Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 15 / ISSUE 2 / MAY 2020 / 93-121
© 2020 Universidad de Talca - Chile

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762020000200107

Conformance Checking of the Activity Network with the Social Relationships Structure in the Context of Social Commerce

Leila Esmaeili[1] and Alireza Hashemi Golpayegani[2*]

[1] Amirkabir University of Technology, Department of Computer Engineering and Information Technology, Tehran, Iran, Leila.Esmaeili@aut.ac.ir
[2] Amirkabir University of Technology, Department of Computer Engineering and Information Technology, Tehran, Iran, Sa.Hashemi@aut.ac.ir

Received 1 December 2018; received in revised form 8 June 2019; accepted 10 June 2019

Abstract

Control and management of e-markets are one of the essential needs in social commerce context. Social commerce systems include not only informational flows, financial and goods/services but also social relationships and interactions. The focus of this research is to find the meaningful relationship between social relations and the network of various commercial flows among human resources (i.e., activity network) in social commerce context. If the activity network conforms with social relations, then business processes and, in general, the e-market can be steered by social relations. The proposed research framework has been constructed from a combination of process mining concepts, multi-layered network modeling, network analyzing, and social theories. Such a combination in the conformance checking field is an original contribution of this study, not yet explored. Two simulated sample datasets in four different versions are applied for the evaluation. Furthermore, Friedman's statistical test is employed to proof of the research hypothesis and gain more confidence in the obtained results. The results indicate that, despite the structural mismatch of the social network with the activity network, social relations affect the formation of business interactions among human resources. This effect is also different for various flow types. Multi-layered network modeling demonstrates this difference.

Keywords: Social commerce, Conformance checking, Social network analysis, Multi-layered network, Triadic closure

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762020000200107

1 Introduction

Social commerce (s-commerce) points out to the usage of the Web 2.0 and its applications in electronic commerce (e-commerce) [36], [37], [102]. Despite the absence of a standard definition for "social commerce" [57], this term is defined as an Internet-based commercial application that makes use of Web 2.0 technologies and social media; and it supports user-generated content and social interactions. S-commerce considers the network of the buyers and sellers as a single platform that includes the trade activities and all related interactions and transactions before, during and after selling/buying (such as customer decision making, purchasing behaviour, buying goods/services, product support, etc.). All of these commercial activities and transactions are carried out inside online communities and markets. Therefore, an electronic social commerce system (ESCS) prepares a platform for commercial and working activities along with non-commercial or social activities. All stakeholders, including business personnel (i.e., human resources), customers, partners, services, etc. can interact with ESCS and with each other via considered interfaces. There are two main architecture models for ESCS [52], [63], but they have two common objectives: 1) realization of commercial goals, and 2) promoting collaboration and interaction among stakeholders [29].

In the context of business systems (e.g., electronic social commerce systems), various data are generated. Businesses analyze part of this data for a deeper understanding of user behaviour, customer interests, business processes, market trends, business values, and so on. Nonetheless, there is always a significant portion of the data that is not used for various reasons (For example, we are not aware of the data existence). In such a way, there is one portion of data within the main data source which is not in regular use, but it can help in decision making, market management and forecasting. This portion is known as "Dark Data" [74]. Dark data is generally in an ideal state, and it is a subset of big data [17]. In today's systems, dark data is about 80% of the total data [80]. Unstructured data usually are considered dark data. Log files are one of the dark data categories [74]. Log files are collected for any type of event, at any place and any time. Of course, the ultimate goal is not to store more events, but to convert event data into real values [84]. Social commerce contains data types including dark data. In social commerce researches, the use of dark data is not considered [29]. Therefore, studying the event logs (a dedicated log file that is sometimes generated from other logs and data [83]) and social relations in the context of social commerce can be substantial.

Studies express the importance of interactions among stakeholders in the context of s-commerce [21], [29], [82], [98]. By taking a closer look at the studies, it is specified that interaction is an important feature and one of the main elements of s-commerce [61]. Social commerce uses the capabilities of Web 2.0 to support interactions; as far as, some researchers have defined s-commerce as a transaction of interaction between the company and customer [33] or the result of interactions among customers [41]. All previous studies are focused on the interactions between customers and companies or interactions among customers. Furthermore, interactions have mentioned as encouraging of sharing information to increase trust in shopping or to increase the trust of customers to brands in the step before buying (from the customer viewpoint). Decisions to buy and buying also have considered as the result of interactions [33], [41]. Hence, there is a research gap from the business management viewpoint. Study of human resources' social relations is strongly essential for businesses, business process designers, business process managers, and even governments. What is the effect of social relations on the business interactions of human resources? How can we get benefit from the social relations for the improvement of the business processes in the social commerce context? But no significant attempt has still been made to answer the above questions.

The formation of business processes in the context of social commerce is based on stakeholders interactions [29]. Hence, the study of the impact of social relations on business interactions can enable us to predict the behaviour of the business market based on social relations. If the conformity of business interactions with social relations are determined, and the impact of social relations on business interactions is specified, the business market can be controlled and managed through changing social relations. Thus, social market management is the motive of our research in the scope of the problem.

In line with research motivation, the main goal of our research is to prove the impact of the social relations among human resources (social networks) on the human resources' commercial interactions network (activity network). The human resources interactions network is the result of various flows formed in the commercial context. Our research hypothesis is based on the fact that in the context of social commerce, commercial interactions of human resource are formed based on social relationships and influenced by social relations. Human resources business interactions result from the execution of at least one process. In fact, human resources tend to collaborate or refer to people they know. Since many years ago, many studies have confirmed the effect of individuals' friendships and informal relations on their cooperative relationships [38], [58], [67]. Accordingly, we seek to find a relationship between the social network and the activity network in the context of social commerce. The main research questions include:

- *Q1. Is there a difference among various networks of human resources interactions (for each type of flows)?*
- *Q2. Does the study of the influence of the structure of social relations network on the structure of the various networks of human resources interactions (for each type of flows) yield more reliable results compared to when there is no distinction between flows?*

Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 15 / ISSUE 2 / MAY 2020 / 93-121
© 2020 Universidad de Talca - Chile
This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762020000200107

Trades are based on processes in the context of social commerce and electronic markets. Moreover, processes are accomplished by individuals (human resources). Due to the execution of the process by various people (human resources), a network of work flows is created between them. Owing to the relationships among human resources are created on the basis of the process, process mining is proposed. Therefore, process mining is our research motivation from the viewpoint of solution. Social relations among individuals also have a network structure. Therefore, taking into account the network nature of the problem, social network analysis is applied. Our research methodology is based on the concepts and methods of social network analysis. Previously, the concepts and basic methods of social network analysis have been used in organizational mining [2], [5], [43], [53], [90], [91]. Organizational mining is one of the approaches of process mining. But multi-layered network modeling and social theories, such as triadic closure, have not been used yet. We use them in this research. Moreover, we use statistical analysis and hypothesis testing to verify the validity of the research hypothesis. Also, to answer the research questions, we use the social data and event logs of one simulated process in the context of a social commerce system.

