Expert Systems With Applications 111 (2018) 107–119

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Making sense of organization dynamics using text analysis

Jiexun Li[a], Zhaohui Wu[b], Bin Zhu[b], Kaiquan Xu[c,*]

[a] College of Business & Economics, Western Washington University, Bellingham, WA 98225, USA
[b] College of Business, Oregon State University, Corvallis, OR 97331, USA
[c] School of Business, Nanjing University, Nanjing, 210093, China

ARTICLE INFO

Article history:
Received 1 May 2017
Revised 5 October 2017
Accepted 5 November 2017
Available online 6 November 2017

Keywords:
Supply networks
Organizational sense-making
Text mining
Social network analysis
Discourse analysis
Sentiment analysis

ABSTRACT

Being able to understand the implicit power structures and dynamics among members plays a crucial role for the management of the organization. This paper introduces a novel and comprehensive approach to analyzing organizational discourse data. The proposed approach provides a holistic view of the power structure implied by the communications. The paper contributes to the domain of text mining by integrating various text-mining techniques to demonstrating different aspects of a power structure within an organization. It also contributes to the domain of supply chain management by using the conventional communication discourse method as the guideline for the development of the tool. We applied the proposed approach to a seven-year collection of meeting minutes from a co-op and our findings were largely confirmed by members of the organization. We provide a roadmap of using the multi-aspect approach to analyzing organizational discourse data in supply networks.

1. Introduction

A supply network is a system composed of complex relationships among firms that belong to the same supply chain. Being able to understand the implicit power structures among the firms with the same supply chain network plays a crucial role for the management of the supply network. And such knowledge could also help the network members to better position themselves within the network. Organizational discourse is a commonly used method to uncover the social and organizational structures embedded in the communications among individuals/firms within a network (Ellis, 1999). For example, some people/organization could be more influential than others because of their expertise, reputation, social connection, etc. (Allaho & Lee, 2013; Bird, Gourley, Devanbu, Gertz, & Swaminathan, 2006; Diesner & Carley, 2005). There are also sub-groups existed within a network system and the members of a subgroup support each other's agenda. Organizational discourse methods usually uncover aforementioned power structures through analyzing people's communications within a network system.

However, when it comes to identifying power structure existed in supply network the conventional way of conducting organizational discourse could only reveal the structure power at the collective level. Because of the difficulty in collecting all communications among firms in the same supply chain network, the interactions among those firms were usually reflected by productions in micro-practices of communication, which prevent the identification of power structures. On the other hand, the occurrence of Internet and remote conference technologies make it possible to document the interactions among firms. Nevertheless, the conventional manual approach of the organizational discourse limits its scope of investigating the archives of digital communications.

At the same time, text mining has become a mature method to make sense of text-based information. Methods such as natural language processing, entity extraction, sentimental analysis, and network construction has been widely used in marketing (Netzer, Feldman, Goldenberg, & Fresko, 2012), social media (Chen, Chiang, & Storey, 2012), etc. However, we have not yet found the application of text mining method in finding power structures existed in a supply chain network.

In this study, we combined organizational discourse analysis methods with text mining technologies to introduce a novel approach to identifying hidden organizational structures. We analyzed a large volume of longitudinal records of meeting minutes collected from an agricultural production and marketing cooperative (co-op). The overall design of our proposed approach consists two stages involving multiple methods. First, we extracted information about people, sentiments and topics from the texts of years of meeting minutes of the co-op and then mapped them in networks for analysis. Second, the patterns discovered from network

* Corresponding author.
E-mail addresses: Jiexun.Li@wwu.edu (J. Li), Zhaohui.Wu@bus.oregonstate.edu (Z. Wu), Bin.Zhu@bus.oregonstate.edu (B. Zhu), xukaiquan@nju.edu.cn (K. Xu).

https://doi.org/10.1016/j.eswa.2017.11.009
0957-4174/

analysis was assessed by a sociologist who has a deep knowledge of the co-op and by two key members of the co-op, which allows us to establish validity and the boundary condition of the methods in interpreting organizational dynamics. The assessment largely corroborates the findings uncovered from our text analysis.

The contribution of this paper is twofold. On one hand, this is the first study we know that applies text-mining techniques to identifying the power structures existed within the same supply chain network under the guidance of conventional organizational discourse principles. It contributes to the field of text mining by suggesting that an application specific text-mining solution could be developed by integrating with well-established methodologies used by academic researchers in other domains. On the other hand, the paper contributes to the domain of supply chain management by suggesting several important patterns and "flash points" that provide important information the network such as intensity of debate, general organizational structure of the network political and critical turning points.

The rest of the paper is organized as follows. Section 2 provides the review of related literature in supply chain network, organizational discourse principles, and related text mining methods. Section 3 introduces a 2-stage design of our methodology and describes how we apply it to analyze data collected from a co-op organization. Section 4 presents our results and evaluation. Finally, we conclude our paper in Section 5 with a roadmap of using the proposed approach in organizations.

2. Literature review

Our work roots from several fields in the literature of text and network analytics. In this section, we will review related work and their limitations that we aim to address in our study.

2.1. Knowledge management in organizations

In the literature of knowledge management, researchers have developed organizational memory information systems (OMIS) to record and maintain memories from the past, which may be reusable in the future for increasing organizational effectiveness (Dow, Hackbarth, & Wong, 2013). Davenport and Prusak (2001) define knowledge as "a fluid mix of framed experience, values, contextual information, and expert insight that provides a framework for evaluating and incorporating new experiences and information." In practice, organizational memories include experiences and skills about projects, products and decisions, as well as directory knowledge of who should know what (Nevo & Wand, 2005). Memories reside in a variety of retainers, including individual, culture, rules, procedures, roles, and so on. Memory is not just an artifact with a static state. Instead, it is often dynamic and embedded in many processes (Ackerman & Hadverson, 2000). Rather than stored as explicit knowledge in formal repositories of manuals and reports, memories are often embedded in employees' minds or in organizational routines, norms and culture as tacit knowledge (Atwood, 2002). Studies show that managers only get one-third of their knowledge from formal documents and the other two-third are from meetings or conversations (Davenport & Prusak, 2001).

In organizations, preserving and maintaining such explicit and tacit knowledge is a challenging task due to the process and informal perspective of the memories. Developing an OMIS often requires manual codification of knowledge for storage in a knowledge base. A number of OMIS initiatives fail due to the reluctance of employees to share knowledge through such systems (Kankanhalli, Tan, & Wei, 2005). The process perspective of knowledge about how to do something is often difficult to codify and represent. Many researchers have also raised concerns with the loss of contextual information in efforts of preserving organizational memories. Contextual information can include many aspects, including time, relevant projects or events, members involved and their relationships, etc. Capturing contextual information is crucial because knowledge is only useful in a specific context. Once that context is lost, the knowledge loses its value as well. Furthermore, OMIS often has search and retrieval functions that find relevant knowledge for review and reuse. How to present knowledge with its contextual information to users in a meaningful and understandable format remains a challenge. Failing to addresses these issues often leads to low usage of electronic knowledge repositories and even failures (Kankanhalli et al., 2005).

To better acquire and retain tacit knowledge from both formal and informal bodies of organizational records, we need more effective and efficient approaches. Hence, we looked into data-mining techniques that can mitigate human codification effort and uncover hidden structural patterns in organizations. One related technique is process mining. As a technology of achieving business process intelligence, process mining is aimed at automatically extracting information from event logs of an information system to discover models describing processes, organizations, and performance (Van der Aalst, 2012; Van der Aalst & Weijters., 2004). With process mining techniques, we can discover not only business process models but also the organizational aspects of processes, such as the social network of workers, interaction patterns, and network of work transfers. For the complex structure of enterprise data systems, mining process models from log data is a non-trivial task. Li, Wang, and Bai (2015) developed an intelligent approach to extract process models by mining text in policy documents. Nevertheless, this study only focused on mining elements and relationships from formal documents to build business processes. We have not seen any work that attempts to extract tacit social structures in organizations by mining text documents.

2.2. The identification of power structure

Researchers in communication literature have used discourse analysis to explore implicit organizational structure (Corman, Kuhn, Mcphee, & Dooley, 2002; Putnam & Fairhurst, 2001). Ellis (1999) calls discourse the "empirics of social organization and structure" and argues that we can only understand collective level constructions by studying their production in micro-practices of communication. Discourse and communication among all members of an organization overtime can help us understand organization-wide processes and detect pattern of change and relationships.

Centering resonance analysis (CRA) is a network-based approach to analyzing large quantities of written text and transcribed conversation (Corman et al., 2002). This approach is aimed at identifying the textual "centers," i.e., words that contributing the most to the topics of the text, by analyzing the structural measures in a word network. A comparison between two text collections in terms of their respective centers can reveal the "resonance" or similarity.

