; the advantage of this method is that it retains the word order and semantic information of descriptive text (Bengio et al., 2003).

Using the annual reports of China's listed firms, we measured the degree of competition annually between firms, and constructed the text-based competition network for each focal firm. Compared with the industry classification system, some competitors obtained by our method were also the industry competitors for focal firms, indicating that the nonfinancial disclosures in annual reports represent the main operating activities of firms. In addition, many leading firms in the competition network receive increasing attention from other firms, and they may control the development of the whole market. According to the analysis of economic links, our method better captures the relative importance between the focal firm and its competitors in some financial indicators. In further tests, the characteristics of the text-based competition network have a significant impact on the future performance of the focal firm, demonstrating that the interactions between firms determines whether the focal firm will achieve the goals of long-term development. Moreover, we used stock return comovement to discuss the number of competitors for each focal firm and showed that the manager of the focal firm will pay more attention to some important competitors.

The rest of this paper is organized as follows. The Methodology section introduces the methods used in this paper, including the method of textual analysis and the construction of a competition network. The Data section describes the data, including the data sources and the data description. The Results section presents the results, including the analysis of network characteristics and economic links, the discussion of focal firms' future performance, and the number of competitors. The conclusion will be drawn in the Conclusion section.

METHODOLOGY

Textual Analysis

The Application of Textual Analysis in Identifying Competitors

The method of textual analysis is the process of applying an algorithm to change the representation of text data and obtain the vectors to represent these texts (Loughran & Mcdonald, 2016). Compared with financial data, text data in information disclosure more objectively shows the operations of firms and helps investors understand firms' development goals. Many researchers apply the textual analysis method to measure the competition from two dimensions: competition-related and product-related words (Hoberg & Phillips, 2016; Li et al., 2013).

In terms of competition-related words, Li et al. (2013) used the frequency of such words in information disclosures to measure the degree of competition perceived by managers, and showed the external pressure faced by individual firms. This method builds the competitive environment based on the perceived pressure of managers and relies on the quality of information disclosure to measure the intensity of competition (Shi et al., 2018). In terms of product-related words, Hoberg & Phillips (2016) used the similarity of product words used in annual reports to measure the competition between firms, and create a unique competitor set for each focal firm. Similar product-related words represent the similar customers and suppliers faced by some firms, which is also the core of Porter's (1989) theory. The extraction of specific words depends on the quality of information disclosure, which guarantees that the application of textual analysis will capture the relationships between firms (Loughran & Mcdonald, 2011).

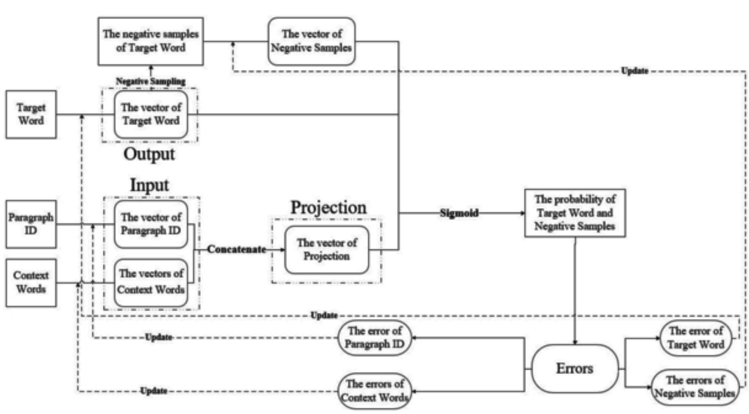
The Method of Paragraph Vector

To capture the semantics of text data, the neural network language model proposed by Bengio et al. (2003) creates links between lexical symbols and lexical semantics by using fixed-length vectors. As an important application of the neural network language model, word2vec uses the vector space to map semantically similar words on similar vector representations (Mikolov et al., 2013). However, word2vec focuses on the semantics of words, so it cannot take into consideration the word order in sentences or documents. Faced with this dilemma, Le & Mikolov (2014) propose paragrah2vec to obtain paragraph vectors for capturing the semantics of sentences or documents for texts of any length.

The paragraph2vec method mainly includes two models: Distributed Memory model (DM) and Distributed Bag of Words model (DBOW). DM focuses on the word orders and semantics of documents, while DBOW focuses on the overall semantics of documents. Because the annual reports of listed firms are composed of professional text, different word orders of the text may change readers' understanding of information disclosure. In this situation, DM captures the word order and semantic information contained in nonfinancial information disclosure and better shows the operation and development of listed firms. Therefore, we selected DM to vectorize the descriptive text of annual reports for each firm. The process of DM is shown in Figure 1.

The core idea of DM is to use the vector of Paragraph ID which is the index of document for firm i $\left(v_{DM_{firm_i}}\right)$ and the vectors of Context Word $\left\{v_{w_{p-n_{window}}},\dots,v_{w_{p-1}},v_{w_{p+1}},\dots,v_{w_{p+n_{window}}}\right\}$ to predict the

Figure 1. The process of Distributed Memory (DM) model



vector of Target Word $\left(v_{w_p}\right)$, where $v_{DM_{firm_i}} \in R^{n_{size}}$ represents the paragraph vector of firm *i* trained by the Distributed Memory model; n_{size} represents the dimension of the paragraph vector; v_{w_p} represents the vector of the Target Word with an index *p*; and n_{window} represents the number of context words for the Target Word. For the word order information in the descriptive text of each annual report, we concatenated the paragraph vector and the Context Word vectors to construct a vector with a higher dimension:

$$V_{input} = concat\left(v_{DM_{firm_i}}, v_{w_{p-n_{window}}}, \ldots, v_{w_{p-1}}, v_{w_{p+1}}, \ldots, v_{w_{p+n_{window}}}\right) \tag{1}$$

In Equation (1), $concat\left(\cdot\right)$ represents the concatenation of the paragraph vector with the context of 2 n_{window} words. Then, we obtained the vector of projection with the dimension of $\left(2n_{window}+1\right)\times n_{size}$, that is, $V_{input} \in R^{\left(2n_{window}+1\right)\times n_{size}}$.

During the training process of DM, the prediction of Target Word w_p is to obtain the probability of the occurrence of the paragraph vector and the Context Word vectors, that is, $p\left(w_p | DM_{firm_i}, w_{p-n_{window}}, \ldots, w_{p-1}, w_{p+1}, \ldots, w_{p+n_{window}}\right)$. For the descriptive text of firm *i*'s annual report, the probability of the occurrence of this document is expressed by the joint probability of all words in this document, that is, $p\left(DM_{firm_i}\right) = p\left(w_{i,1}, w_{i,2}, \ldots, w_{i,g}\right)$, where $w_{i,1}$ represents the first word in the descriptive text of firm *i*'s annual report and *g* represents the number of words in this document. According to the distribution theory proposed by Harris (1954), the probability of the occurrence of one word is related to its context words. Therefore, the probability of the descriptive text of firm *i*'s annual report is estimated as follows:

$$p\left(DM_{firm_i}\right) = \prod_{w_p \in corpus_i} p(w_p \mid DM_{firm_i}, w_{p-n_{window}}, \ldots, w_{p-1}, w_{p+1}, \ldots, w_{p+n_{window}}) \tag{2}$$

In Equation (2), $corpus_i$ represents the set of words contained in the descriptive text of firm *i*'s annual report.

Maximizing $p\left(DM_{firm_i}\right)$ is achieved by maximizing the probability of each word, that is, maximizing $p(w_p \mid DM_{firm_i}, w_{p-n_{window}}, \ldots, w_{p-1}, w_{p+1}, \ldots, w_{p+n_{window}})$. According to the method of Le & Mikolov (2014), we used Negative Sampling and gradient ascent to improve the training speed of the paragraph vector to obtain the paragraph vector of firm *i* under the Distributed Memory model, that is, $v_{DM_{firm_i}}$.