One of the important subjects of process mining is conformance checking. The main contribution in the problem domain is conformance checking of the structure of human resources commercial interactions with their social relations' structure in the context of social commerce systems. Social commerce systems include information, goods/services, and financial flows [9], [68]. These systems also incorporate social relations [11], [29], [34]. Thus, they are different with information systems which only contain information flows [10], [24], [39]. In the solution domain based on the organizational perspective, our contribution also provides a framework for conformance checking of the two networks - the human resources' commercial interactions network based on various flows and the social network - in the context of social commerce systems. Thus, the formulation of the problem based on multi-layered networks is done in this research for the first time. Our proposed framework is based on the concept of link prediction and triadic closure, which has not been considered so far.

The article is organized as follow. Sections 2 and 3 explain the related work and the theoretical background respectively. Definition and modeling of the problem are presented in section 4. Section 5 also explains the research methodology, the proposed research framework, and evaluation strategy. Dataset and the results of the experiments are presented in section 6. Finally, section 7 implies discusses, and section 8 outline some conclusions.

2 Related Works and Theoretical Backgrounds

In this section, we review the research backgrounds, e.g., event logs concept, the studies of conformance checking, and the activity network phrase.

2.1 Event Logs

Event data is a type of dark data [74]. Regarding the business process management perspective, event data is related to the execution of a particular business process, and it is recorded in a wide variety of data sources such as databases, flat files, message logs, transaction logs, enterprise resource planning (ERP) systems and document management systems. Each event refers to a particular activity. An activity is a well-defined step in the business process. Each event can be recorded with extra information such as executive resource (e.g., person, device, machine or computer servers), initiator of the activity, time label or other recorded data factors in the event (e.g., the execution place of activity) [5]. Consequently, doing any activity can lead to record an event. The activities are not only limited to the business processes. From a systemic perspective, web or mobile application systems (e.g., social networks) also include a set of activities. Messaging, sharing content, commenting, etc. are among the possible activities on social networks [48]. A fragment of event logs is displayed in table 1. An event (Pid, a, r, t) describes something happened during the execution of a specific business process *P* [85]. It is characterized by a set of mandatory characteristics. They have written in the header row of table 1.

Table 1: A fragment of event logs

Business Process ID (*Pid*)	Activity Name (*a*)	Timestamp (*t*)	Resource Name (*r*)
1	Registeration request	2018-05-04: 16:52	Pete
1	Check ticket	2018-05-04: 16:55	Sara
2	Registeration request	2018-05-04: 16:53	Mike
2	Check ticket	2018-05-04: 16:56	Server 1
2	Examine causality	2018-05-04: 16:58	Ellen
1	Examine causality	2018-05-04: 16:58	Mike
3	Registeration request	2018-05-04: 17:10	Pete
3	Check ticket	2018-05-04: 17:14	Server 1
…	…	…	…

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762020000200107

2.2 Conformance Checking

Process mining has been introduced since 2012 as an integral part of data science, which was constructed by event data [83], [87]. Techniques, tools and process mining methods could extract knowledge from event logs which usually exist in today's information systems [87]. Knowledge extraction leads to understanding the behaviour of individuals, organizations, machines, and systems [83]. The benefits of the process mining come from several different viewpoints [6], [75], [87]: control flow perspective, executive perspective, organizational perspective, and case. One of the important issues in all approaches of process mining is conformance checking. Conformance checking has been mostly studied from the control flow viewpoint [3], [7], [75], [85], [87], [96].

One of the important goals in business process management (consequently, process mining) is conformance checking. Conformance checking compares the model extracted from the event logs with the primary default model [85]. From the control flow viewpoint, the model is a business process model or workflow model. In this model, the priority of activities and their logical connection are determined. For instance, there may be a process model indicating that purchase orders of more than one million Euro require two checks. Analysis of the event logs will show whether this rule is followed or not. Another example is the checking of the so-called *four-eyes* principle stating that particular activities should not be executed by one and the same person. By scanning the event logs using a model specifying these requirements, one can discover potential cases of fraud. Hence, conformance checking may be used to detect, locate and explain deviations, and to measure the severity of these deviations [85]. Some studies with this approach are discussed here.

Rozinat and van der Aalst examined the matching of the business process model and event logs in process-aware information systems. They developed a matching checker in the ProM framework [70]. van der Aalst and de Medeiros advocated the use of process mining techniques (Alpha algorithm) to detect anomalous process executions and check process conformance [88]. Furthermore, Accorsi and Stocker used conformance checking with the control flow viewpoint for security audits [1].

The author in [86] identified two ways to create and/or maintain the fit between business processes and supporting information systems: delta analysis and conformance testing. Delta analysis compares the discovered model with some predefined processes model. Conformance testing attempts to quantify the *fit* between the event logs and some predefined processes model. van der Aalst et. al studied the behaviour of the web services through conformance checking. Their study addresses the problem of checking and quantifying how much the actual behaviour of a web service, as recorded in event logs, conforms to the expected behaviour as specified in a process model [89]. In recent years, other similar studies also performed conformance checking regard to process model discovery aspect [16], [51], [79], [92].