In addition to revealing structural relationships between words and topics, analyzing discourse of communication can also help to uncover structural characteristics of social networks. There are many studies exploring the communications between open-source developers, and investigating the structures of the open-source community. For example, Xu, Christley, and Madey (2006) uses the SourceForge data to construct the social network between developers, and finds the small world phenomenon, and the topological characteristics' influences on the information diffusion and collaborations. Kankanhalli et al. (2005) adopts the datasets from Github.com and Ohloh.net and finds power-law degree distributions and exhibit small-world characteristics. Their study also shows high degree nodes tend to connect more with low degree nodes, suggesting collaborations between experts and newbie developers. In addition, Antwerp and Madey (2010) and Bird, Pattison, D'Souza, Filkov, and Devanbu (2008) also find the power structures of open-source communities. There also exist some work utilizing email communication to explore the structures in organization. For example, Diesner and Carley (2005) investigates the Enron email dataset from social network perspective, and identifies key players across time. They find that during the Enron crisis the network had been denser, more centralized and more connected than during normal times, and the top executives had formed a tight clique with mutual support and highly. Bird et al. (2006) uses the email archives, and finds that a few members account for the bulk of the messages sent, and the bulk of the replies. The in-degree and the out-degree distribution of the social network exhibit typical long-tailed, small-world characteristics.

Yet, these studies take a piecemeal approach. We argue that to understand what is going on in an organization or a social network, one would need to understand the overall network structure, the key players in the network and their network positions, key issues and how the issues are perceived by members. Information emerged from these varied approaches provides different pieces of information. They can be triangulated to offer a realistic depiction of what is going on in the organization or network. Although CRA can extract main topics from a corpus, these identified topics, mostly nouns or noun phrases, do not carry subjective sentiments people express. Hence, we believe that sentiment analysis is necessary to provide an additional angle of looking into the underlying power structure of an organization.

2.3. Sentiment analysis in social networks

Human decisions and behaviors are always driven by people's opinions. People express their opinions and sentiments about a topic or an event via oral or written communications. In recent years, especially with the fast growth of user-generated contents on social media, sentiment analysis has received tremendous attention in both industry and academia for mining and tracking user opinions. This need has instigated a growing number of research on sentiment analysis or opinion mining (Bo Pang & Lee., 2008; Liu, 2006; Raghu & Chen, 2007). Sentiment analysis approaches try to analyze text and determine its sentiment polarity. It can be conducted at the document to determine the overall sentiment of a document (e.g., a review) (Abbasi, Chen, & Salem, 2008; Pang, Lee, & Vaithyanathan, 2002). When a document consists of sentences expressing a mixture of opinions, a sentiment-level analysis of sentiment is more appropriate. Furthermore, sentiment analysis can be carried out at the attribute level to extract opinions toward specific attributes or aspects of a topic or product) (Liu, 2006). Sentiment analysis techniques can be classified into two types: unsupervised approaches and supervised approaches. The unsupervised approaches require a lexicon of sentiment words to determine the polarity of text by counting the positive and negative words (Hatzivassiloglou & Wiebe, 2003). The supervised approaches, on the other hand, train a classifier using labeled data to predict polarity of unlabeled text (Pang et al., 2002; Wiebe, Bruce, & O'Hara, 2002; Yu & Hatzivassiloglou, 2003). Most existing sentiment analysis work focuses on analyzing customer reviews or social media content. Often times these approaches deal with one person's opinion/sentiment on one topic at a time. By contrast, in our study, we analyze a collection of meeting minutes that archive oral discussions and communications between multiple people on a variety of topics. These increase the complexity of sentiment analysis.

In social network analysis (SNA), researchers have developed a variety of approaches to uncovering how people influence and connect with each other in a power structure. To identify nodes with great influential power, we often look at different centrality measures. Specifically, degree centrality is defined as the number of links incident upon a node. A person with a high degree centrality tends to influence (or be influenced by) a large number of other users. Other centrality measures include betweenness, which considers a node's position in terms of how it acts are bridges other nodes together, and closeness, which takes into account the distance between a node with others in the network. There are also studies that aim to detect communities hidden in a network structure based on measures such as edge betweenness and modularity (Newman and Girvan). Most work in SNA mainly focuses on analyzing link structures in social networks. Only a few recent studies consider sentiment as a factor in social network analysis. For instance, Wang, Li, Xu, and Wu (2017) look at both link structure and node-level sentiments (e.g., ratings on products) in a social network in order to detect "sentiment communities," i.e., closely connected users who also share common sentiments. Adding sentiments into social networks can enrich the information and help to gain more insights into the relationships between individuals and how they interact with each other. We consider our study an early attempt to discover social networks of members and their sentiments towards various topics by mining text of transcribed conversation.

2.4. Summary

Based on reviewing related literature, we argue that, both formal and informal communications could be instrumental in uncovering the tacit power structure in an organization. Transcribed conversations recorded in meeting minutes contain rich information about members' exchange of opinions on various topics. We have identified the following challenges in finding tacit power structures from this data source:

- The large amount of unstructured text makes manual codification an impractical effort. Some automated text-mining approach is desirable.
- Most exiting analytical techniques focus on discovering only one type of knowledge from data. For example, CRA focuses on identifying word centers; social network analysis focuses on analyzing link structures; and sentiment analysis focuses on determining sentiment polarity. There is no unified approach that analyzes information from all three aspects and presents a comprehensive view of an organization's power structure.

In order to address these challenges, we introduce a novel approach to mining transcribed communication text for uncovering hidden dynamics in organizations.

3. Methodology

This section introduces a two-stage design for discovering organizations' power structure by mining text from meeting minutes. In our two-stage design, Stage 1 is text mining and analysis, seeking to uncover hidden patterns in organizations by mining text documents and analyzing network structures. Stage 2 involves assessment and interpretation of the text analysis findings based on interviews with three key members of the co-op. The purpose of Stage 2 assessment of finding is not to validate the finding; rather, it is to evaluate the efficacy of our text analysis method. Before introducing our two-stage design, we provide a brief introduction of the research context and data source.

3.1. Research setting and data source

The research setting involves Country Natural Beef, an agricultural marketing and production cooperative (co-op henceforth). It sells natural beef products — the products do not contain hormone or antibiotics. The co-op is recognized as a leader in practicing holistic land management and environmental practices in ranching. Such value-driven co-op is characterized by strong voluntary member participation (Ashforth & Reingen, 2014). Culture plays a salient role in the functioning of a co-op. Fierce debate is common in co-op's democratic decision process. Such culture is manifested in lively decision deliberation and value-oriented rhetorics in debates.

A production and marketing co-op is considered a network-based organization (Shaffer, 1987; Zusman, 1992). In a co-op, each node as a production unit is an independent business entity. In ranching, a ranch family often is made up of multi-generation members who are involved in ranching business. The family businesses are private enterprises. Many are incorporated into a limited liability corporations (LLCs). Each business may also employ temporary and seasonal workers. In our research setting, the large ranch operations have more than 1000 mother cows and the smaller ones has less than 200. According to the by-law of the co-op, each production unit/family has one vote in the co-op.

The co-op is managed by a chairman and four managers who are in charge of marketing and production. Specific routine tasks are carried out by advocacy teams nominated by the members. In decision-making, the co-op adopts a democratic decision-making process where each business entity has one vote. Hence the co-op has a rather flat governance structure. Presumably each member has equal power in the functioning and decision of the co-op.

Because the co-op members live in distance rural areas across eight western states, weekly meetings are held over teleconferences. The co-op also holds two-day-and-half-day face-to-face meetings every year when major issues are deliberated and critical decisions made. Minutes of all meetings and teleconferences are transcribed verbatim and distributed to all members through emails. We were provided access to all minutes. From this co-op, we obtained a corpus of 502 documents (e.g., minutes, agendas) that record all meetings held between 2005 and 2012.

One research member of this study had extensive interaction with the co-op in a rural sociology research program. He pointed out that between 2007 and 2011 the co-op has gone through critical transformations. Members had engaged in intense debate and consequent decision had resulted in rapid growth of the co-op; at the same time, the growth came at the expense of member dissatisfaction and membership drop. Thus, we conclude weekly meeting minutes over this five years would capture the political dynamics of decision-making and network structure. Thus, we use complete meeting minutes of these five years to analyze tacit power structure of the co-op. Based on insights from the rural sociologist, the five years, four critical decisions are made — (1) taking position on labor union issue associated with one of its supplier (2007–2008); (2) insurances policy on ranchers with high rate of sick cattle (2008–2009) (3) adoption of an animal welfare standards (2009) (4) adoption of a bylaw change (2010–2011).

3.2. Stage one. Text mining and analysis

In our two-stage research design, Stage One involves using text-mining techniques to extract information from meeting minutes and further uncovering hidden power structure. Fig. 1 illustrates our overall analytical methodology, including five main modules: (1) *text mining*, (2) *social network analysis*, (3) *sentiment/emotion analysis*, (4) *topic network analysis*, and (5) *sentiment cluster analysis*. In this section, we will describe each of these modules in detail.

3.2.1. Text mining

The meeting minutes are complete transcription of dialogues of weekly teleconference calls and annual meetings from 2005 to 2012. They are a rich repository of information and opinion exchange between members in the co-op. We use text-mining techniques to extract key information from these meeting minutes.

Each document of meeting minutes document records a variety of information including names of attendees, agenda items, conversations on certain topics, etc. For our purpose of finding hidden organizational structures, we focus on extracting three types of information: *persons, topics* and *sentiments*. Identifying individual person is certainly critical in that we need to differentiate individuals in a power structure. Since meetings vary in issues discussed and people's engagements in these discussions also vary due to expertise and interests, extracting topics is also critical to differentiate people's power distribution across subjects. Furthermore, it is common to see people expressing different and opposite sentiments on the same topic. Hence, sentiments should be extracted from text as another important dimension.