The Construction of the Competition Network

The Identification of Competitors

The similarity of the descriptions of operations and development shows the degree of competition between firms (Medlin & Ellegaard, 2015). Increasing their similarities will further aggravate the competition intensity, which is also reflected in some measurements of competition, such as the HHI (Shi et al., 2018). To describe the degree of competition more accurately, we used the paragraph vector to represent the descriptive text in information disclosure and chose the cosine similarity function to measure the degree of similarity between the focal firm and other firms, as shown in Equation (3):

$$Similarity_{i,j} = \frac{V_{DM_{firm_i}} \cdot V_{DM_{firm_j}}}{\left\| V_{DM_{firm_i}} \right\| \times \left\| V_{DM_{firm_j}} \right\|} \quad (3)$$

In Equation (3), $Similarity_{i,j}$ indicates the similarity of the descriptive text semantics of firm *i* and firm *j*. $\left\| V_{DM_{firm_i}} \right\|$ indicates the length of the paragraph vector for firm *i*. According to the cosine similarity function, we gradually calculated the similarity between focal firm *i* and other firms and obtained firm *i*'s similarity results, as shown in Equation (4):

$$SIM\_Results_i = \left\{ Similarity_{i,1}, Similarity_{i,2}, \ldots, Similarity_{i,n} \right\} \quad (4)$$

In Equation (4), $SIM\_Results_i$ represents the similarity results of firm *i* and *n* represents the number of listed firms. Based on Equation (4), we calculated the similarity between focal firm *i* and other firms year by year and obtained the similarity results of each firm during the sample periods.

According to the definition of competition, there is a close relationship between a focal firm and its competitors, which is captured by similar descriptive text related to the operations and development in information disclosure (Hoberg & Phillips, 2016). Most managers only pay more attention to some competitors. Referring to the method of Lee et al. (2015), we defined the top *k* similar firms of focal firm *i* in similarity results as its competitors, as shown in Equation (5):

$$Competitors_i = \left\{ c_{1,i}, c_{2,i}, \ldots, c_{k,i} \right\} \quad (5)$$

In Equation (5), focal firm *i* will face *k* competitors in each year, and their relative importance can be measured by their similarities.

Directed Competition Networks

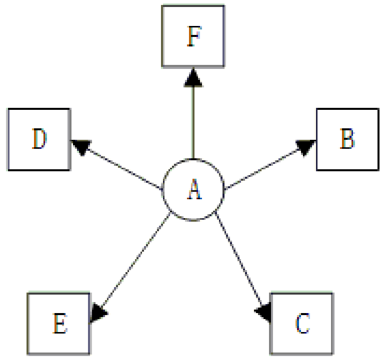
Based on the idea of identifying competitors, we constructed the directed competition network for each focal firm. According to the set of competitors shown in Equation (5), there was a close relationship between the focal firm and its competitors. For these competitors, the manager of the focal firm will pay more attention to them, which is regarded as some directed links from the focal firm to its competitors (Kang et al., 2022). To better describe the competitive environment of individual firms, we constructed an ego network for each listed firm and showed the group of competitors faced by the focal firm, as shown in Figure 2.

In Figure 2, we selected firm A as the focal firm and showed an example of a directed competition network. To show the links between the focal firm and its competitors, we set *k* as 5 and identify five competitors (B, C, D, E, and F). In this directed competition network, the manager of firm A will pay constant attention to these competitors, so that there may be some interactions between such firms.

Undirected Competition Networks

In the construction of a directed competition network, we used the method of identifying competitors to determine the direction of links. According to social network theory, the links between firms are also bidirectional; that is, competitors may choose the focal firm as their competitor (Eklinder-Frick et al., 2011). These bidirectional links reflect the reciprocity between the focal firm and some of its competitors, which are seen as stable relationships (Wincent et al., 2010). For example, Ford Motor

Figure 2. The directed competition network of focal firm A (k = 5)

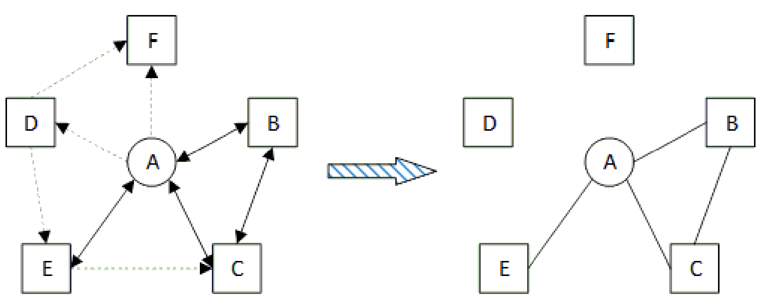


Co. and Volkswagen Group did not regard Tesla Inc. as a competitor in the early stage, so they did not respond to Tesla's actions. In contrast, Tesla responded to the actions of Ford and Volkswagen. When Tesla became Ford and Volkswagen's competitor, Tesla's actions caused Ford and Volkswagen to react. In this situation, bidirectional links motivated more interactions between the focal firm and its competitors.

Based on the construction of the directed competition network, we selected the bidirectional relationships between the focal firm and its competitors to construct the undirected competition network, as shown in Figure 3. First, we chose the directed competition network of firm A and set its competitors as new focal firms to obtain their directed competition networks. Then, based on the competitive relationship among all firms, we further obtained the bidirectional relationships. Finally, according to the characteristics of reciprocity, we retained the bidirectional links between the focal firm and its competitors and constructed the undirected competition network of firm A.

Figure 3 shows an example of an undirected competitive network. There was a closer competitive relationship between firm A and some of its competitors (firm B, firm C, and firm E), while there was no bidirectional relationship between firm A and firm F (firm D). Combined with the reciprocity of the competition network, there was a close network for firm A, firm B, and firm C, which is called network closure. According to the theory of network closure, firm C will engage in special reactions to the decisions of firm A or react to the actions of firm B, which may respond to firm A's actions. In this situation, network closure creates a more stable link between the focal firm and its competitors by facilitating information exchanges (Skilton & Bernardes, 2015).

Figure 3. The construction of an undirected competition network



DATA

Data Sources

Due to the requirements of the CSRC, listed firms need to introduce their operations and development goals in information disclosure. In this paper, we chose China's listed firms on the Shanghai Stock Exchange and Shenzhen Stock Exchange from 2007 to 2018 as the research sample and further constructed the text-based competition network for each firm. Because the operations and products of firms in the financial and insurance industries are different from those of nonfinancial firms, we needed to exclude financial firms from our research sample, which guaranteed that the focal firm and its competitors have similar development goals. For the text data on information disclosure, the annual reports of sample firms were obtained from the websites of the Shanghai Stock Exchange and Shenzhen Stock Exchange. Based on the "Standards for the Contents and Forms of Listed Firms' Information Disclosure" released by the CSRC in 2005, we extracted the "Management Discussion and Analyses" (MD&A) section from the annual reports of individual firms and obtained the descriptive texts. Because the MD&A section may contain some redundant information, we saved the descriptive text related to products or services in the preprocessing of text data. Therefore, there are fewer words used in the neural network language model than in the annual reports.

To evaluate the economic links between the focal firm and its competitors, financial indicator data were obtained from the China Stock Market & Accounting Research Database and Wind Database, including return on assets (ROA), asset turnover, price-to-earnings ratio, return on equity (ROE), and stock return. Based on the empirical analysis, the competitive relationship in our competition network was further explained and represented the interactions between the focal firm and its competitors.