While the conformance checking is not limited to control flow [85], [101], so far, the conformance checking has been studied from the perspective of the control flow viewpoint [3], [7], [16], [51], [75], [79], [83], [85], [92], [96]. In organizational mining, the model can be an organizational structure [84], [86]. The resources interactions in the event logs can be mapped to a network based on the handover of works scenario and its conformity to the organizational structure can be investigated. Although some studies have used the old clustering methods to explore the organizational structure [2], [31], [53], [75], [90], none of them have focused on conformance checking.

In this research, we seek to find a structured and meaningful relationship between the social network and the human resources' interactions network in the context of social commerce. To this end, while modeling the problem with the multi-layered network, we employ the concepts and methods of link prediction and triadic closure. Conformance checking between the two networks (social network and activity network), and also using statistical analysis fulfills our goal and confirms our approach. Accordingly, some of our research differences in the problem domain are:

- Conformance checking studies have so far been addressed in the context of information systems and web services interactions, and have not been used in other areas such as electronic commerce and social commerce. The study of social commerce systems differs from that of information systems. Social commerce is the evolution of e-commerce, which includes both business and social activities together [10], [24], [39]. Therefore, while information systems in the organization are analyzed only based on the information flow, in social commerce systems there is not only information flow but also there are financial and goods/services flows and can be analyzed [9], [68]. Social relations are also available in these systems [11], [34]. Hence, it is possible to examine each of the flows independently and all three flows together with social relations; studying them can create a complete understanding of the various dimensions of the business.

- Previous studies have focused on conformance checking from the control flow point of view, while it can be investigated from the organizational viewpoint and based on the relations among process resources [85]. Therefore, by considering the impact of social relations on business interactions and people's co-operation [38], [58], [67], the initial model in conformance checking can be the structure of human resources' social relations (or any structure based on formal relations such as organizational structure or informal relations like

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762020000200107

a friendship network). As a result, the matching of business interactions among human resources in the event logs is compared to that of the initial model, and vice versa. In this research, we are looking for this objective.

Consideration of the mentioned differences in the problem domain, for the following reasons we can not employ the approaches of conformance checking that are proposed in process model discovery studies. Conformance checking in the process model approach is performed between the process model discovered from the event log and the initial process model [16], [51], [70], [79], [85], [92]. A process involves a set of tasks that between every two tasks exists a logical link [86]. The event log is generated based on the structure of the initial process. A discovered process model which generates an event log that is similar to the event log of the initial model is appropriate. Also, the initial model's event log should be tracked by the discovered model (see the conformance checking metrics in [16], [51], [79], and [92]). Therefore, the input of these methods is the event log, and their output is the process model discovered from the event log. Finally, the conformance rate is reported based on the evaluation metrics.

In this research, equivalent to the process model, we have a network of social relations among human resources. But the logical structure does not govern social relations. On the other hand, there is a process which makes some interactions among human resources, and the execution of the process is stored in the event log indirectly. We aim to study the influence of the social relations network on the execution of the process and, in particular, we seek to know how the work interactions form among human resources. Thus, regards to the goal of this research, we do not discover the process model or organizational structure from the event logs. Hence, the methods and approaches of previous studies that focus on discovering a process model or discovering an organizational structure are not effective in our research. In our study, the input is the event log and the social relations network. These two are from two separate contexts. We also have two outputs. 1) The intermediate output includes the activity network extracted from the event logs and the activity network generated based on social relations. The second activity network is based on methods and concepts of social network analysis. Activity networks consist of human resources. Both human resources are linked to each other based on the activities they have performed. As a result, activity network is different from the process model. 2) The final output which is the values of the conformance of the activity network and the network of social relations. Therefore, the methods used in the process model discovery researches are not appropriate for our research.

2.3 Activity Network

The term *network activity* is also used in researches on social network analysis. An activity network in social network analysis is a result of activities that users do in social network sites. For example, commenting, liking, following and so on. In fact, the activity network in the social networks domain includes users as nodes and what they do in their interactions, is an edge [8], [20], [44], [48], [72]. For example, if user A sends a message to user B, a directional edge is defined from A to B. Usually, the activity network is a directed weighted network in social network analysis researches. But in general, the type of activity network depends on its construction scenario [48].

In researches of social network analysis, the activity network has not formed based on the business process. In other words, there is no assumption that the interactions among users on social networks are purposeful. Most researchers have structurally compared the network of users' friendships with users' activity network; for example, network evolution, degree distribution, degree correlation, the strength of relations, etc. Some also compare the characteristics of two networks over time. Khadangi has summarized these studies in [48]. For more information, refer to his study.

The activity network is formed based on one/several processes in our research; the process which is predefined or can be ad-hoc. Therefore, a set of rules governs it. For example, the tasks that a human resource can do is limited and specified. That is, the resource can not do all the tasks in the business context. The activity network can be extracted from the event logs according to different scenarios (see Section 3.1). In the previous researches, the term "sociogram" [91] was used instead of the activity network. But in this research there is also the network of social relations, so we used the activity network term to avoid confusion. Also, since relations among nodes are based on activities that human resources have done in the business process, the term activity network is more appropriate than sociogram.

To the best of our knowledge and review of process mining studies, there has been no research on the study of the relationship between social relations and activity network so far, and we believe this study is the first one.

3 Theoretical Backgrounds

In this section, the theoretical backgrounds such as social network analysis and statistical test are reviewed.

3.1 Social Network Analysis

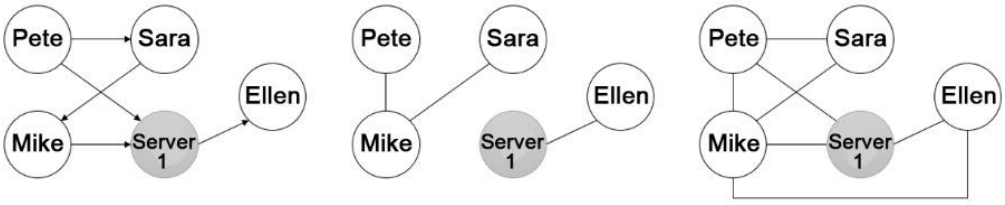
Social network analysis (SNA) or graph mining explores the social relations and network structures concerning graph theory. Social network analysis includes methods and metrics to extract information and knowledge from network structures. According to the graph theory, a single-layered network $SN = (V, E)$ is formed by a not-null set of vertices

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762020000200107

(V) (i.e., members or actors) and a set of edges (E) (i.e., homogeneous relationships) which indicates an explicit or implicit relationship between two vertices [99]. Sets V and E are defined based on a scenario. Defining different scenarios to generate a network leads to the formation of different types of networks [28]. For example, the set of vertices can be defined as the resources of a process and the set of edges could be determined based on the handover of work. After the extraction of network from the event logs (e.g., based on the handover of work, joint activity, working together, etc. [90], [91]), analysis of the network based on the NA methods and metrics could be accomplished.