As shown in Fig. 1, the Text Mining module contains three steps: (1) Sentence segmentation: for each document, we first segment it into individual sentences as the basic unit for further analysis; (2) Sentence parsing: we parse each sentence using a linguistic parser, which allows us to identify its syntactic structures; and (3) Information extraction: based on the lexical and syntactic characteristics of each sentence, we extract three types of information, persons, sentiments, and topics, using the following methods:

- *Persons*: We recognize all people's names, including full names ("John Doe"), first names ("John"), last names ("Doe"), and initials ("J. Doe"). Based on the complete member list of the co-op and several simple rules, we conduct name disambiguation and map each names occurring in the text to its corresponding full name as given in the member list.
- *Sentiments*: We extract sentiment words from sentences to infer individuals' opinions towards topics. For this purpose, a learning-based sentiment analysis model would require a large training set of labeled data, which is be costly but still cannot guarantee high accuracy due to the large variety of topics involved. Hence, we decide to choose a simple lexicon-based method for sentiment analysis in this research. There are several existing sentiment lexicons, including some general ones (e.g., Baccianella, Esuli, & Sebastiani, 2010; Hu & Liu, 2004; Stone, Dunphry, Smith, & Ogilvie, 1966; Tausczik & Pennebaker, 2010; J. Wiebe, Wilson, & Cardie, 2005), and some domain-specific ones (e.g., Asghar, Ahmad, Qasim, Zahra, & Kundi, 2016; Hamilton, Clark, Leskovec, & Jurafsky, 2016). In this study, we choose to use a general-purpose lexicon of subjectivity clues developed by Wiebe et al. (2004) because it not only indicates each word's sentiment polarity (positive vs. negative) but also its strength (strong vs. weak). Specifically, we assign each subjectivity clue a numeric weight as follows: strong positive (1), weak positive (0.5), strong negative (−1), and weak positive (−0.5), respectively. In this research, we ignore all negation cues in sentences. Last, we sum up the weights of all occurring sentiment words to get an overall sentiment score for each sentence as follows:

$$\begin{aligned}\text{Sentiment} = {} & (\#\,\text{of strong positve}) + (\#\,\text{of weak positve}) \times 0.5 \\ & - (\#\,\text{of strong negative}) + (\#\,\text{of weak negative}) \times 0.5\end{aligned}$$

- *Topics*: In linguistic analysis, noun phrases are often considered the key indicators of the subjects and/or objects of a sentence. Hence, from each sentence, we extract all noun phrases (NP's) based on the syntactic structure created by the parser in the previous step. After filtering out people's names, we consider the remaining NP's to represent key topics discussed during meetings. If an organization has some specific topics of interest, it can also define a list of corresponding keywords and identify them from sentences, in the similar manner as how we identify persons' names.

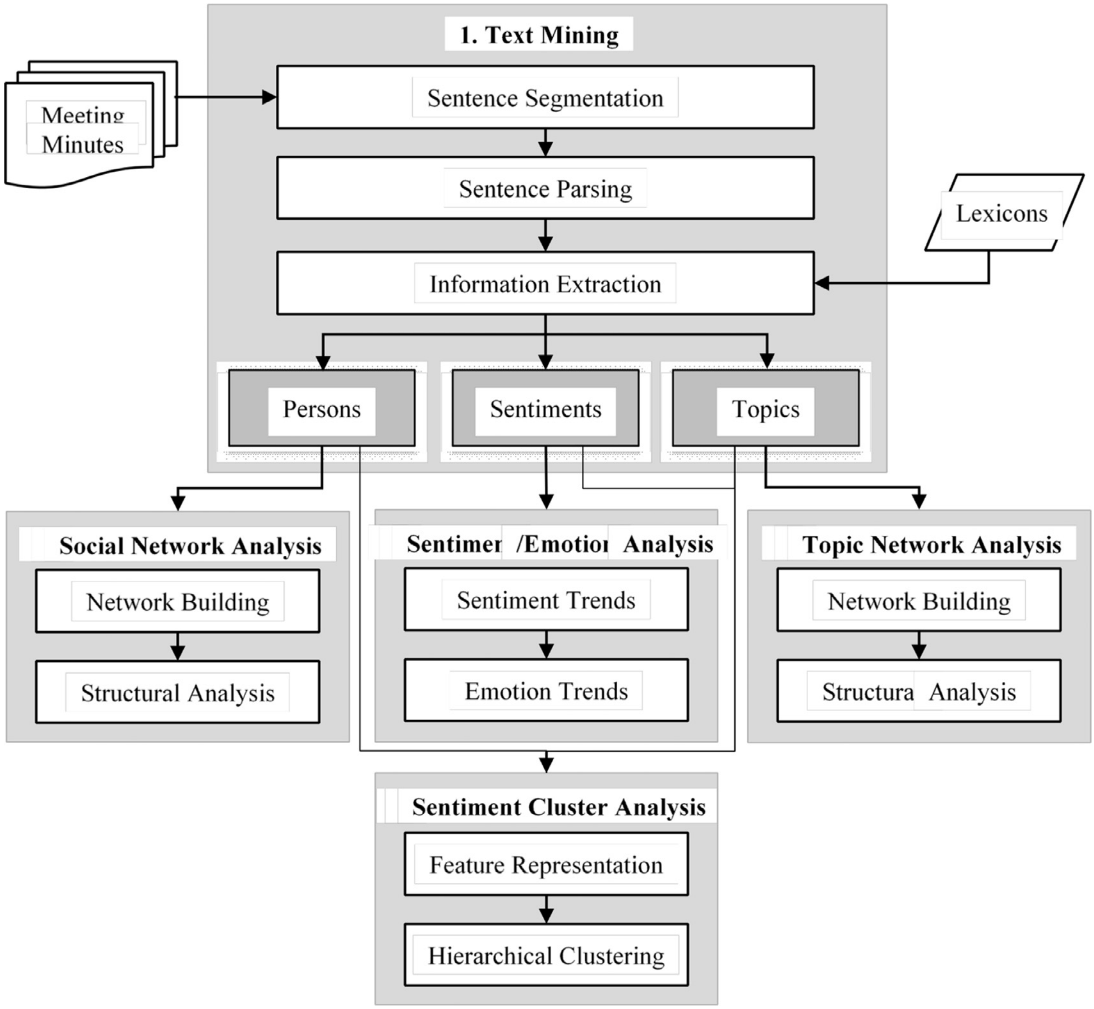


Fig. 1. Overall methodology of text mining and analysis of discourse.

It is worth noting that information extraction is only one step out of our overall methodology. Our research does not focus on developing new algorithms for information extraction or text mining. Hence, for the sake of simplicity and efficiency, we only adopt some basic and existing text-mining techniques in this research. We consider them to suffice in terms of demonstrating the effectiveness of our proposed approach. Depending on the needs and contexts, more advanced techniques could be utilized or developed here for text mining. For example, the sentiment analysis step could be improved by using a lexicon that is more specific to the domain of interest or by using an aspect-level sentiment analysis method (Asghar et al., 2016; Hamilton, et al., 2016); for topic identification, topic modeling techniques such as Latent Dirichlet Allocation (LDA) (Blei, Ng, & Jordan, 2003) can be used to extract topics at a higher semantic level.

Once these three types of information are extracted from meeting minutes, we carried out four types of analysis. Specifically, social network analysis is to discover the co-occurrence relationships between members. Topic network analysis is carried out to identify the most critical topics discussed in these meetings. Next, we perform a sentiment and emotion analysis is on the meeting minutes to detect sentiment/emotion trend over the years. Lastly, by combining all three types of information together, we aim to discover groups of members who share common opinions on the same topics.

3.2.2. Social network analysis

To understand the social relationships between members/nodes in the co-op we build a social network based on the co-occurrence analysis. Members vary in the way of participating in meetings and topics. If two members were both involved in the discussion of the same topic at some point, we consider that they share some common interest and therefore add a link between them. Over the course of meetings in years, two members may have been involved in discussion of multiple topics. Thus, we assign a weight to each link to represent the number of co-occurrences of the two members. If two members are often involved in discussing of certain topics, they tend to have a strong link (edge) between them.

From our entire collection of meeting minutes, we built a co-occurrence network of 155 nodes and 2279 edges. Out of the 155 nodes, about 110 are co-op members whereas the remaining nodes represent people who are outside the co-op but participated the meetings. In the co-occurrence network, each edge/link indicates the co-occurrence of two members in the same topic(s).