Data Description

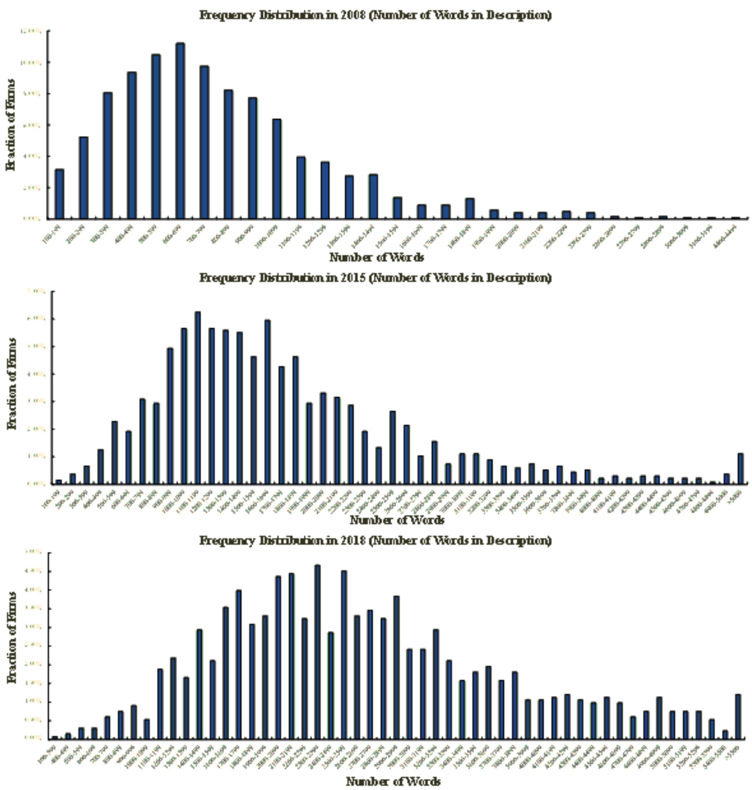
The text data contained in the information disclosures are used to capture the competitive relationship between firms and construct the text-based competition network. Since the CSRC implemented the Regulations on Information Disclosure of Listed Companies in 2007, we selected 2007 as the initial year of the sample period. In terms of the performance of China's listed firms in information disclosure, we performed a statistical analysis for the descriptive text of annual reports and obtained the frequency distribution of words in different years (2008, 2015, and 2018), as shown in Figure 4.

In Figure 4, the number of words used by most firms in 2008 was approximately 700, and it increased to 1,500 and 2,500 in 2015 and 2018, respectively. The change in words used in annual reports reveals that an increasing number of listed firms have begun to pay attention to the descriptive text contained in information disclosure, and they aim to use more nonfinancial information disclosure to introduce their operations and development goals. In this situation, whether firms are engaged in similar businesses or goals is measured through the similarity of descriptive texts in the annual reports. In other words, the more similar the descriptive texts of the information disclosures, the more competitive environment these firms will face.

Referring to Lee et al. (2015), we set the number of competitors as 10 ($k = 10$); that is, each focal firm will pay attention to 10 competitors. Compared with the traditional industry classification, we obtained the distribution of competitors in the same industry as the focal firm. In terms of the industry classification system, the standard industry classification released by the CSRC in 2012 was selected as the benchmark and an analysis from the primary sector (IC1) and tertiary sector (IC3) of the CSRC's industry classification system was performed. The results of the distribution of competitors in the directed competition network are shown in Table 1.

Table 1 shows summary statistics on the degree of correspondence between our method and standard industry classifications. According to the results of Panel A, 55% of the most similar competitors belong to the primary sector of the focal firm and 48% of those belong to its tertiary sector, indicating that most competitors in the text-based competition network belong to the same industry as the focal firm. With the decrease in similarity, an increasing number of competitors

Figure 4. Frequency distribution of words in the descriptive texts of annual reports



belong to different industries, and there are interindustry correlations between the focal firm and its competitors, revealing that our measure of competition has gradually eliminated the industry boundaries. Panel B shows that the proportion of focal firms and their competitors belonging to the same industry reached the lowest point in 2009. The main reason for this may be the global financial crisis and the unwillingness of China's listed firms to disclose more businesses and goals in their annual reports. Since 2012, many competitors in the text-based competition network have belonged to the same industry as the focal firm, indicating that the new industry classification encourages firms to disclose more information related to their operations.

In Table 1, some competitors obtained by the similarity of nonfinancial information disclosure may come from different industries, demonstrating that some economically related firms of the focal firm have broken the boundaries of the traditional industry. Considering the impact of external shocks on information disclosure, the changes in descriptive texts reflect managers' preference for businesses and goals, which also shows the characteristics of dynamic competition.

RESULTS

Characteristics of the Competition Network

Analysis of the Firm-Level Competition Network

According to the construction of the competition network proposed in The Construction of the Competition Network section, we gradually constructed the whole competition network for all sample firms, including both the directed competition and undirected competition networks. To show the

Table 1. Correspondence with standard industry classifications

Panel A: Correspondence by degree of closeness of the top 10 competitors		
Competitor rank	**Same CSRC_IC1**	**Same CSRC_IC3**
1	0.55	0.48
2	0.49	0.41
3	0.47	0.39
4	0.45	0.37
5	0.42	0.34
6	0.39	0.31
7	0.38	0.30
8	0.37	0.29
9	0.36	0.28
10	0.34	0.26
Total	0.42	0.34
Panel B: Correspondence by year for the top 10 competitors		
Year	**Same CSRC_IC1**	**Same CSRC_IC3**
2007	0.49	0.38
2008	0.40	0.31
2009	0.32	0.25
2010	0.42	0.33
2011	0.36	0.28
2012	0.48	0.41
2013	0.44	0.38
2014	0.44	0.37
2015	0.44	0.36
2016	0.41	0.33
2017	0.44	0.35
2018	0.42	0.33
Total	0.42	0.34

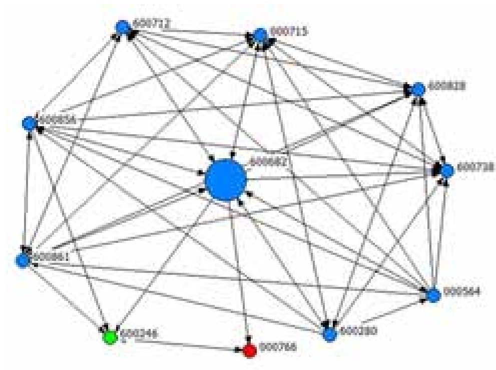
Note: CSRC is the abbreviation of the China Securities Regulatory Commission

competitive environment faced by the focal firm, an analysis of the firm-level competition network was performed for the two kinds of focal firms' ego networks.

In the directed competition networks, the top 10 similar firms of the focal firm are identified as its competitors, and they are also the nodes of the focal firm's ego network. Specifically, the directed competition network of Nanjing Xinjiekou Department Store (NJXB, 600682) in 2014 was selected as an example to show the advantages of the text-based competition network, as shown in Figure 5.

Figure 5 shows the competitors faced by NJXB (600682) in 2014. In this network, the blue node represents that the firm belongs to the retail industry. The red node indicates that the firm belongs to the pharmaceutical manufacturing industry. The green node represents that the firm belongs to the real estate industry. According to the links of competition in Figure 5, most of the competitors of NJXB belong to the retail industry, while Vantone NeoDev Group (600246) and Tonghua Golden-horse

Figure 5. A directed competition network in 2014



Group (000766) belong to the real estate and pharmaceutical manufacturing industries, respectively. It is noteworthy that these two competitors belong to different industries from NJXB. Through the annual report of NJXB in 2014, this firm disclosed that "NJXB has gradually transformed from the traditional department store to the multi sectors shopping mall; after investing in the Healthcare Solutions of Natali, NJXB will try to enter the field of medical services." Based on the descriptive text of NJXB, it is clear that this firm participates in multiple markets, including the real estate and pharmaceutical industries. Therefore, the method of identifying text-based competitors can effectively help focal firms obtain competitors in different markets, and then break the boundaries of traditional industry classifications.

Based on the directed competition network, we obtained the undirected competition network by maintaining some stable links. Moreover, two ego networks were used to show the characteristics of undirected competition network, and Shanghai Yanshi (600696) and Shandong Jintai Group (600385) are selected as examples, as shown in Figure 6.