Since the social network analysis is based on the graph theory, and the graphs are categorized according to the features of the vertices and edges, the network extracted from the event logs could be varied as well (figure 1). Usually, the derived network based on the event logs is a graph with weighted and directed edges if the edges define based on the handover of works. The weight of each edge is equivalent to the frequency of occurrence of that edge. This graph has the homogeneous relationships with one type (e.g., the handover of work) and due to the existence of one vertex type (e.g., resources), it is considered as a one-mode network. When the edges are defined based on working together or joint activity, the graph will be undirected but can be weighted. Regards to figure 1 (a), Pete refers a case to Sara. Thus, a directed edge has drawn among them. In figure 1 (b), Pete and Mike have an edge since they do one joint activity entitled with *registration request*. Likewise, in figure 1 (c), due to the collaboration among Pete, Sara, and Mike for $Pid = 1$, these three persons are connected.

These social networks are named sociogram in other studies [91]. In this research, we denominate them activity network (see section 5.1). The tools such as Prom, NetMiner, and USINET have been employed hitherto in the research to analyze the networks extracted from the event logs.



(a) based on handover of work (b) based on joint activity (c) based on working together

Figure 1: Extracted networks based on different scenarios for event logs in table 1
(The grey node is an inhuman resource)

Applying metrics and methods of NA, one can compare the individuals, organizations, and communications. These metrics are divided into two main categories: 1) micro-level metrics, i.e., metrics that examine only one specific node and 2) macro-level metrics, i.e., metrics that analyze the entire network.

Micro-level metrics: the role and effect of a specific vertex could be determined based on the metrics of micro-level. For instance, it could be figured out whether a vertex has the leadership role or it is only an isolated node relative to the entire network; whether a node is a bridge for connecting two key nodes or not [4]. Some of the important micro-level metrics include degree centrality (in-degree and out-degree), betweenness centrality [4], [75], closeness centrality (in closeness and out closeness), eigenvector centrality [4], and clustering coefficient. Each one of the given metrics is interpreted according to the application context. We have employed degree centrality metric in this research. It is defined as the number of links incident upon a node (i.e., the number of ties that a node has) [99]. In figure 1 (b) the degree centrality of node Mike is two.

Macro-level metrics: macro-level metrics provide insight and understanding of the overall network structure. These metrics include component, density, clustering coefficient, and centrality. Each metric has a different interpretation depending on the application context. For example, as there are fewer central nodes in the network, it will lead to increase the network centrality. One or a few of nodes administrate the network with high centrality. Therefore, if because of any reason this node is removed from the network, the network is divided into some disconnected sub-networks. The network with high centrality is not a good sign owing to all the credit and power of the network is accumulated in one person. In this research, the metrics of the component, density, and clustering coefficient are used.

In an unconnected graph, each connected part of the graph is called a component [99]. Furthermore, in an undirected simple graph, the density is defined as the ratio of the number of edges of the graph to the number of possible edges among nodes [99]. The clustering coefficient is defined based on triplets of nodes. A triplet is three nodes that are connected by either two (open triplet) or three (closed triplet) ties. The clustering coefficient is equal to the ratio of the number of closed triplets to the total number of triplets (open and closed triplets) [27], [95]. Nodes in a graph with a high clustering coefficient have more tendency to form clusters or communities.

3.1.1 Multi-layered Network

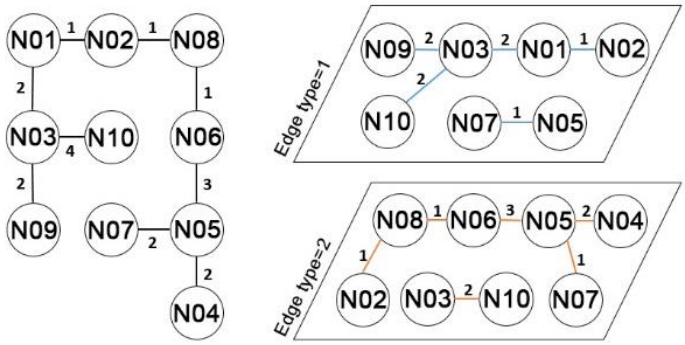
There is not always one type of relation (E) among the vertices (V) of a network. For example, there may be three types of relations within a network in which nodes are individuals: friendship, cooperation, and fellow-citizen. Modeling such a network can be done in two ways. The first is to weigh each of the relations (weights of the edges can be equal

Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 15 / ISSUE 2 / MAY 2020 / 93-121
© 2020 Universidad de Talca - Chile
This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762020000200107

or different), then we define a simple network with weighted edges (the weight of each edge is equal to the total weights of each type of edges). Obviously, in this situation, the type of relation cannot be determined using the weight of the edges. Therefore, the reality is not fully modeled and the analysis may not be accurate. The second method is to consider each relationship as a simple network. Consequently, the network is represented as three simple networks or, in other words, a 3-layered network [23], [47]. In such a case, there may be differences between the vertices that exist on each layer with the vertices of the other layers. Also, depending on the application field, there may be edges between the vertices in two different layers [23]. Figure 2 shows an example of two modeling methods. There are no edges among different layers in this example.