We used three common centrality measures as an indicator power and influence in network analysis (Freeman 1979). In the social network analysis, *degree centrality* is defined as the number of direct interactions a member has with others. *Betweenness centrality* measures the extent to which a node lies between other nodes in a network. It is measured as the number of geodesics (shortest paths between two nodes) passing through it. In a network, members with high betweenness scores often serve as bridges and brokers between different sub-groups in the network. They are the key communication channels for spreading information or resolving conflicts. Lastly, *closeness centrality* of a node is defined as the sum of the length of geodesics between this node and all others in a network. Members with low closeness scores are often peripheral outliers who are inactive and difficult to communicate with in the network (Xu and Chen 2005). These three centrality measures together capture the level of power and influence of nodes in a co-occurrence network from different aspects (Wasserman & Faust, 1994). We consider members with a higher centrality ranking to be opinion leaders in the organization. That is, members with high centrality scores tend to be the most active and out-spoken ones that are involved in a variety of matters in the organization. It is worth noting that, in such a social network, a co-occurrence link between two members only indicates their co-involvement in topics; it does not indicate whether they agree or disagree with each other. In other words, the link itself does not carry information regarding their opinions or sentiment. We will address this issue in sentiment cluster analysis.

3.2.3. Sentiment/emotion analysis

Given the sentiments extracted from meeting minutes, we can aggregate them by meeting, by month, by event, or by year in order to uncover the sentiment trend as a dynamic process in the organization. By comparing the usages of positive vs. negative words over time, we may identify some overall trends or peaks that indicate critical periods corresponding to events such as fights, celebrations, etc. In order to further differentiate more subtle sentiment traits, we use an emotion lexicon created by National Research Council Canada (NRC). The emotion lexicon contains a list of words and their associations with eight emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) and two sentiments (negative and positive). Annotations in this lexicon were manually curated through Amazon's Mechanical Turk (Mohammad & Turney, 2013). Emotion analysis allows us to find out how emotions fluctuated during several specific periods that define the historical development of the organization.

3.2.4. Topic network analysis

To understand what topics are important and how the topics evolve overtime, we adopt the centering resonance analysis (CRA) approach (Corman et al., 2002). CRA is a network-based approach in analyzing large quantities of written test. By analyzing the structural measures in a word network, CRA aim to identify the textual "centers," i.e., words (mostly nouns) that contributing the most to the topics/themes of the text. Furthermore, a comparison between two texts in terms of their respective centers can reveal the "resonance" or similarity.

Here, we adopt a three-step process of CRA, consisting of (1) selection, (2) linking, and (3) indexing. First, at step (1) selection, we reduce each sentence in meeting records as its noun phrases. These noun phrases are considered good representative of the main topic(s) of each sentence. Next, at step (2) linking, we build a topic network by adding an edge between two topic nodes if they co-occur in at least one sentence. In this sense, the topic network is essentially a co-occurrence network of topics. In addition, we add another dimension to CRA by calculating the overall sentiment score associated with each topic. In our module 1 "*text mining*," we have already estimated the sentiment of each sentence. Then, for each topic *t*, we calculate its overall sentiment score by summing up the sentiments of all sentences in which it appears. We posit that adding the sentiment score to each topic can enrich the topic network by providing new dimension about the overall tone on this topic in the organization. Finally, at step (3) indexing, using the network centrality measures, we analyze the structure of the topic network based on three centrality measures (degree, betweenness and closeness) so as to identify the "centers" that contribute the key topics of the text.

Table 1
Search conditions for events.

Keywords for all events: relationship, trust, fair, independence, customer, whole foods, feedlot, market, cattle

Events	Years	Unique Keywords
Union crisis	2007–2008	union
Out-of-program	2008–2009	OP, out of program
Animal welfare	2009	GAP, standard, animal, welfare
By-law	2011	by-law

Given a large collection of documents, CRA often results in a complex network consisting of hundreds or even thousands of connected nodes, which makes the topic network barely readable. Nevertheless, if a user has a more specific target of interest, he/she can apply some search criteria to narrow down the scope of analysis. We reason that this approach reflects practicality in a real organizational setting. A new comer or a manager would have basic information of its organization in terms of key events and the timeframe of these events. For our particular co-op dataset, we are aware a priori of four critical events. Therefore, it would make more sense to conduct CRA on each event, separately. Specifically, for each event, based on insights of the rural sociologist, we specified two filtering criteria: (1) the time period of the event, and (2) a list of keywords potentially related to the event. The first condition filters out the potentially relevant documents of meetings within the specified time period. With the keyword list, the system checks the occurrence of these words in each sentence and only keeps those containing *at least* one of these words in further analysis. In the meantime, we caution that the key words choice, based on suggestion of the rural sociologist, would invariably influence the structure of the network. Thus, we follow a principle of parsimony in choosing such key words. More specifically, in the context of agricultural cooperative, we create a set of general key words that capture the essential principles of co-ops (cite) instead of key words that are very specific to the events. These words are: *relationship, trust, fair, independence, market* and *customer*. We include a few more words to capture the names of the co-op, the key customers and the nature of the business including feedlot and cattle. Lastly, for each event, we add terms such as "*union*" and "*by-law*" that identify the events. Table 1 shows the search conditions, i.e., years and keywords, we specify for each event.

For each event, we will visualize central words as a topic network. In this network, each node is a keyword identified from meeting minutes and each link between two nodes indicates their co-occurrence in the same sentence(s). We follow Corman et al. (2002) and use betweenness centrality to identify the most central keywords in each topic network.

3.2.5. Sentiment cluster analysis

So far, based on simple co-occurrence analysis on sentences from meeting minutes, we create networks of members and networks of topics, respectively. Such a social network can show connections between members that indicate their co-participation of certain topics, whereas such a topic network shows connections of central topics discussed in meetings. These two types of networks tell the same story from two different dimensions, i.e., members and topics, respectively and separately. However, they cannot really reveal groups of members who not only care about the same issues in an organization but also share common opinions/sentiments towards these topics. In order to achieve this, we need to combine all three types of information, including people, topics and sentiments, together and determine the similarity between members. In this study, we apply cluster analysis to group members that ex-

Table 2
A summary of analyses performed on information extracted from meeting minutes.

	Information used			Techniques & measures	What to find
	Persons	Sentiments	Topics		
Social network analysis	√			Co-occurrence; centrality	Key influential members
Sentiment/emotion analysis		√		Distribution & trend analysis	Sentiment and emotion trends
Topic network analysis		√	√	CRA; centrality	Key topics
Sentiment cluster analysis	√	√	√	Hierarchical clustering	Groups of members with common opinions

press similar sentiment on common topics into the same cluster while separating those who have less in common in different clusters.

Given the three types of information (i.e., persons, topics, and sentiments) extracted from meeting minutes, we represent each person as a feature vector, in which each feature represents a topic and the value of each feature is his/her overall sentiment score to the topic:

$$\text{Person}_i :< \text{Topic}_{i1}, \text{Topic}_{i2}, \ldots, \text{Topic}_{in} > .$$

where the value of each feature is the sentiment score of the person to the topic.

Given all people represented as such a feature vector, we calculate the similarity between each pair of individuals, A and B, using a similarity measure. In this study, we choose the cosine similarity function because it is commonly used in the field of information retrieval and text mining:

$$\text{similarity}\,(A, B) = \cos(\theta) = \frac{\mathbf{A}\cdot\mathbf{B}}{\|\mathbf{A}\|\,\|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} (A_i)^2}\sqrt{\sum_{i=1}^{n} (B_i)^2}}$$

Next, we apply agglomerative hierarchical clustering to group people into a hierarchy of clusters. To avoid chaining effect in the process of hierarchical clustering, we use a complete linkage method to calculate the similarity between two clusters. The clustering process results in a hierarchy of clusters, where members holding similar opinions on certain topics tend to be grouped together while those with opposite opinions are separated in different groups. The clusters of members can reveal the hidden structures within the organization.

Table 2 summarizes the different types of analysis we perform on the information extracted from meeting minutes. Our social network analysis is to uncover the relationships between members and consequently identify individuals who are in the center and influential in the decision dialogue within the organization. Our sentiment/emotion analysis is to discover overall distributions and trends of different sentiments and emotions in the organization. The topic network analysis based on CRA is to identify key topics covered in meetings as well as their associated sentiments. In our topic network analysis, we incorporate sentiments associated with each topic and aggregate the sentiment scores at the node/topic level. Furthermore, we propose to use the sentiment cluster analysis is to reveal the similarity/difference between people's opinions towards different topics and uncover hidden opinion-sharing groups in an organization.

3.3. Stage two assessment and interpretation of the analytical results

The second stage of our research involves a mini-case study involving the focal co-op. The purpose of the case study is proof-of-concept. Our objective is to evaluate to what extend a manager can use our tools to diagnose an organization with rich text data.

To do so, we sought out assistance from three members of the co-op to evaluate our findings. The three members have a long membership history and extended knowledge of the events, individuals and social dynamics in the co-op. All of the three are founding members of the co-op. They are among the original 14 members and have been with the co-op since its founding 27 years ago. They have had administrative responsibility and actively participated in decision-making and committee works. Given the democratic nature of co-op governance, this suggests the three members are highly respected by members of the co-op.

We first debriefed these three members the overall objective of our research. Then we ask each of them separately to evaluate our findings. More specifically, we asked the three co-op members to evaluate the following: (1) whether top 10 members with high centrality rankings are indeed among the most influential members of the co-op; (2) whether the sentiment and emotion words describe the overall organizational mood and organizational atmosphere of the co-op; (3) whether the identified keywords and their relationships truly reveal the central issues during different events; and (4) whether the sentimental cluster analysis is able to detect camps or cliques of the co-op. Ultimately, when these analyses are put together, a manager can offer a general sense of the power center, political groups and overall "organizational health."