Figure 6. The undirected competition networks in 2018

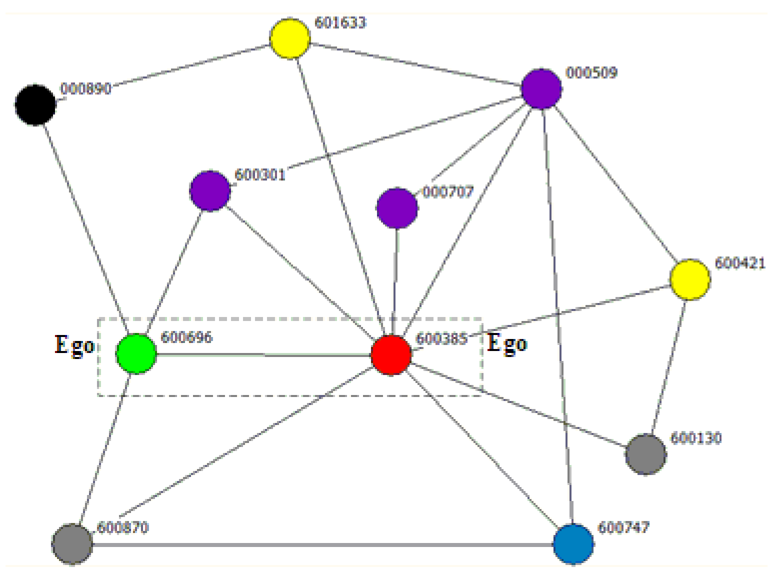


Figure 6 shows the undirected competition networks of Shanghai Yanshi (600696) and Shandong Jintai Group (600385) in 2018. According to the undirected competition networks in Figure 6, Shanghai Yanshi belongs to the real estate industry, while Shandong Jintai Group belongs to the pharmaceutical manufacturing industry. It is worth noting that there is a competitive relationship between these two firms, but they do not belong to the same industry as their competitors. By analyzing the descriptive text of Shanghai Yanshi's annual report in 2018, this firm disclosed that "our financing needs are high in different industries, including the medical, industrial, real estate and other industries; our business will focus on high-tech firms in the medical, industrial, real estate and other industries." Similarly, Shandong Jintai Group disclosed that "the income from the home rentals and internet services is low; the major changes in operating costs are mainly due to the trade in gold and jewelry during this reporting period." Based on the descriptive text of these two firms in 2018, both Shanghai Yanshi and Shandong Jintai Group participated in multiple markets, which forced them to face some competitors in different industries. Therefore, undirected competition networks lead to a stronger competitive relationship between firms, and their links also represent the participation of such firms in different markets.

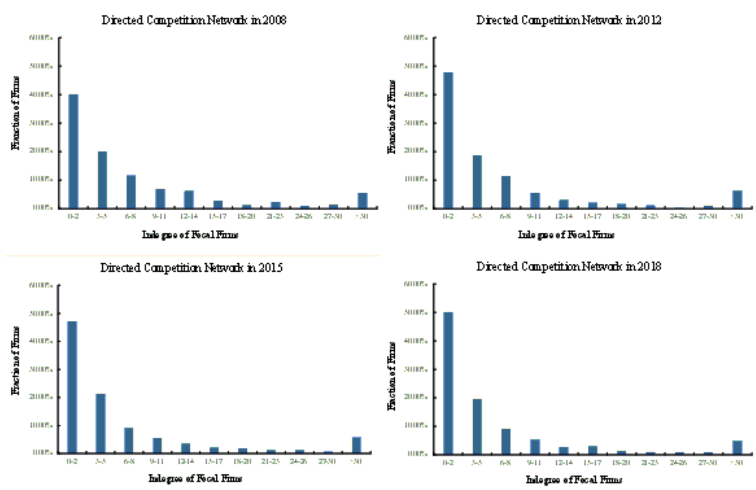
Analysis of the Whole Competition Network

Based on the method of identifying competitors proposed in this paper, we used the competition network at the firm level to construct the whole competition network year by year. The changes in the whole competition network establish the trend in the degree of competition among all firms and show the clusters of some firms. In this situation, we further analyzed the whole directed competition and whole undirected competition network.

First, the directed competition networks of focal firms allow them to face 10 competitors, which will make the outdegree of each focal firm 10. In this type of competition network, the indegree of the focal firm reflects the degree to which the focal firm is a concern of other firms, that is, the number of firms that identify this focal firm as their competitor. With the increase in indegree, this focal firm will gradually become an important node in the whole network and may even become a leader in some markets. Therefore, we focused on the indegree of each focal firm in the whole directed competition network to show the change in competition among all firms, as shown in Figure 7.

Figure 7 shows the changes in the indegree of focal firms in different periods. In 2008, the distribution of focal firms' indegree was relatively balanced. However, after 2012, fewer and

Figure 7. The indegree of focal firms in different periods

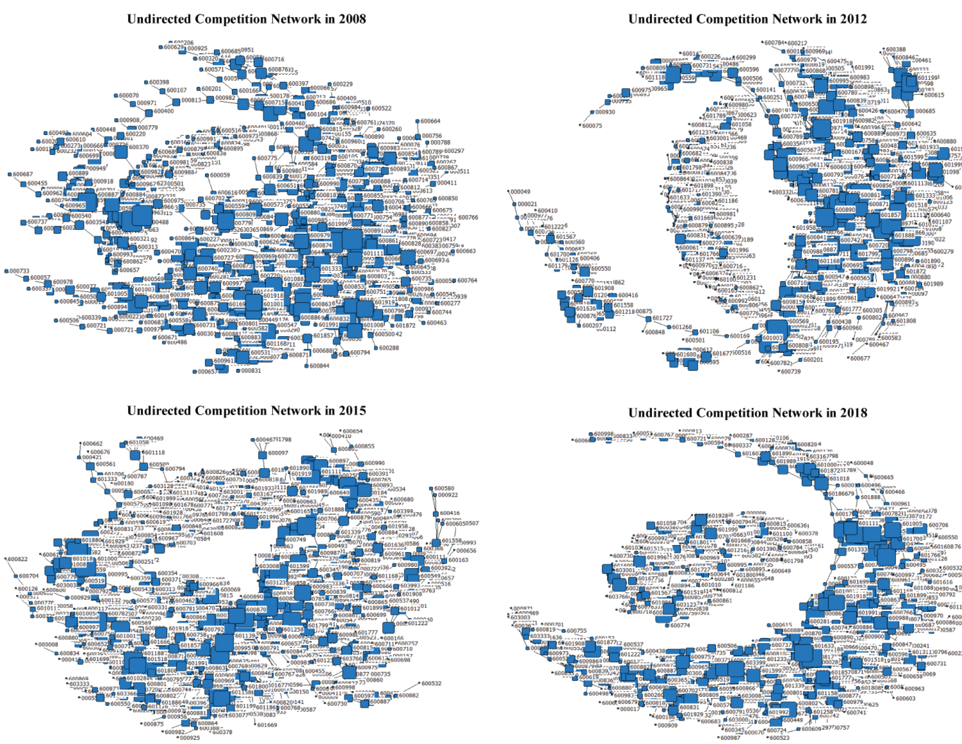


fewer firms paid attention to some focal firms, and the fraction of focal firms with indegrees in the range of 0 to 2 reached 50% in 2018. Because the total indegree and outdegree of the competition network are equal, the increase in firms attracting less attention indicates that some firms attracting more attention appears in the whole network, and these firms may become leaders in some fields.

Second, the undirected competition networks of focal firms are based on the links of directed competition networks, and the outdegree of each focal firm is not constant. These links are stable competitive relationships, reflecting the equivalence between the focal firm and its competitors. In this situation, there may be some isolated firms in the whole undirected competition network. Therefore, we focused on the changes in firms' locations in the whole undirected competition network to show the competitive environment faced by each firm, as shown in Figure 8.

Figure 8 shows the graphs of the undirected competition network in different periods, and the changes in the undirected competition network are captured in four years (2008, 2012, 2015, and 2018). The size of a node in a competition network is determined by the number of its competitors; that is, a large node will face more competitors, and a small node will face fewer competitors. In 2008, many firms owned some competitors in the undirected competition network, but there were increasingly more clusters of firms in this kind of network after 2012, and some firms with a high degree gathered in some groups. It is worth noting that some important firms with more competitors are gathered in some groups, and the whole market is obviously divided into two parts. These results demonstrate that some leading firms receive more attention from other firms, and they may control the development of the whole market.

Figure 8. Graph of the undirected competition network in different periods



The Economic Links Between the Focal Firm and Its Competitors

In the resource-based view, the competition between firms can be further transformed into a game of rare resources, which will support the long-term development of such firms (Li et al., 2013). Competitive relationships represent the economic links between firms, which also make the relationship between the focal firm and its competitors stronger (Dierynck & Verriest, 2020).