Source	Target	Edge Type	Weight
N01	N02	1	1
N01	N03	1	2
N04	N05	2	2
N05	N06	2	3
N07	N05	2	1
N03	N09	1	2
N10	N03	1	2
N08	N06	2	1
N10	N03	2	2
N07	N05	1	1
N02	N08	1	1

(a) Adjacency matrix



(b) Simple network (c) Multi-layered network

Figure 2: A sample of two types of network modeling

3.1.2 Community

There are various issues in social network analysis. Detecting communities is one of these issues. Each community is a subset of the network in which individuals have highly internal relations, and they have less communication with individuals outside of their community [31]. Actually, in the field of information systems and systems of business process management, people inside the community can have the similar roles and responsibilities and represent the organizational units (it depends on the scenario of the network definition). Various methods and algorithms have been suggested by researchers for community detection [4], [100]. Communities' detection in the field of information systems and business process management systems helps to understand the organization and improvement the organizational collaboration. Despite the importance and application of community detection methods based on the network structure, the organizational structure discovery has focused on traditional clustering methods. Only Appice (2018) has used community detection methods to discover and analyze the organizational units [5].

3.1.3 Link Prediction and Triadic Closure

Relationships among nodes in networks are continually changing, new edges and nodes are added to the network over time, and old ones are deleted. Link prediction as a critical issue in social network analysis has attracted a lot of attention since link prediction is important for exploring and analyzing the evolution of networks [94], [97]. Current methods for predicting links are divided into both supervised and unsupervised categories. The unsupervised approach predicts the relationships without any background knowledge or step of learning. On the other hand, the supervised approach employs a trained model and the background knowledge which is acquired from the network to predict links. Without requiring the training step, the unsupervised methods often use the value of similarity or closeness and network structural features to predict links [73], [97].

We also focus on unsupervised methods, in this study. One of the essential mechanisms for building a new relationship in network structures is triadic closure [27], [32].

As a theoretical background, a triadic closure is a concept in the theory of social networks which initially proposed by German sociologist Georg Simmel [41]. Additionally, triadic closure is one of the famous theories of collaboration mechanism [12], [14]. The triadic closure is a feature among the three nodes A, B, and C, in a way that in the case of the existence of a relationship between A-B and A-C, there will be a relation between B-C. Thus, the underlying hypothesis is that if two persons have a common friend, the probability that they will become the friend is more [41], [93], [101]. The correctness of this feature in excessively large and complex networks is too extreme, but it is a useful simplification of a fact that can be employed to understand and predict networks [27]. The triadic closure mechanism is the most important and powerful model for the growth and development of social networks among other mechanisms [101].

In this research, not only do we use the basic concepts and metrics of social network analysis in simple networks for identifying organizational units and roles, but, we also apply multi-layered network modeling to study the different

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762020000200107

workflows and its relation with the social network. Likewise, we use link prediction methods to check the conformance and investigating the effect of social relations network on commercial interactions among resources. Based on the link prediction methods, the structure of cooperation relations which is expected to be formed according to the social relationships could be predicted (research hypothesis). Then, the anticipated structure of cooperation relations could be compared to the real structure of cooperation relations. The study of these two networks determines the degree of conformance of social relations with working relationships. Statistical analysis is employed to investigate the validity of the research hypothesis, moreover, to prove that the results are not obtained randomly.

3.2 Statistical Analysis

Many natural phenomena, systems, and algorithms are inherently random in various fields. Examples of these areas include social processes, image processing, signal processing, information theory, computer science, and social sciences [15], [22]. For example, publishing a Tweet on the Twitter social network has an inherently random mechanism [46] and it depends on the probability of occurrence of some events. Also, by changing the probability, the propagation process may change. Therefore, statistical analysis is required when a phenomenon is random [25].

Nonparametric statistical methods are statistical methods used to determine whether a random phenomenon has occurred based on an accident or a chance. In fact, these methods determine whether a phenomenon has occurred based on a non-chance and meaningful pattern or behaviour [64]. Nonparametric statistical tests or nonparametric hypothesis tests are divided into two classes: pairwise comparison and multiple comparisons [25]. The pairwise comparison is a comparison between the two algorithms (e.g., the Wilcoxon Signed Rank Test [64]), while the multiple comparisons compare more than two algorithms (e.g., the nonparametric Friedman test [64]). More information on nonparametric statistical tests is available in [25].

Similar to nonparametric test methods, some statistical measures are used in social network analysis to identify frequent patterns (i.e., motifs). Two z-score and p-value measures are applied to identify motifs. The pattern with the higher z-score and the lower p-value is the more important one [35], [45], [49], [66], [69]. Indeed, it confirms that this pattern is a meaningful pattern and is not formed randomly or by chance.

In this regard, in this paper, Friedman's famous nonparametric test is employed to confirm the validity of the research hypothesis using the proposed conformance checking method.

4 Problem Definition

In this section, we formally define the problem.

Definition 1: according to social network definition [99], the social network could be the representation of any structure of official/unofficial relationships among the individuals in our problem. Here, we consider the social network based on the friendship relationships. Therefore, $SN = (V_{SN}, E_{SN})$ is defined as a static, undirected and unweighted social network. V_{SN} is a set of people, and $E_{SN} \subseteq V_{SN} \times V_{SN}$ is a set of friendship relationships among these people.

Definition 2: based on the event logs definition [85], $EventLog = (C, A, R, T, L)$ is event logs for structured processes in the context of social commerce systems. The processes have the loop and iterative structures. C is a set of case identifiers; A is a set of activities (activities don't have the same name), R is a set of resources (including human and inhuman ones), T is a set of timestamps and L is a set of content labels or flow types. Each resource performs one to n activity based on his predefined role. Also, each activity has one type of flow. The label determines the type of flow which is transferred to the resource r_k by the resource r_j by doing the activity a_i (the type of work interaction between the resources r_j and r_k). The items of set L are $L = \{i, g, f\}$, which i is the label of the informational flow; it means that information is transferred between the resources r_j and r_k. Value g is the label of the goods flow, saying that a physical good is transferred between the resources r_j and r_k. Likewise, value f is financial flow; it means that some money has been exchanged between the resources r_j and r_k.

Various roles have no common tasks and resources.Due to there is no limitation for the number of human resources (HRs), the number of human resources in the various execution of business processes is variable.

In the s-commerce context, in addition to the human resources of any business, there are activities which are conducted by machinery resources or the people outside of business such as customers. For instance, searching for the desired product of the customer is conducted by the server and the filling out the ordering form by the customer. So, we have:

- Members of set V_{SN} are all human.
- There is a vertex i in R that does not exist in V_{SN}, so the vertex i is an inhuman resource.

Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 15 / ISSUE 2 / MAY 2020 / 93-121
© 2020 Universidad de Talca - Chile
This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762020000200107

- There is a vertex j in V_{SN} that does not exist in R, therefore, the vertex j is not a human resource, or has no role in the process execution.
- The number of customers is more than the total number of human resources.
- Moreover, to simplify the problem, the resources cannot have multiple roles.

Formally, we have such a definition of the main problem of the research:

Problem 1 (Conformance checking): regarding the data of the network SN and $EventLog$ in context of social commerce systems, our problem is finding a function like f which concludes the recorded commercial interactions among human resources in the $EventLog$ occur based on the relations of social network SN. Actually f can determines the correlation or conformance level of social network SN with commercial interactions of human resources by the value set of X.

$$\mathrm{f}: (\mathrm{SN}, \mathrm{EventLog}) \rightarrow \mathrm{EventLog\ conforms\ with\ SN\ by\ the\ value\ set\ of\ X} \quad (1)$$

5 Research Methodology

The research methodology is drawn in figure 3. We explained the research problem definition and its formulation in section 1 and 4. The proposed framework of conformance checking is represented in details in section 5.1. Section 6.1 describes the dataset. Results of research, analysis of them, results validation and evaluation are also described in sections 6.2 and 6.3. Dataset creation is independent of the proposed research framework. Therefore, they perform in parallel.

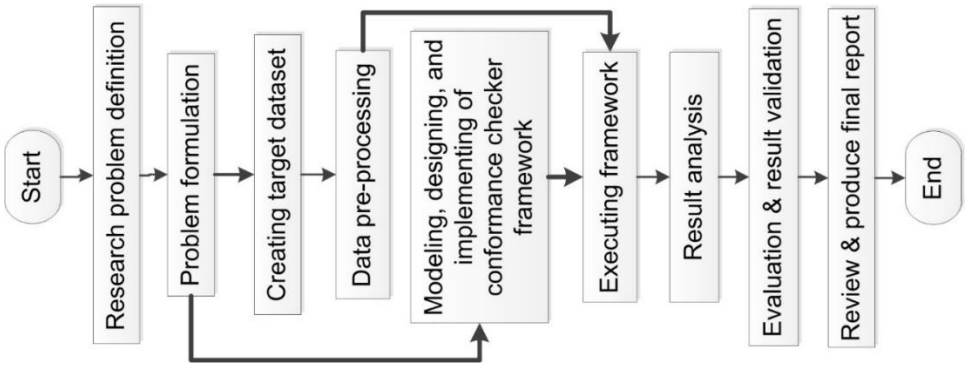


Figure 3: The flowchart of research methodology

5.1 Proposed Framework

The proposed framework of our model is drawn in figure 4. Function f has three sub-functions in the proposed framework: pre-processor, structural conformance checker, and behavioural conformance checker. Each function is explained in the following.

5.1.1 Pre-processor (SF1)

According to Definition 1, the human resources social relations (i.e., network $\boldsymbol{SN}$) have a graph-based structure. This is while the event logs do not have a graph-based structure. As a result, according to research questions, the interactions among human resources recorded in event logs should be mapped to a graph-based structure. Consequently, the two graph structures are comparable. In fact, we can check the conformation between the social network $\boldsymbol{SN}$ and the human resources interactions network (i.e., activity network) which is based on $\boldsymbol{EventLog}$. Activity network can be created based on different scenarios from the event logs. Due to the social network is an undirected network, a scenario for constructing activity network should be selected to create an undirected network. Therefore, the two networks will have comparison conditions. Scenarios of working together and joint activities create an undirected activity network [5], [31], [75], [77], [91]. Since the roles and workgroups of each resource are not specific, a scenario is required to identify the clusters of the role or workgroup of each resource. Thus, the scenario of joint activities is the most appropriate scenario in our problem.

Definition 3: According to the multi-layered network definition [23], [48], $AN_L = (V_{AN_L}, E_{AN_L}, L)$ is defined as a multi-layered, static, undirected and weighed activity network. L is the same set of flow type in $EventLog$. In other words, it specifies the interaction type among resources. Its values include $L = \{i, g, f\}$, when the value of index L is not set, all three values of L are considered. V_{AN_L} is a set of human resources in one of the layers of set L, and we have $V_{AN_L} \subset V_{SN}$ and $V_{AN_L} \subset R$ (in other words $V_{SN} \cap R = V_{AN}$ and $V_{SN} \cap V_{AN} \neq \phi$). $E_{AN_L} \subseteq V_{AN_L} \times V_{AN_L}$ is a set of relations based on

Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 15 / ISSUE 2 / MAY 2020 / 93-121
© 2020 Universidad de Talca - Chile
This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762020000200107

the joint activities among the human resources in one of the layers of set L. As an example, there is an edge between two vertices i and j in the network AN_g when nodes i and j at least have one same activity like A_n in various executions of business processes, in a way that the activity A_n is accompanied by goods exchange between nodes i and j. If financial transactions are carried out electronically, there may be no flow type f among human resources. Network AN_L is constructed through P01 in figure 4.

Therefore, a link is formed between two vertices i and j at AN_L when these two vertices have carried out one same activity at least in one of the layers of set L. Two vertices i and j do the same activity due to they have the same role. Each role includes a set of legitimate activities. On the other hand, when the vertices i and j have a link in one of the layers of network AN_L, it means that a single human resource has referred the work to them or two different human resources with the same role have referred the work to the vertices i and j. The number of joint activities between i and j is defined as the weight for their relationship.