4. Findings and evaluation

In this section, we report our findings and describe how the co-op members assess these findings.

4.1. Social network analysis results

For each of the four events, we created a social network of members based on their co-occurrence in discussing about same topics during meetings. Table 3 summarizes the basic characteristics along with topological measures of the four social networks.

The size of co-occurrence networks varies across the four event periods. Among the four periods, Out-of-Program (2008–2009) contains the most nodes (108) and the most edges (1390), suggesting this event involves broad participation of members and external stakeholders.

By examining the centrality measures in these networks, we observed an increase in average degree centrality between 2007 to 2009 when the score reached a peak (0.2607) in the year 2009. Afterwards, the density started to decrease and reached the lowest score of 0.1803 in 2011. We also observed a similar pattern in average closeness centrality. In many social networks, it is evident that nodes tend to cluster together and create highly dense groups. We calculated and compared the clustering coefficient of the four social networks to measure the degree to which nodes in a graph tend to cluster together. The clustering coefficient also reached its peak (0.6893) in 2009 and dropped to the lowest (0.6491) in 2011. These overall patterns indicate that the co-op experienced an increase in intensity of interactions until around 2009–2010, which abated since then.

Table 4(a) summarizes the top 10 members in the network built based on all meeting records. We ranked the top 10 members that have the highest centrality measures. The choice of 10 is arbitrary, it represents about 10% of the co-op's membership. Granted it may

Table 3
Structural analysis of co-occurrence networks.

Events	Phase	# nodes	# edges	Average centrality scores			Clustering coefficient
				Degree	Betweenness	Closeness	
Union crisis	2007–2008	106	1137	0.2043	0.0114	0.4831	0.6752
Out-of-program	2008–2009	108	1390	0.2406	0.0095	0.4943	0.6603
Animal welfare	2009	89	1021	0.2607	0.0121	0.5145	0.6893
By-law	2011	86	659	0.1803	0.0123	0.4705	0.6491

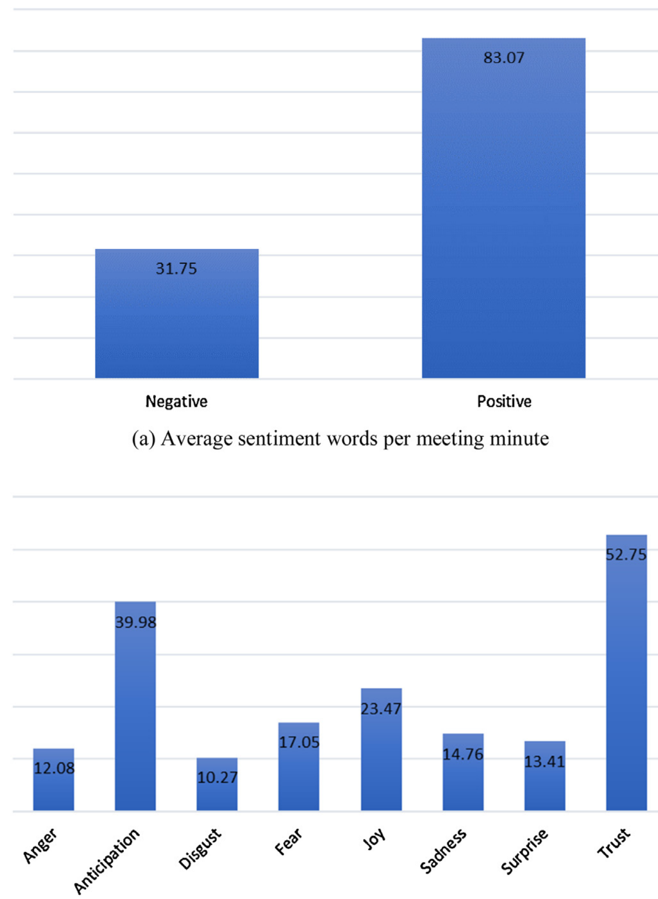
not present an exhaustive list of key players, it is a practical number for outsiders like a new manager to identify important individuals in the co-op. Table 4(b)–(e) summarize the top 10 members in each of the networks built for the four events.

Looking across the three centrality measures of all records in Table 4(a), we find five individuals that are ranked top 10 based on all three measures: David P., Sarah D., Cheryl H., Pam K., and Bryan H. We consider them to be the most vocal, active members and opinion leaders of the co-op. In addition, we also find four other members who appear in at least two out of the three top-10 list based on different measures: Margo B., Rachel H., Trevor P., and Mathew M. Furthermore, when we look into each individual event, the key players vary slightly with almost the same sets of members appear in each event. While we may include several more members (e.g. Ryan W., Roger H., Wanda P., and Jimmy W.) as influential individuals, the frequency of the same individuals gives us confidence that core individuals have been identified.

We present the results and our interpretation of the results to three co-op members/evaluators for evaluation. They validate our interpretation with the following assessment of our interpretation that the individuals identified as influential member of the co-op based on centrality measures. One caveat is that, while the evaluators agree that we identified the influential individuals of the co-op, it does not necessarily mean we have identified all the most influential individuals. The evaluator simply agreed with the list of people we presented to them.

In a different study, we collect social metric data through survey on the social network of the co-op members. The survey was sent to 81 active members, which eventually yielded 78 usable responses for a 96.3% response rate. The survey asked each members of the co-op two questions: (1) with whom "your ranch has directly interacted" and "who influenced your ranch's business operations in areas such as ranching practices, business decision-making, range and animal management etc.," and (2) with whom "your ranch has directly interacted" and "your ranch has discussed and provided suggestions about their business operations in areas such as ranching practices, business decision-making, range and animal management etc." Based on the response to these two questions, we build a social network of 78 members and seek to identify "influential members" of the co-op by analyzing centrality measures.

Given all top ranked members in Table 4(a), we remove all duplicates and have 15 unique members. By comparing the results from the two separate analyses, we found that 12 out of the 15 identified by the text analysis are also the individuals with the highest centrality scores in the social network. The three exceptions are: Rachel H, Rick K, and Trevor P. A closer look by our evaluator reveals that these three are indeed important players within certain clusters. The social network analysis largely reflects geographical characteristics of social interaction. Although the three are active in their own regions, they did not come out as the highest among all members by the social network analysis. Nevertheless, our text analysis was able to identify them as central members despite the geographical separation.



(a) Average sentiment words per meeting minute

(b) Average emotion words per meeting minute

Fig. 2. Average sentiments and emotions in meeting minutes.

Although the survey questions identify centrality based on the notion of advice-giving and, provisioning of assistance, they do identify members who are influential and active in the critical decisions and affairs of the co-op as shown in our text analysis of these key events. Therefore, the social metric data does not provide a proof, though it does provide a validation of our text analysis.

4.2. Sentiment/emotion analysis results

Fig. 2 shows the average occurrences of sentiment and emotion words in all meeting minutes. As shown in Fig. 2(a), the overall sentiment of the co-op is dominantly positive (83.07) against negative (31.75), which probably reflects the democratic form of governance and voluntary participation. Members are motivated by not only economic reason but also value and mission. Fig. 2(b) shows that, among the eight emotions, trust and anticipation are the two with the highest values, 52.75 and 39.98, whereas disgust and anger have the two lowest values, 10.27 and 12.08. These indicate the overall collegial and optimistic atmosphere in the co-op.

Table 4
Top 10 central members in social networks.

Rank	Degree	Betweenness	Closeness
(a) All records (fake names)			
1	David P	Sarah D	David P
2	Mathew M	Bryan H	Sarah D
3	Sarah D	David P	Mathew M
4	Cheryl H	Kimmy P	Ryan W
5	Ryan W	Amy H	Cheryl H
6	Trevor P	Jimmy W	Rachel H
7	Rachel H	Remy E	Bryan H
8	Margo B	Peri H	Kimmy P
9	Bryan H	Cheryl H	Trevor P
10	Kimmy P	Rick K	Margo B
(b) Union crisis			
1	David P	Sarah D	David P
2	Mathew M	Bryan H	Sarah D
3	Sarah D	David P	Mathew M
4	Cheryl H	Kimmy P	Ryan W
5	Ryan W	Amy H	Cheryl H
6	Trevor P	Jimmy W	Rachel H
7	Rachel H	Remy E	Bryan H
8	Margo B	Peri H	Kimmy P
9	Bryan H	Cheryl H	Trevor P
10	Kimmy P	Rick K	Margo B
(c) Out-of-program cattle			
1	David P	Sarah D	David P
2	Mathew M	Bryan H	Sarah D
3	Sarah D	David P	Mathew M
4	Cheryl H	Kimmy P	Ryan W
5	Ryan W	Amy H	Cheryl H
6	Trevor P	Jimmy W	Rachel H
7	Rachel H	Remy E	Bryan H
8	Margo B	Peri H	Kimmy P
9	Bryan H	Cheryl H	Trevor P
10	Kimmy P	Rick K	Margo B
(d) Animal welfare standards			
1	David P	Sarah D	Ryan W
2	Ryan W	Bryan H	David P
3	Sarah D	Kimmy P	Sarah D
4	Nathan B	David P	Kimmy P
5	Cheryl H	Amy H	Trevor P
6	Trevor P	Derrick L	Rachel H
7	Peri H	Peri H	Nathan B
8	Mathew M	Ryan W	Peri H
9	Kimmy P	Leo C	Cheryl H
10	Rachel H	Dean B	Mathew M
(e) Purchasing cattle and bylaw change			
1	Roger H	Ray E	Cheryl H
2	Cheryl H	Dean B	Roger H
3	Bryan H	Roger H	Wanda P
4	Peri H	John S	Ray E
5	Ray E	Cheryl H	Bryan H
6	Wanda P	Peri H	Kimmy P
7	Shawn R	Larry F	Peri H
8	Kimmy P	Jimmy W	Shawn R
9	David P	Wanda P	David P
10	Jason W	Sarah D	Jason W