Capturing economically related firms is a conceptual idea, but it is difficult to measure firms' relative importance. Referring to the method of Lee et al. (2015), we measured the economic links between the focal firm and its competitors by using their relative importance; that is, the focal firm's competitors should exhibit greater contemporaneous correlation with this focal firm in some financial indicators. The industry classification released by the CSRC in 2012 (CSRC_IC) was selected as a benchmark identification method. Furthermore, we used the idea of TNIC to identify competitors for focal firms by product-related words (CN_TNIC), and this method was selected as another benchmark identification method. CN_TNIC is constructed by the method of term frequency–inverse document frequency (TF-IDF), which is helpful in capturing the relationship between focal firms and competitors from the dimension of product-related words. The quarterly data of China's listed firms from 2012 Q2 to 2016 Q1 were gathered to perform the tests of economic links, and the cross-sectional model is shown in Equation (6):

$$Varible_{i,q} = \alpha_i + \beta_q Variable\_Average_{Competitors_i,q} + \varepsilon_{i,q} \tag{6}$$

In Equation (6), $Varible_{i,q}$ represents the financial indicators of focal firm *i* in period *q* and $Variable\_Average_{Competitors_i,q}$ represents the average financial indicators of firm *i*'s competitors in period *q*. To capture the economic links between the focal firm and its competitors, the average R^2 from quarterly cross-sectional regressions is used to measure their relative importance in terms of financial indicators.

Based on the existing studies, some financial indicators were selected in this section, namely, risk-taking (Risk1 and Risk2), price to sales ratio (PS), price to earnings ratio (PE), profit margin (PM), and market value to tangible assets ratio (MT) (Hoberg & Phillips, 2016; Lee et al., 2015, 2019). In terms of corporate risk-taking, this variable represents the ability of firms to bear potential risks, which is an important measure of firms' profitability. We used the method proposed by John et al. (2008) to construct the variables of corporate risk-taking in three periods (T = 3), as shown in Equations (7)–(9). The variables of price to sales ratio and price to earnings ratio show the performance of individual firms in the stock market, as well as the well-being of stockholders. The profit margin variable represents the actual profitability of firms, and these firms' advantages can be shown by the level of market value to tangible assets ratio:

$$Adj\_ROA_{i,q} = \frac{EBIT_{i,q}}{ASSET_{i,q}} - \frac{1}{Num} \sum_{j \in Competitor_i}^{Num} \frac{EBIT_{j,q}}{ASSET_{j,q}} \tag{7}$$

$$Risk1_{i,q} = \sqrt{\frac{1}{T-1} \sum_{q=1}^{T} \left(Adj\_ROA_{i,q} - \frac{1}{T} \sum_{q=1}^{T} Adj\_ROA_{i,q} \right)^2} \mid T = 3 \tag{8}$$

$$Risk2_{i,q} = Max\left(Adj\_ROA_{i,q}\right) - Min\left(Adj\_ROA_{i,q}\right) \mid T = 3 \tag{9}$$

According to Equation (6), we ran the cross-sectional regression to obtain the average R^2 of 16 quarters from 2012 Q2 to 2016 Q1. The results are reported in Table 2.

Table 2 compares the performance of the text-based method and other benchmark methods in terms of economic links between the focal firm and competitors. Column (1) reports the average R^2 of text-based competitors in 16 quarters, and columns (2) and (3) report the average R^2 of CSRC_IC and CN_TNIC competitors. By comparing the results of columns (1)–(3), it can be seen that text-based competitors outperform CSRC-IC and CN-TNIC competitors for most financial indicators. Column (4) tests for the significance of the differences in average R^2 between text-based and CSRC_IC competitors. According to the results of column (4), the differences are statistically significant at the 1% and 5% levels for corporate risk-taking and at the 1% level for the price to sales ratio. For the profit margin and market value to tangible assets ratio, the differences are statistically significant at the 10% level. Column (5) tests for the significance of the differences in average R^2 between text-based and CN_TNIC competitors. According to the results of column (5), the differences are statistically significant at the 10% level for the price to sales ratio, and the price to earnings ratio is significant at the 5% level. Combined with the results of columns (5) and (6), the differences in these variables are positive, which shows that text-based identification method better captures the economic links between focal firms and competitors.

The Impact of the Competition Network on Future Performance

The text-based competition network focuses on the competitive environment faced by the focal firm, and each focal firm will face a unique set of competitors. In this situation, the external pressure brought by these competitors will directly affect the decision-making of the focal firm's manager, and

Table 2. Comparison of R^2 for text-based competitors and other competitors

	Text	CSRC_IC	CN_TNIC	The differences between (1) and (2)	The differences between (1) and (3)
	(1)	(2)	(3)	(4)	(5)
Risk1	0.0300	0.0070	0.0173	0.0230***	0.0127
	(0.0074)	(0.0021)	(0.0030)	(0.0077)	(0.0080)
Risk2	0.0305	0.0102	0.0178	0.0203**	0.0126
	(0.0070)	(0.0041)	(0.0031)	(0.0071)	(0.0076)
PS	0.0265	0.0026	0.0159	0.0239***	0.0107*
	(0.0043)	(0.0013)	(0.0039)	(0.0045)	(0.0058)
PE	0.0158	0.0100	0.0034	0.0059	0.0124**
	(0.0045)	(0.0071)	(0.0013)	(0.0084)	(0.0047)
PM	0.0053	0.0003	0.0007	0.0050*	0.0047
	(0.0028)	(0.0001)	(0.0004)	(0.0028)	(0.0029)
MT	0.0161	0.0022	0.0140	0.0140*	0.0022
	(0.0074)	(0.0012)	(0.0068)	(0.0075)	(0.0100)
Number of quarters	16	16	16	16	16

Note: CSRC_IC is the identification of competitors by the industry classification released by the CSRC. CN_TNIC is the identification of competitors by product-related words. Risk1 and Risk2 denote the variables of corporate risk-taking; PS denotes the variable of price to sales ratio; PE denotes the variable of price to earnings ratio; PM denotes the variable of profit margin; and MT denotes the variable of market value to tangible assets ratio. ***, **, * represent the significance at the 1%, 5%, and 10% levels, respectively. Standard errors are reported in parentheses.

even affect its long-term development. Therefore, we will further explore the impact of text-based competition networks on the future performance of focal firms.

Based on social network theory, we used two network variables to show the characteristics of text-based competition networks, including the intensity of competition and the density of networks. First, firms in a more competitive environment tend to make important investment decisions sooner, indicating that the intensity of competition will affect some firms' profitability (Akdoğu & MacKay, 2008). According to industrial organizations, the concentration of production reflects the strength of competitive relationships between the focal firm and its competitors, and the focal firm will pay more attention to some competitors with higher similarity (de Bodt et al., 2018). Referring to Zhang & Guan (2019), the average of the similarities between the focal firm and its competitors is used to measure the intensity of competition in the directed competition network, as shown in Equation (10):

$$Competition_{i,t} = \frac{1}{k} \sum_{j \in Competitors_{i,t}}^{k} Similarity_{i,j,t} \tag{10}$$

In Equation (10), $Competition_{i,t}$ represents the intensity of competition faced by focal firm i in year t; $Similarity_{i,j,t}$ indicates the similarity between focal firm i and its competitor j in year t; $Competitors_{i,t}$ represents the set of competitors of enterprise focal firm i in year t; and k represents the number of competitors.

Second, the structure of the competition network determines the transmission of information about firms' decisions, which also enables some firms to respond quickly to the actions of competitors. In the theory of network closure, the connections between competitors improve the speed of information transmission for focal firms' actions, and this special social capital limits the decision-making of focal firms (Zhang & Guan, 2019). The undirected competition network can better show the information interactions between firms, and the structure of this network may have an impact on the future performance of the focal firm. Referring to Skilton & Bernardes (2015), the density of the competition network is measured by the percentage of possible competitive links between firms of the ego network that are actually realized, as shown in Equation (11):

$$Density_{i,t} = \frac{Actual\,ties\,among\,competitors_{i,t}}{Max\,ties\,among\,competitors_{i,t}} \tag{11}$$

In Equation (11), $Density_{i,t}$ represents the density of firm i's competition network in year t; $Actual\ ties\ among\ competitors_{i,t}$ represents the actual number of competitive links between firm i's competitors; and $Max\ ties\ among\ competitors_{i,t}$ represents the maximum number of possible competitive links between firm i's competitors.