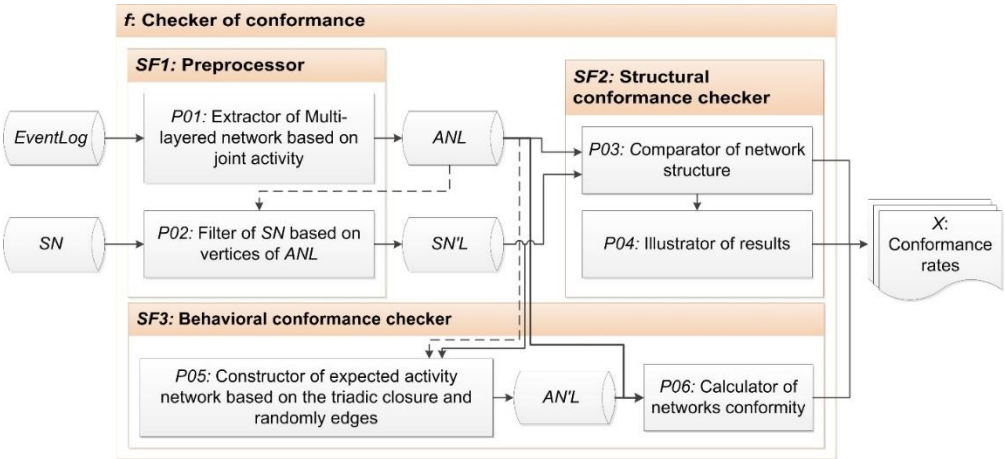


Figure 4: The proposed framework of the conformance checking based on the multi-layered networks

Due to all persons in network SN do not exist in network AN_L, filtering the network AN_L based on the network SN is necessary for computing reduction.

Definition 4: regarding the mentioned definitions, $SN'_L = (V_{SN'_L}, E_{SN'_L})$ is defined as a static, undirected and unweighted social network in a way that SN'_L is a sub-graph of SN. $V_{SN'_L}$ is a subset of V_{SN}, and it is a set of people who exist in network AN_L and their friends. Likewise, $E_{SN'_L} \subseteq V_{SN'_L} \times V_{SN'_L}$ is a set of friendship relationships among individuals. Network SN'_L is constructed by P02 in figure 4.

5.1.2 Structural Conformance Checker (SF2)

In response to research questions, two networks SN'_L and AN_L are examined and compared from different structural aspects. For this purpose, the basic metrics of social network analysis and some of the concepts of social science are applied. Structural conformance involves structural features of the network, homophily, social role (popularity), and link dependency. We will explain them in more detail. In Figure 4, P03 performs structural conformance checking calculations based on the NetworkX library in Python and VBA in Excel. Finally, P04 maps out charts and graphs.

Network structural characteristics: The structural characteristics of the network, are the main features that the network structure is defined by that. These features include the number of vertices $|V|$, the number of edges $|E|$, type of vertices, type of edges, density D, number of connected components $\#CC$, graph clustering coefficient GCC, and degree distribution. For each equal value of the features in networks AN_L and SN'_L, a value *True* is added to set X.

Homophily: Homophily principle [50] proves that people with similar characteristics (e.g., social situation, age and so on) tend to work together. Especially, we study the link homophily. We examine whether two people (human resource) who have mutual friendship edges in network SN'_L will tend to collaborate with each other in network AN_L. Is having more common friends in SN'_L equivalent to have a relationship in the network AN_L? The answer to these questions is examined for different flows. If there is a linear relationship between the number of mutual friends and the probability of connection, value *True* will add to set X.

Popularity: To the popularity, we investigate whether two popular individuals in social network SN'_L are connected to each other in the activity network AN_L or not? If more than 10 percent of relations are based on the popularity, *True* as another value will add to set X. Degree centrality is applied for measuring popularity [81]. The *popular* person has defined in a way that his/her degree centrality is among the top 10 percent ones. The other individuals are ordinary. 2-bit binary codes (XX) (X = 0 or 1) are employed to represent the dual status. Zero means an ordinary person and 1

Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 15 / ISSUE 2 / MAY 2020 / 93-121
© 2020 Universidad de Talca - Chile
This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762020000200107

denotes a popular one. Likewise, popularity status is determined based on the network SN'_L and the links are checked in the network AN_L for each different flow. As an example, how much is the probability of linking in network AN_i for status 11? Owing to the networks are undirected, there are three status of 00, 11 and 10 or 01.

Link dependency: Is referral of a case of work from one human resource to another one depending on their social relations? In fact, what fraction of human resources' commercial interactions have been formed based on the relation between two human resources who are connected in the social network SN'_L? We studied it for different flows in event logs. The value of the link dependency adds to set X as one of the values.

5.1.3 Behavioural Conformance Checker (SF3)

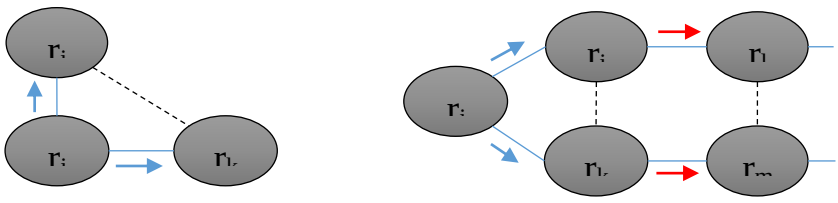
To answer the second question of the research, we suggest SF3. In fact, with the assumption of the effect of social relations on the formation of interactions between human resources, we anticipate the secondary network of the name AN'_L based on triadic closures. The relations of network AN'_L are based on this human resources' behaviour that they refer jobs to their social friends. Then we calculate the conformance rate of the real activity network AN_L with the AN'_L which is formed based on the social network SN'_L. The similarity of these two networks is reported as the behavioural conformance rate between networks AN_L and SN'_L (because network SN'_L influenced the formation of network AN'_L relations). Then, by evaluating and statistical analysis of the results, we prove that the assumption of the influence of social relations on the formation of the interactions among human resources is correct. In addition, network AN'_L has not been created in terms of chance and luck, and network SN'_L relations have affected them. Since the network AN'_L is formed on the basis of work referral behaviour which is based on the research hypothesis, this step of the study is called behavioural conformance checker.

A salient point in the triadic closure mechanism's generalization is that the triadic closure has been introduced in the context of one (social) network [40], [93], [101], but we expand it to the context of two different networks. If social relations influence the collaborations among the individuals in the business process, then, considering the generalization of the triadic closure mechanism, the expected activity network AN'_L is defined for each layer (each flow) based on definition 5. The extraction of network AN'_L is carried out by P05 in figure 4.