Fig. 3 illustrates how average frequency of sentiment and emotion words have evolved over time from 2005 to 2012. First, both positive and negative emotions rise and fall at the same time (Fig. 3(a)). More specifically, Fig. 3(b) shows that largely consistent trend among the eight types of emotions. Starting from 2005 to 2008, the emotions are generally mild and stable; and they started to pick up from 2008 and reached they peak values around the year 2010; afterwards, they dropped again to lower levels. This observation offers two insights. First, the synchronized movement of emotion points to competing opinion and positions in the co-op, which corroborates the duality of value-oriented cooperative (Ashforth & Reingen, 2014). Second, the climax of heated debates in 2010 probably suggests a dramatic moment of the co-operative and perhaps an inflection point as the emotion abated precipitously.

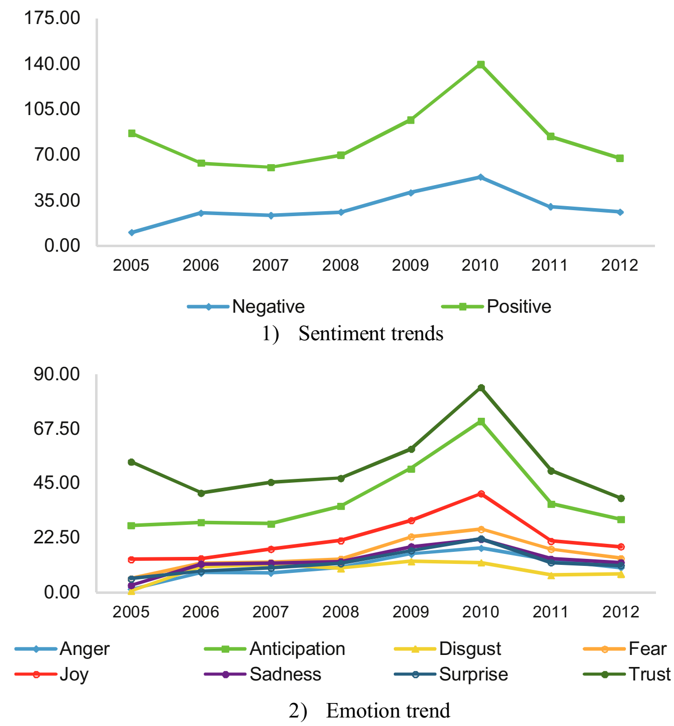


Fig. 3. Sentiment and emotion trend by year.

4.3. Topic network analysis results

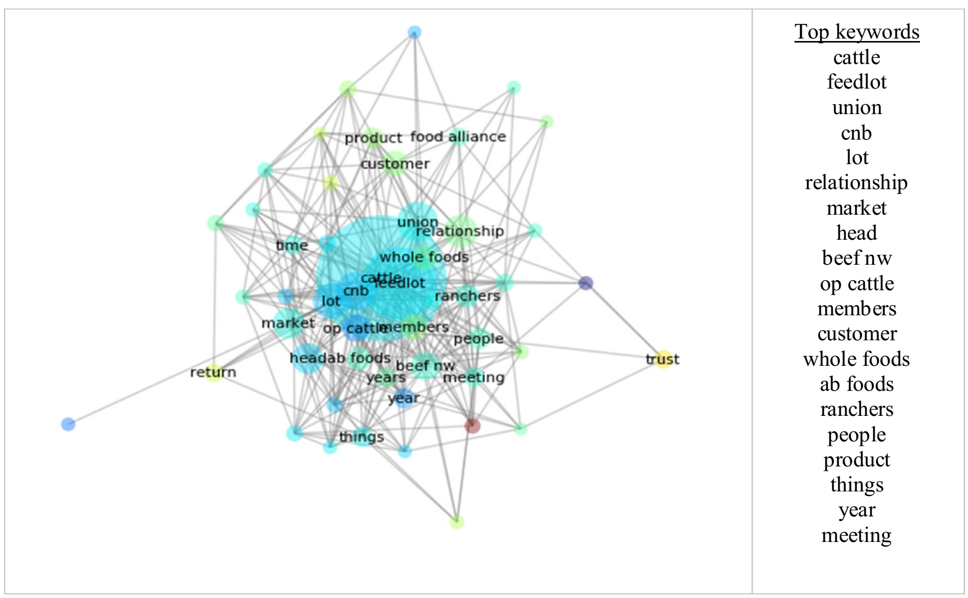
Fig. 4(a)–(d) show topic networks we built for the four major events. In a topic network, each node represents a keyword identified from related meeting minutes and each link indicates their co-occurrence in the same sentence(s). The size of a node indicates its betweenness centrality (Corman et al., 2002) and the color indicates the overall sentiment of sentences that contains the keyword. Beside each network, we also list the top 20 keywords ranked by centrality degree. Our results show that keywords such *cattle, feedlot, relationship, customers* and *ranchers* are among the top 20 in all four topic networks. These words are always be the focal topics in this co-op given the nature of this organization. In addition, for each particular event, we also identify several unique keywords with high centrality degree. For instance, during the event of union issue, the keyword *union* is ranked number three among all. During the event of animal welfare, keywords such as *GAP, animal compassion* and *standards* tend to gain more attention and discussion. Our evaluators also confirm these identified keywords to be the main topics involved during these four events.

Comparing the sentiments across the four events, we see that the overall sentiments differ. There are three interesting observations. First, the overall sentiment color of Bylaw Change is the "coolest" and the warm-cold color contrast is low. This is followed by the Union Issue. On the contrary, there are warmer sentiments in the Out-of-Program and Animal Welfare. This suggests that overall negative sentiment in the co-op during By-Law Change and Union Issue, whereas Out-of-Program and Animal Welfare generates more positive sentiments. Juxtaposing observation corroborates with the Patterns of Emotions and Sentiment Change analysis above, it appears that the dramatic emotional flare-up and precipitous drop in the last two events (2009–2011) the Animal Welfare and By-Law Change suggests a negative turn of the co-op – a period of intense negative emotion followed by a negative subdued emotional vibe.

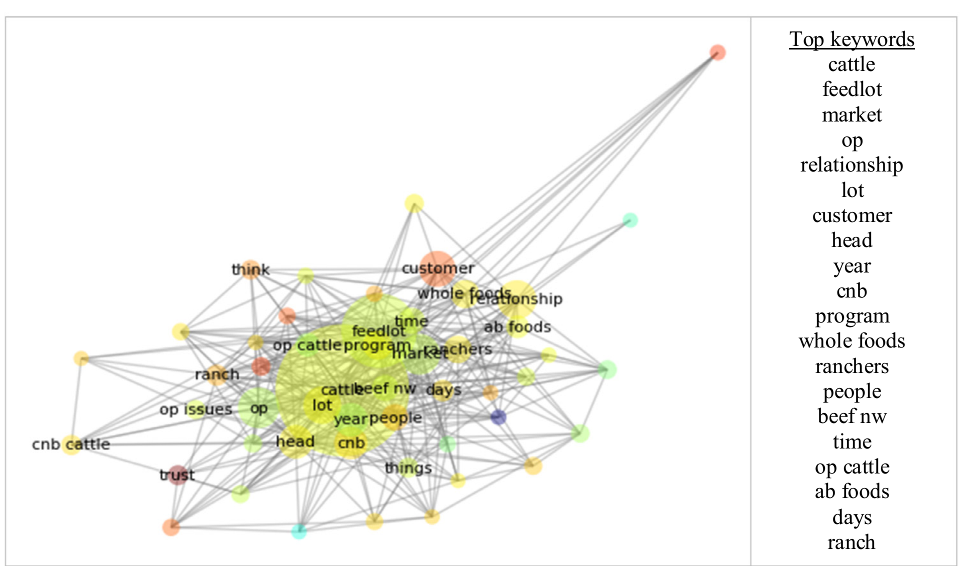
The three evaluators concur with our findings of the sentiment and emotion change over the observed period. When we asked them to explain our findings, they pointed out that in general the first two events did not bring up as strong emotion as the emotion in the third event - the Animal Welfare issue, where heated debate and conflicts persists among members as the event unfolded. The evaluators also agree that the Union Issue and By-Law Change instigate overwhelming negative sentiments of the majority of the co-op members, which is reflected the cooler colors shown in the topic networks. In other words, the uniformity of color, either cool color or warm color, suggests consensus in member sentiments; and the presence of contrasting colors indicates arguments and disagreement.