Considering the dynamic interactions between the focal firm and its competitors, changes in the competition network structure may have an impact on the focal firm's operations and even on its profitability (Gupta et al., 2021). In different competitive environments, the decision-making of firms may reflect their reactions to their competitors, and this process may determine the ability of the focal firm to bear potential risks (Ferris et al., 2017; Luo et al., 2016). According to Li et al. (2013), future performance is captured by some variables related to return on operating assets. Thus, corporate risk-taking represents the profitability of one firm and its ability to bear potential risks (John et al., 2008). Therefore, we used corporate risk-taking to describe the future performance of the focal firm and constructed the following empirical model:

$$RiskTaking_{i,t+1} = \alpha_1 + \beta_1 NetworkVariable_{i,t} \\ +\gamma Controls_{i,t} + Year_{fixed\ effect} + Region_{fixed\ effect} + \varepsilon_{i,t} \tag{12}$$

In Equation (12), $Risk\ Taking_{i,t+1}$ represents the risk-taking level of focal firm *i* in year *t*+1, including the variables of Risk1 and Risk2; $Network\ Variable_{i,t}$ represents the characteristics of the competition network structure of focal firm *i* in year *t*, including the intensity of competition and the density of the network; and $Controls_{i,t}$ represents the control variables, including the leverage, stock returns, the proportion of independent directors and institutional shareholding, the growth of operating revenue, and net profit.

In a further test, the square terms of the competition network structure are added to the basic empirical model in Equation (12), and then the nonlinear association between the characteristics of the network structure and the future performance of the focal firm is tested in Equation (13):

$$RiskTaking_{i,t+1} = \alpha_1 + \beta_1 NetworkVariable_{i,t} + \beta_2 NetworkVariable_{i,t}^2 \\ +\gamma Controls_{i,t} + Year_{fixed\ effect} + Region_{fixed\ effect} + \varepsilon_{i,t} \tag{13}$$

Before the empirical analysis, we performed a descriptive statistical analysis of all variables and reported the results in Table 3. The standard deviation of the risk-taking variables is smaller than that of most of the other variables, indicating that the future performance of the focal firm will not change greatly. In addition, the mean value of the competition intensity is close to its maximum value, which suggests that many firms are in a competitive environment. Moreover, the maximum value of network density is 5 times its mean value, demonstrating that some firms are in a dense competition network.

According to the empirical models based on Equations (12) and (13), we first performed a regression analysis on the linear and nonlinear association between competition intensity and the focal firm's future performance. The regression results are reported in Table 4.

In Table 4, columns (1)–(3) explore the impact of competition intensity on the variable of Risk1. From the results of columns (1)–(2), it can be seen that there is a positive association between competition intensity and corporate risk-taking, significant at the 1% level. According to the results of column (3), there is a significant U-shaped relationship between competition intensity and corporate

Table 3. Descriptive statistics of the variables

Variable	N	Mean	Standard Deviation	Minimum	Maximum
Risk1	15,672	0.0321	0.0407	0.0000	0.2364
Risk2	15,672	0.0608	0.0767	0.0000	0.4479
Competition	15,672	0.4956	0.0873	0.2354	0.6844
Density	15,672	0.1723	0.2865	0.0000	1.0000
Leverage	15,672	0.5281	0.2158	0.0820	1.2235
Return	15,672	0.2729	0.8429	–0.7414	3.6971
Independ	15,672	0.3699	0.0528	0.3000	0.5714
Institution	15,672	0.0658	0.0805	0.0000	0.3813
Sales	15,672	0.2399	0.8859	–0.7323	6.9373
Profitability	15,672	–0.2161	5.3727	–32.5902	21.0563

Table 4. The impact of competition intensity on the focal firm's future performance

	Dependent Variable: Risk1			Dependent Variable: Risk2		
	(1)	(2)	(3)	(4)	(5)	(6)
Competition	0.0179***	0.0226***	–0.2795***	0.0327***	0.0415***	–0.5295***
	(3.02)	(3.91)	(–7.66)	(2.91)	(3.80)	(–7.69)
Competition^2			0.3058***			0.5779***
			(8.36)			(8.38)
Leverage		0.0142***	0.0137***		0.0269***	0.0260***
		(6.64)	(6.43)		(6.66)	(6.44)
Return		0.0042***	0.0042***		0.0081***	0.0080***
		(6.03)	(6.01)		(6.09)	(6.07)
Independ		0.0040	0.0045		0.0068	0.0079
		(0.66)	(0.75)		(0.60)	(0.69)
Institution		–0.0519***	–0.0480***		–0.0979***	–0.0905***
		(–15.29)	(–14.16)		(–15.30)	(–14.18)
Sales		0.0013**	0.0013**		0.0024**	0.0024**
		(2.37)	(2.34)		(2.35)	(2.32)
Profitability		–0.0009***	–0.0009***		–0.0016***	–0.0016***
		(–9.32)	(–9.27)		(–9.23)	(–9.18)
Constant	0.0311***	0.0160***	0.0805***	0.0594***	0.0308***	0.1527***
	(11.70)	(4.26)	(9.14)	(11.82)	(4.37)	(9.20)
Year	Yes	Yes	Yes	Yes	Yes	Yes
Region	Yes	Yes	Yes	Yes	Yes	Yes
p value	.00	.00	.00	.00	.00	.00
R^2	0.0283	0.0630	0.0687	0.0288	0.0634	0.0690
N	14,172	14,172	14,172	14,172	14,172	14,172

Note: Competition^2 is the square terms of the competition network structure. ***, **, * represent the significance at the 1%, 5%, and 10% levels, respectively. The *t* statistics are reported in parentheses.

risk-taking, indicating that the impact of competition intensity on the focal firm's future performance will change with the degree of competition. Columns (4)–(6) explore the impact of competition intensity on the variable of Risk2. The results of columns (4)–(5) show that there is also a significant positive association between competition intensity and corporate risk-taking. In column (6), there is a significant U-shaped relationship between competition intensity and corporate risk-taking, which supports the result of column (3). Based on the results in Table 4, competition intensity improves the future performance of the focal firm, and this is strengthened by an increasing competitive environment.

We then used the empirical model constructed by Equations (12) and (13) to discuss the linear and nonlinear associations between network density and the focal firm's future performance. The empirical analysis of network density is consistent with that of competition intensity. The regression results are reported in Table 5.

In Table 5, columns (1)–(3) explore the impact of network density on the variable of Risk1. From the results of columns (1)–(2), it can be seen that there is a negative association between network density and corporate risk-taking, significant at the 10% level. According to the results in column

Table 5. The impact of network density on the focal firm's future performance

	Dependent Variable: Risk1			Dependent Variable: Risk2		
	(1)	(2)	(3)	(4)	(5)	(6)
Density	–0.0023*	–0.0022*	–0.0056	–0.0044*	–0.0043*	–0.0108
	(–1.89)	(–1.88)	(–1.57)	(–1.92)	(–1.91)	(–1.62)
Density^2			0.0041			0.0079
			(1.00)			(1.04)
Leverage		0.0139***	0.0139***		0.0263***	0.0263***
		(6.48)	(6.48)		(6.50)	(6.50)
Return		0.0042***	0.0042***		0.0081***	0.0081***
		(6.05)	(6.05)		(6.11)	(6.11)
Independ		0.0037	0.0036		0.0063	0.0060
		(0.61)	(0.59)		(0.55)	(0.53)
Institution		–0.0515***	–0.0514***		–0.0971***	–0.0969***
		(–15.21)	(–15.18)		(–15.23)	(–15.20)
Sales		0.0013**	0.0013**		0.0025**	0.0025**
		(2.38)	(2.37)		(2.35)	(2.35)
Profitability		–0.0009***	–0.0009***		–0.0016***	–0.0016***
		(–9.33)	(–9.33)		(–9.24)	(–9.24)
Constant	0.0372***	0.0237***	0.0240***	0.0705***	0.0451***	0.0456***
	(22.01)	(7.55)	(7.61)	(22.11)	(7.62)	(7.69)
Year	Yes	Yes	Yes	Yes	Yes	Yes
Region	Yes	Yes	Yes	Yes	Yes	Yes
p value	.00	.00	.00	.00	.00	.00
R^2	0.0279	0.0622	0.0623	0.0284	0.0626	0.0627
N	14,172	14,172	14,172	14,172	14,172	14,172

Note: Density^2 is the square terms of the competition network structure. ***, **, * represent the significance at the 1%, 5%, and 10% levels, respectively. The *t* statistics are reported in parentheses.

(3), the nonlinear association between network density and corporate risk-taking is not significant. Columns (4)–(6) explore the impact of network density on the variable of Risk2. The results of columns (4)–(5) show that there is a significant negative association between network density and corporate risk-taking. In column (6), there is no significant nonlinear association between network density and corporate risk-taking. Based on the results in Table 5, the density of the competition network may inhibit the ability of firms to bear potential risks, suggesting that a denser competition network limits the decision-making process of the focal firm.

Combined with the results in Tables 4 and 5, the structural characteristics of the competition network significantly affect the future performance of the focal firm. Specifically, a more competitive environment improves the ability of the focal firm to bear risks, but this may also weaken the manager's risk preference for the focal firm. It is worth noting that when the degree of competition reaches a specific level, the promotion of competition intensity on the future performance of the focal firm is strengthened, which further supports that increasing the concentration of production reduces the potential risks faced by such firms (Engelberg et al., 2018).

Number of Text-Based Competitors

For any firm, it is difficult for managers to pay attention to the operations and development goals of all competitors, which forces managers to pay more attention to those competitors with a high level of economic similarity (Luo et al., 2016). As the external pressure generated by competitors affects the decision-making of the focal firm, the number of competitors will have a potential impact on the economic links between such firms (Lewellen & Metrick, 2010).

Based on the economic relevance between the focal firm and its competitors, we used stock return comovement to discuss their relative importance (Lee et al., 2015). Because our method was mainly based on the annual reports of listed firms, discussing the number of competitors at the monthly level improves the comovement of the focal firm and its competitors in the stock market. Motivated by this idea, we discussed economic relevance by using the explanation of the average monthly returns of competitors for the monthly return of the focal firm and analyzed the different numbers of competitors. The empirical model of stock return comovement is as follows:

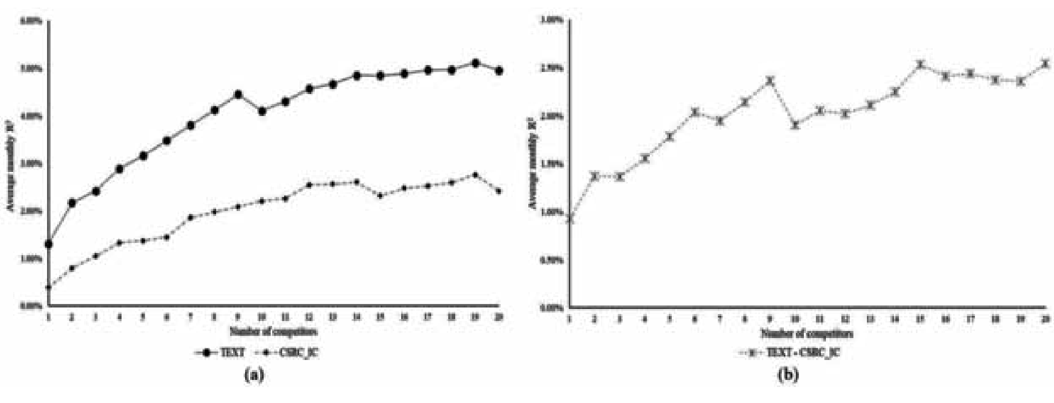
$$Return_{i,m} = \alpha_i + \beta_m Return\_Average_{Competitors_i,m} + \varepsilon_{i,m} \tag{14}$$

In Equation (14), $Return_{i,m}$ represents the stock return of focal firm i in month m and $Return\_Average_{Competitors_i,m}$ represents the average stock returns of focal firm i's competitors in month m. According to the identification of competitors proposed in *The Identification of Competitors section*, we varied the value of k from 1 to 20. In terms of industry competitors, we randomly selected k firms from the same industry of focal firm i.

Considering the new industry classification released by the CSRC in 2012, we chose the sample period from May 2013 to April 2016 (a total of 36 months) and estimated the cross-sectional regression based on Equation (14) (Lee et al., 2015). Moreover, we utilized the R^2 measure to show the changes in economic relevance between the focal firm and its competitors with different k. The results of cross-sectional regression are shown in Figure 9.

Figure 9 (a) shows the changes in the average monthly R^2 of competitors with different numbers of competitors. In these two methods, the stock price comovement between the focal firm and its competitors will increase with the growth of the number of competitors. The average monthly R^2 obtained by text-based competitors (TEXT) is significantly higher than that obtained by industry competitors (CSRC_IC), supporting that our method better captures the economic links between the focal firm and its competitors. When the number of competitors reaches 10, the average monthly R^2

Figure 9. The average monthly R^2 and the differences in R^2 for different numbers of competitors



begins to show a stable state, indicating that it is reasonable to set k to 10; that is, each focal firm will pay more attention to 10 competitors. Figure 9 (b) shows the differences in average R^2 between text-based and industry competitors with different k. According to the results in Figure 9 (b), our method is superior to the industry classification under the different number of competitors, and the difference shows a steady state under the condition of 10 competitors. The results of Figure 9 demonstrate that the text-based competition network helps each firm identify a group of economically related firms, and the economic links obtained by our method are stronger than those obtained by industry classification.

CONCLUSION

Competition is the key factor in determining whether individual firms can achieve the goal of long-term development and has an impact on the survival of most firms. From resource-based views, the core of competition lies in obtaining unique and rare limited resources, which will also determine firms' positions in different markets. Under the differentiation strategy, an increasing number of firms begin to enter multiproduct markets, which forces firms to face competitors from different fields or industries. Although the competitors identified by industry classification systems are focused on their main economic activities, industry relationships ignore the competition between firms in multiple markets. To solve this dilemma, identifying the appropriate competitors and constructing the competition network for individual firms is an important step in exploring long-term development strategies, which also describe the dynamic change in different markets.

Based on the theory of competition, we explored the characteristics of the competition between firms and analyzed the impact of the competition network on the operations of firms from a micro perspective. Our method uses descriptive texts in information disclosures to show the similarity of the operations and goals of firms and measures their relative importance. To better show the semantic similarity of nonfinancial information disclosure, we adopted a neural network language model to capture the overall meaning of descriptive texts and identify the similarities between firms. Considering the actual competition between the focal firm and its competitors, some relatively important firms are identified as its competitors, and the text-based competition network is constructed by this competition relationship.

This kind of competition network shows the competitive environment faced by the focal firm at a micro level, and the weight of links reflects their relative importance. From some financial indicators, the text-based competitors identified by our method better explains the financial performance of the focal firm, which demonstrates that the text-based competition network captures the economic links between the focal firm and its competitors. In addition, the structure of the competition network also has a significant impact on the future performance of the focal firm, which suggests that the competition mechanism of the network is a special kind of social capital that influences firms' decision-making process. Considering the number of competitors, we further showed that managers are only able to pay constant attention to some competitors, which supports the effectiveness and practicality of the text-based competition network proposed in this paper.

Compared with the traditional identification of competitors, our method is more focused on the competitive environment at a micro level and identifies a time-varying and nontransitive competitive relationship between firms. The text-based competition network helps managers identify some competitors they truly care about and promote the sustainable development of firms from the perspective of social interactions.

AUTHOR NOTES

The authors of this publication declare there is no conflict of interest.

This research was supported by the National Natural Science Foundation of China (Grant number 72172020), the Fundamental Research Funds for the Central Universities (Project No. 2022CDJKYJH047).