4.4. Sentiment cluster analysis results

Fig. 5 shows the hierarchy of sentiment clusters mined from all the meeting minutes. In this visualization, each circle represents a member or a cluster of members who share similar opinions on certain topics. The size of a "leaf" circle (i.e., a member) means his/her occurrences in all meeting records. Overall, the clustering hierarchy shows a somewhat concentric pattern, which means individual members or small sub-clusters are gradually added to form a bigger cluster. In the inner area of the view, there are several more clearly separated clusters. The hierarchy in Fig. 5 shows

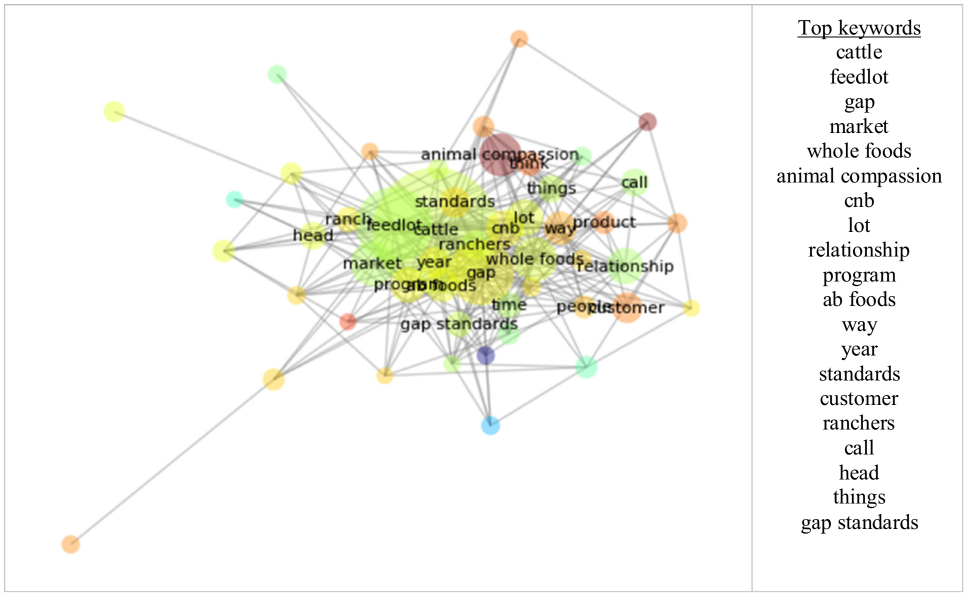


(a) Union Issue (2007-2008)

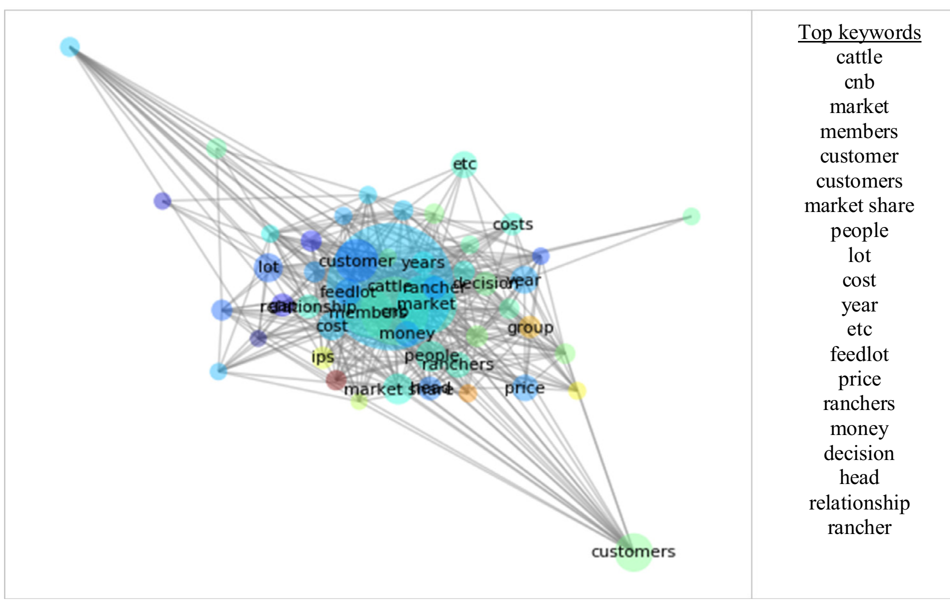


(b) Out-of-Program Issue (2008-2009)

Fig. 4. Topic networks of the four events.



(c) Animal Welfare Standard (2009)



(d) By-law Change (2011)

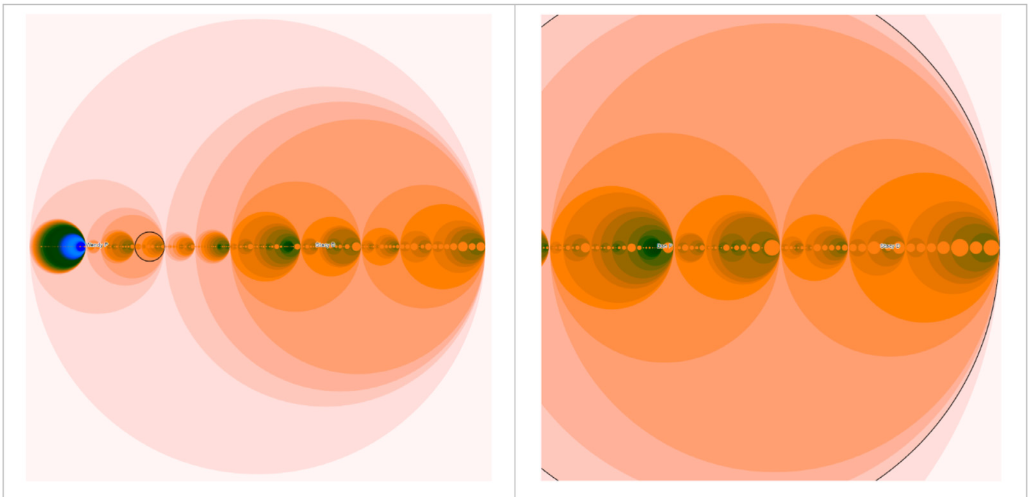
Fig. 4. Continued

a clear partition of two clusters. In our clustering hierarchy, at any level, each cluster is represented by the most active member, i.e., the one who has most frequently appeared in meeting records. In this particular hierarchy, the two top-level clusters are represented by *Wanda P* and *Sarah D*, respectively. This suggests two “cliques” or “camps” of members. Our system also provides an interactive feature that allows users to explore the clustering hierarchy by zooming in and out the visualization.

When reviewing our results of sentiment clusters, the three evaluators agree that the co-op is represented by two opposing groups (cliques/camps) led by the two individuals. The evaluator indicates that these two individuals represent polar value systems indicating the duality of co-op identity (Ashforth & Reingen, 2014); yet they are not the opinion leaders and champions even though they are not the ones with the strongest administrative power in the co-op organizations.

The objective of a manager is to understand the latent structure of the organization. While aggregating all discourse data would detect long-term cliques and camps of the organization, the same type of analysis can be conducted with a focus on a particular event of interest. Event-based sentiment cluster analysis would get at the cliques and opinion leaders of a particular historical event.

Discussion with the co-op evaluators also informs us that these analytical methods should be used together to provide a general sense of the co-op. One should not rely on a single method to draw a finite conclusion of the co-op. Rather, these analytical methods need to be used together to extrapolate. Repeated information such as the appearance of individuals across multiple lists of top influential members ranked by different centrality measures and the joint consideration of emotion and topics all together sheds light into the importance of certain individuals and the overall dynamics of the co-op.



(a) Overview (b) Zoom-in view

Fig. 5. Sentiment cluster hierarchy for all meeting records.

5. Conclusions

Discourse of organizational communication is a valuable data source for uncovering tacit structures and dynamics among members. However, organizational discourse analysis often requires process and understand a large volume of unstructured data about formal or informal communication. To address these challenges, in this study, we introduce a novel and comprehensive approach to analyzing organizational discourse data using text-mining techniques. The proposed approach provides a holistic view of the power structure implied by the communications. It not only suggests the opinion leaders but also demonstrates how people allied with each other on different issues within the network community. More importantly, by combining different text mining and analysis techniques, our approach can help to reveal both key topics related to certain events and people's sentiments towards these issues over time.

This research contributes to the domain of text mining by innovatively integrating and applying various text-mining techniques to uncover different aspects of a power structure from communication discourse in an organization. At the same time, the paper also contributes to the domain of supply chain management by introducing a novel analytical approach with the conventional communication discourse method as a guideline. This indicates the value and importance of integrating text mining into the research methodology of other management disciplines. While the conventional method could be limited by its manual approach, text-mining could certainly provide a more scalable solution. Text mining can only be useful when it starts from a specific information need in a managerial domain. The fact that our findings from analyzing a seven-year collection of meeting minutes were largely confirmed by members of the organization validates the effectiveness of the proposed approach. It is evident that our approach is of great potential for large companies/organizations to discover hidden structures from their big data of communication discourse.

In order to successfully carry out our approach, we recommend one follow this roadmap:

(1) **Information extraction**: Our approach extracts three types of information from discourse, including *persons, sentiments* and *topics*. Each type of information captures one aspect of the discourse data, which allows further analysis from multiple aspects. At this step, some lexicon or dictionary that can control or limit the scope of interest would be helpful to generate clean and interpretable results.