REFERENCES

Akdoğu, E., & MacKay, P. (2008). Investment and competition. *Journal of Financial and Quantitative Analysis*, *43*(2), 299–330. doi:10.1017/S0022109000003537

Andrevski, G., Brass, D. J., & Ferrier, W. J. (2016). Alliance portfolio configurations and competitive action frequency. *Journal of Management*, *42*(4), 811–837. doi:10.1177/0149206313498901

Barney, J. B. (1986). Types of competition and the theory of strategy: Toward an integrative framework. *Academy of Management Review*, *11*(4), 791. doi:10.2307/258397

Bengio, Y., Ducharme, R., Vincent, P., & Jauvin, C. (2003). A neural probabilistic language model. *Journal of Machine Learning Research*, *3*(6), 1137–1155. doi:10.1162/153244303322533223

Cao, J., Liang, H., & Zhan, X. (2019). Peer effects of corporate social responsibility. *Management Science*, mnsc.2018.3100. 10.1287/mnsc.2018.3100

Chen, S. S., Huang, C. W., & Lin, C. Y. (2017). Observational learning from acquisition outcomes of innovation leaders. *Financial Management*, *46*(2), 487–522. doi:10.1111/fima.12132

de Bodt, E., Eckbo, B. E., & Roll, R. W. (2018). Corporate rivalry and return comovement. SSRN *Electronic Journal*. 10.2139/ssrn.3218544

Dierynck, B., & Verriest, A. (2020). Financial reporting quality and peer group selection. *Management Accounting Research*, *47*, 100675. doi:10.1016/j.mar.2019.100675

Eklinder-Frick, J., Eriksson, L. T., & Hallén, L. (2011). Bridging and bonding forms of social capital in a regional strategic network. *Industrial Marketing Management*, *40*(6), 994–1003. doi:10.1016/j.indmarman.2011.06.040

Engelberg, J., Ozoguz, A., & Wang, S. (2018). Know thy neighbor: Industry clusters, information spillovers, and market efficiency. *Journal of Financial and Quantitative Analysis*, *53*(5), 1937–1961. doi:10.1017/S0022109018000261

Ferris, S. P., Javakhadze, D., & Rajkovic, T. (2017). CEO social capital, risk-taking and corporate policies. *Journal of Corporate Finance*, *47*, 46–71. doi:10.1016/j.jcorpfin.2017.09.003

Galvin, P., Burton, N., Singh, P. J., Sarpong, D., Bach, N., & Teo, S. (2020). Network rivalry, competition and innovation. *Technological Forecasting and Social Change*, *161*, 120253. Advance online publication. doi:10.1016/j.techfore.2020.120253

Gupta, A., Owusu, A., & Zou, L. (2021). Identifying board of director network influence for firm characteristics. *Physica A*, *581*, 126212. doi:10.1016/j.physa.2021.126212

Harris, Z. S. (1954). Distributional structure. *Word*, *10*(2–3), 146–162. doi:10.1080/00437956.1954.11659520

Hoberg, G., & Phillips, G. (2016). Text-based network industries and endogenous product differentiation. *Journal of Political Economy*, *124*(5), 1423–1465. doi:10.1086/688176

Hrazdil, K., Trottier, K., & Zhang, R. (2014). An intra- and inter-industry evaluation of three classification schemes common in capital market research. *Applied Economics*, *46*(17), 2021–2033. doi:10.1080/00036846.2014.892200

John, K., Litov, L., & Yeung, B. (2008). Corporate governance and risk-taking. *The Journal of Finance*, *63*(4), 1679–1728. doi:10.1111/j.1540-6261.2008.01372.x

Kang, M., Hovav, A., Lee, E. T., Um, S., & Kim, H. (2022). Development of methods for identifying an appropriate benchmarking peer to establish information security policy. *Expert Systems with Applications*, *201*, 117028. doi:10.1016/j.eswa.2022.117028

Katselas, D., Sidhu, B. K., & Yu, C. (2019). Know your industry: The implications of using static GICS classifications in financial research. *Accounting and Finance*, *59*(2), 1131–1162. doi:10.1111/acfi.12285

Kaustia, M., & Rantala, V. (2015). Social learning and corporate peer effects. *Journal of Financial Economics*, *117*(3), 653–669. doi:10.1016/j.jfineco.2015.06.006

Kedia, S., Koh, K., & Rajgopal, S. (2015). Evidence on contagion in earnings management. *The Accounting Review*, *90*(6), 2337–2373. doi:10.2308/accr-51062

Le, Q. V., & Mikolov, T. (2014). Distributed representations of sentences and documents. *31st International Conference on Machine Learning, ICML 2014*, (*vol. 4*, pp. 2931–2939). https://arxiv.org/abs/1405.4053

Lee, C. M. C., Ma, P., & Wang, C. C. Y. (2015). Search-based peer firms: Aggregating investor perceptions through internet co-searches. *Journal of Financial Economics*, *116*(2), 410–431. doi:10.1016/j.jfineco.2015.02.003

Lee, C. M. C., Sun, S. T., Wang, R., & Zhang, R. (2019). Technological links and predictable returns. *Journal of Financial Economics*, *132*(3), 76–96. doi:10.1016/j.jfineco.2018.11.008

Lewellen, S., & Metrick, A. (2010). Corporate governance and equity prices: Are results robust to industry adjustments? *Working Paper*, 1–63.

Li, F., Lundholm, R., & Minnis, M. (2013). A measure of competition based on 10-K filings. *Journal of Accounting Research*, *51*(2), 399–436. doi:10.1111/j.1475-679X.2012.00472.x

Lin, Y., Mao, Y., & Wang, Z. (2018). Institutional ownership, peer pressure, and voluntary disclosures. *The Accounting Review*, *93*(4), 283–308. doi:10.2308/accr-51945

Loughran, T., & Mcdonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance*, *66*(1), 35–65. doi:10.1111/j.1540-6261.2010.01625.x

Loughran, T., & Mcdonald, B. (2016). Textual analysis in accounting and finance: A survey. *Journal of Accounting Research*, *54*(4), 1187–1230. doi:10.1111/1475-679X.12123

Luo, P., Chen, K., & Wu, C. (2016). Measuring social influence for firm-level financial performance. *Electronic Commerce Research and Applications*, *20*, 15–29. doi:10.1016/j.elerap.2016.09.004

Medlin, C. J., & Ellegaard, C. (2015). Conceptualizing competition and rivalry in a networking business market. *Industrial Marketing Management*, *51*, 131–140. doi:10.1016/j.indmarman.2015.05.009

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *1st International Conference on Learning Representations, ICLR 2013 - Workshop Track Proceedings*. https://arxiv.org/abs/1301.3781

Phillips, G., & Mackay, P. (2005). How does industry affect firm financial structure? *Review of Financial Studies*, *18*(4), 1433–1466. doi:10.1093/rfs/hhi032

Porter, M. E. (1989). How competitive forces shape strategy. *Readings in Strategic Management*, 133–143. 10.1007/978-1-349-20317-8_10

Salop, S. C. (1979). Monopolistic competition with outside goods. *The Bell Journal of Economics*, *10*(1), 141. doi:10.2307/3003323

Shi, G., Sun, J., & Zhang, L. (2018). Product market competition and earnings management: A firm-level analysis. *Journal of Business Finance & Accounting*, *45*(5–6), 604–624. doi:10.1111/jbfa.12300

Skilton, P. F., & Bernardes, E. (2015). Competition network structure and product market entry. *Strategic Management Journal*, *36*(11), 1688–1696. doi:10.1002/smj.2318

Wincent, J., Anokhin, S., Örtqvist, D., & Autio, E. (2010). Quality meets structure: Generalized reciprocity and firm-level advantage in strategic networks. *Journal of Management Studies*, *47*(4), 597–624. doi:10.1111/j.1467-6486.2009.00882.x

Zhang, J. J., & Guan, J. (2019). The impact of competition strength and density on performance: The technological competition networks in the wind energy industry. *Industrial Marketing Management*, *82*, 213–225. doi:10.1016/j.indmarman.2019.01.011

Wei Wang received the Ph.D. degree from Nanjing University in China. He is currently a Post-Doctorate in the School of Economics and Business Administration of Chongqing University. His research interests include natural language processing, data mining, and corporate governance.

Fengzhang Chen is a doctoral student at the College of Computer Science of Chongqing University in China. His main research direction is artificial intelligence.

Zewei Long is a doctoral student at the School of Economics and Business Administration of Chongqing University in China. His main research direction is deep learning.

Fengwen Chen received the Ph.D. degree from Chongqing University in China. He is currently a Professor in the School of Economics and Business Administration of Chongqing University. His research interests include social network, corporate governance, and business management research.

Fu-Sheng Tsai received the Ph.D. degree from I-Shou University. He is currently a distinguished adjunct professor in North China University of Water Resources and Electric Power, and a Professor in the Department of Business Administration of Cheng Shiu University. He serves editorial positions in several SSCI/SCI journals. His research interests include knowledge management, network analysis, and strategy management.