(2) **Multi-aspect analysis**: A comprehensive analysis of discourse requires wrangling data from multiple aspects with different approaches. Specifically, by analyzing co-occurrence of members in a social network, we can identify individuals with the most influential powers in the decision dialogue within an organization. Sentiment and emotion analysis reflects the dynamic trends of tones over time. Our topic network analysis can reveal identify key topics as well as their associated sentiments. Furthermore, our sentiment cluster analysis is to uncover hidden groups/camps of individuals with different propositions. Discussion with the co-op evaluators also informs us that these analytical methods should be used together to provide a general sense of the co-op. It is important to note that each of these analyses only provides one angle to look into the organization. One should not rely on a single method to draw a finite conclusion of the co-op. Rather, these analytical methods need to be used together to extrapolate and reveal the dynamics in a more holistic view. Repeated information such as the appearance of individuals across multiple lists of top influential members ranked by different centrality measures and the joint consideration of emotion and topics all together sheds light into the importance of certain individuals and the overall dynamics of the co-op.

(3) **Assessment**: As a tool for discovering and revealing tacit knowledge in an organization, this system should provide an interface for managers to interact with and evaluate on. A manager should be able to specify different search conditions (e.g., time period, events, keywords) to perform certain analysis and assess the results based on evidence.

In our future research, we plan to extend this work from the following directions. First, we will improve our system by enhancing the information extraction module for higher accuracy and quality. Second, we plan to improve the visualization functionality of our system for better interactivity and usability. Third, we will extend our study to analyze even larger amounts and more types of communication data (e.g., social collaboration tools, peer production platforms) for uncovering tacit intelligence and knowledge in businesses.

Acknowledgments

The authors thank the editors and anonymous reviewers for helpful feedback. K. Xu acknowledges financial support from the National Science Foundation of China [NSFC 71622008, 71301071], The China Ten Thousand Talent Program (Youth Talent Support Program), Jiangsu Planned Projects for Postdoctoral Research Funds (1601100C). All remaining errors are the authors own.

References

Abbasi, A., Chen, H., & Salem, A. (2008). Sentiment analysis in multiple languages: Feature selection for opinion classification in Web forums. *Acm Transactions on Information Systems, 26*, 1–34.
Ackerman, M. S., & Hadverson, C. A. (2000). Reexamining organizational memory. *Communications of the ACM, 43*, 58–64.
Allaho, M. Y., & Lee, W.-C. (2013). Analyzing the social ties and structure of contributors in open source software community. *2013 IEEE/ACM international conference on advances in social networks analysis and mining*. Niagara Falls, ONCanada: IEEE.
Antwerp, M. V., & Madey, G. (2010). The importance of social network structure in the open source software developer community. *Hawaii international conference on system sciences*. Honolulu, HawaiiUSA: IEEE.
Asghar, M., Ahmad, S., Qasim, M., Zahra, S., & Kundi, F. (2016). In *SentiHealth: Creating health-related sentiment lexicon using hybrid approach: 5* (p. 1139). SpringerPlus.
Ashforth, B. E., & Reingen, P. H. (2014). Functions of dysfunction: Managing the dynamics of an organizational duality in a natural food cooperative. *Administrative Science Quarterly, 59*, 474–516.
Atwood, M. E. (2002). Organizational memory systems: Challenges for information technology. In *Hawaii international conference on system sciences* (p. 104).
Baccianella, Stefano, Esuli, A., & Sebastiani, F. (2010). SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *the seventh conference on international language resources and evaluation* (pp. 2200–2204). VallettaMalta: European Language Resources Association.
Bird, C., Gourley, A., Devanbu, P., Gertz, M., & Swaminathan, A. (2006). Mining email social networks. In *Proceedings of the 2006 international workshop on mining software repositories*.
Bird, C., Pattison, D., D'Souza, R., Filkov, V., & Devanbu, P. (2008). Latent social structure in open source projects. In *Proceedings of the 16th ACM SIGSOFT international symposium on foundations of software engineering* (pp. 24–35).
Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research archive, 3*, 993–1022.
Chen, H., Chiang, R., & Storey, V. (2012). Business intelligence and analytics: From big data to big impact. *Mis Quarterly, 36*, 1165–1188.
Corman, S. R., Kuhn, T., Mcphee, R. D., & Dooley, K. J. (2002). Studying complex discursive systems centering resonance analysis of communication. *Human Communication Research, 28*, 157–206.
Davenport, T. H., & Prusak, L. (2001). Working knowledge: How organizations manage what they know. *The Journal of Technology Transfer, 26*, 396–397.
Diesner, J., & Carley, K. M. (2005). Exploration of communication networks from the enron email corpus. In SIAM international conference on data mining, *Newport Beach, CA* (pp. 3–14).
Dow, K. E., Hackbarth, G., & Wong, J. (2013). Data architectures for an organizational memory information system. *Journal of the Association for Information Science and Technology, 64*, 1345–1356.
Ellis, D. G. (1999). Crafting society: Ethnicity, class, and communication theory. *Communication Theory, 50*, 179–181.
Freeman, L. C. (1979). Centrality in social networks, Conceptual clarification. *Social Networks, 1*, 215–239.
Hatzivassiloglou, V., & Wiebe, J. M. (2003). Effects of adjective orientation and gradability on sentence subjectivity. In *Conference on computational linguistics* (pp. 299–305).
Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. *the ACM SIGKDD international conference on knowledge discovery & data mining*. Seattle, WashingtonNew York: ACM.
Kankanhalli, A., Tan, B., & Wei, K. K. (2005). Contributing knowledge to electronic knowledge repositories: An empirical investigation. *Mis Quarterly, 29*, 113–143.
Li, J., Wang, H. J., & Bai, X. (2015). An intelligent approach to data extraction and task identification for process mining. *Information Systems Frontiers, 17*, 1195–1208.
Liu, B. (2006). *Web data mining- Exploring hyperlinks*. Contents and Usage Data.
Mohammad, S. M., & Turney, P. D. (2013). Crowdsourcing a word–Emotion association lexicon. *Computational Intelligence, 29*, 436–465.
Netzer, O., Feldman, R., Goldenberg, J., & Fresko, M. (2012). Mine your own business: Market-structure surveillance through text mining. *Marketing Science, 3*, 521–543.
Nevo, D., & Wand, Y. (2005). Organizational memory information systems: A transactive memory approach. *Decision Support Systems, 39*, 549–562.
Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. In Proceedings of the conference on empirical methods in natural language processing, *Philadelphia, PA, USA* (pp. 79–86).
Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval, 2*, 1–135.
Putnam, L. L., & Fairhurst, G. T. (2001). *Discourse analysis in organizations: Issues and concerns*. Thousand Oaks, CA: Sage.
Raghu, T. S., & Chen, H. (2007). Cyberinfrastructure for homeland security: Advances in information sharing, data mining, and collaboration systems. *Decision Support Systems, 43*, 1321–1323.
Shaffer, J. D. (1987). *Thinking about farmers' cooperatives, contracts and economic coordination*. Washington, DC: Agricultural Cooperatives Services.
Stone, P. J., Dunphry, D. C., Smith, M. S., & Ogilvie, D. M. (1966). *The general inquirer: A computer approach to content analysis*. Cambridge, MA: MIT Press.
Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology, 29*, 24–54.
Van der Aalst, W. M. P. (2012). Process mining: Overview and opportunities. *ACM Transactions on Management Information Systems, 3*, 7.
Van der Aalst, W. M. P., & Weijters, A. (2004). Process mining: A research agenda. *Computers in Industry, 53*, 231–244.
Wang, D., Li, J., Xu, K., & Wu, Y. (2017). Sentiment community detection: Exploring sentiments and relationships in social networks. *Electronic Commerce Research, 17*, 103–132.
Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge University Press.
Wiebe, J., Wilson, T., Bruce, R., Bell, M., & Martin, M. (2004). Learning subjective language. *Computational Linguistics, 30*(3), 277–308.
Wiebe, J., Wilson, T., & Cardie, C. (2005). Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation, 39*, 165–210.
Wiebe, J. M., Bruce, R. F., & O'Hara, T. P. (2002). Development and use of a gold-standard data set for subjectivity classifications. *Pediatric Clinics of North America, 13*, 835–862 contd..
Hamilton, WilliamL., Clark, Kevin, Leskovec, Jure, & Jurafsky, Dan (2016). Inducing domain-specific sentiment lexicons from unlabeled corpora. *Empirical methods in natural language processing (EMNLP)*. Austin, Texas: Association for Computational Linguistics (ACL).
Xu, J., & Chen, H. (2005). Criminal network analysis and visualization. *Communications of the ACM, 48*(6), 100–107.
Xu, J., Christley, S., & Madey, G. (2006). Application of social network analysis to the study of open source software. *The Economics of Open Source Software Development, 31*, 205–224.
Yu, H., & Hatzivassiloglou, V. (2003). Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In *Conference on empirical methods in natural language processing* (pp. 129–136).
Zusman, P. (1992). Constitutional selection of collective-choice rules in a cooperative enterprise. *Journal of Economic Behavior & Organization, 17*, 353–362.