Definition 5: $AN'_L = (V_{AN'_L}, E_{AN'_L})$ is defined as a static and undirected activity network. $V_{AN'_L}$ is a set of human resources, and we have $V_{AN'_L} = V_{AN}$. $E_{AN'_L} \subseteq V_{AN'_L} \times V_{AN'_L}$ which is a set of predicted relations based on the social network of human resources. Two conditions for linking two individuals in the network AN'_L are determined. The proposed mechanism is performed by the function *Add_Edge_Predictably_Triadic_Closure* in algorithm 1.

Condition 1: There are three individuals r_i, r_j and r_k in a social network, based on figure 5 (a). There is also, two social relations $r_i r_j$ and $r_i r_k$ between them and no more information about the network are available.

- Human resource r_i has accomplished his task due to his role, and in the next step, he refers the case of work to the resource r_j or r_k with the probability of p (the value of p is based on the link dependency). Resource r_j and r_k are friends of r_i according to the social network structure. Since the authorized activities range of human resource r_i due to his role is limited and specified, his referenced works are also limited, specified and similar. If the resource r_i refer some works to resources r_j and r_k, thus, the resources r_j and r_k will carry out joint activities (regarding the joint activity scenario) while they will be linked to each other in the activity network with the probability of p^2 (equation 1). The relationship between the two resources r_j and r_k in the expected activity network is drawn by a dashed line in figure 5 (a). Thereupon, it is expected that these two nodes will also be in one community. The resources r_j and r_k could be the ordinary or popular individuals in social network SN'_L.



(a) Condition 1 (b) Condition 2

Figure 5: Link formation in network AN'_L based on the relations in network SN' (straight lines are the relations in the network SN', the directed arrows indicate case of work referral, and dashed lines are relations in the network AN'_L)

$$\begin{gathered} P(r_j r_k) = P(\overrightarrow{r_i r_k}) P(\overrightarrow{r_j r_i}) \\ P(\overrightarrow{r_i r_k}) = P(\overrightarrow{r_j r_i}) = p \\ P(r_j r_k) = p^2 \end{gathered} \tag{2}$$

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762020000200107

Condition 2: In accordance with figure 5 (b), there are five individuals of r_i, r_j, r_k, r_l and r_m and four edges of r_ir_j, r_ir_k, r_lr_j and r_kr_m in the social network.

- Resources r_l and r_m are linked to each other if the resources r_j and r_k are connected based on the condition 1, and according to the condition 1, the edges r_lr_j and r_kr_m are not formed. If edges r_lr_j and r_kr_m are predicted and the edge r_lr_m is added based on the condition 2, it means that four individuals r_j, r_k, r_l and r_m have performed joint activities and have the identical role. Hence, referring work from the resource r_j to the resource r_l or from the resource r_k to the resource r_m is not possible. The probability of referring work from the resource r_i to the resources r_j and r_k, also from the resource r_k to the resource r_m and from the resource r_j to r_l is equal to p. Consequently, the probability of establishing a relation in the network AN_L' between the resources r_l and r_m regarding their independence from each other is p^4 (equation 2). Similar to the condition 1, the resource r_l and r_m could be the ordinary or popular individuals in social network SN_L'.

$$\begin{gathered} P(r_lr_m) = P(r_jr_k)\, P(\overrightarrow{r_lr_j}|r_jr_k) P(\overrightarrow{r_kr_m}|r_ir_k) \\ P(\overrightarrow{r_lr_j}|r_jr_k) = P(\overrightarrow{r_kr_m}|r_ir_k) = p \\ P(r_lr_m) = p^4 \end{gathered} \tag{3}$$

The condition 2 can expand; in fact, the referral of work can be cascaded down to other human resources through their social relations. This referral expansion was ignored since the probability for building a relationship in the network AN_L' has dropped significantly.

Since the number of the predicted edges may be less than the number of network AN_L edges and the comparison becomes impossible, thus, the difference between the number of AN_L and AN_L' edges is provided through a random method using the function *Add_Edge_Randomly.*

ALGORITHM 1: Behavioural Conformance checking of network AN_L with SN_L'
Name: Conformance checking based on triadic closure (CCTC)

Input: network SN_L'network AN_L, L
Output: AN_L' , *conformance_rate[]*
Initialize $E_{AN_L'}^{Predictably} = \{\}$, $E_{AN_L'}^{Randomly} = \{\}$, $E_{AN_L'} = \{\}$, $V_{AN_L'} = V_{AN_L}$, $edge_counter = 0$, $p = value$
Begin
$E_{AN_L'}^{Predictably}$= *Add_Edge_ Predictably_Triadic_Closure* ($SN_L', V_{AN_L'}, p, edge_counter, |E_{AN_L}|, L$)
$E_{AN_L'}^{Randomly}$= *Add_Edge_Randomly*($V_{AN_L'}, edge_counter, |E_{AN_L}|$)
$E_{AN_L'} = E_{AN_L'}^{Predictably} \cup E_{AN_L'}^{Randomly}$
conformance_rate[] = *Calculate_Conf_Rate*(AN_L, AN_L')
return AN_L', *conformance_rate[]*
End

After constructing the network AN_L', the study of conformity and structural comparison are conducted between two networks AN_L and AN_L' for different layers. Regarding that structural comparison of two networks AN_L and AN_L' is considered and two sets $V_{AN_L'}$ and V_{AN_L} are equal together. Therefore two sets $E_{AN_L'}$ and E_{AN_L} must be compared with each other. This operation accomplishes by the P06 in figure 1 and the function *Calculate_Conf_Rate* in algorithm 1. The metrics used to measure the conformity or similarity of these two sets are Jaccard and Sørensen-Dice coefficient. The Jaccard index, also known as the Jaccard similarity coefficient, is a statistic to compare the similarity and diversity of sample sets. The Jaccard coefficient measures the similarity among finite sample sets (for example *X* and *Y*), and it is defined by the size of the intersection's sample sets divided by the size of the union of the sample sets (equation3). This metric is in the interval [0, 1] [42], [78].

$$J(X,Y) = |X \cap Y| / |X \cup Y| \tag{4}$$

The Sorensen-Dice index is a statistic for comparing the similarity of two sample sets. It was developed independently by the botanists Thorvald Sorensen [76] and Lee R. Dice [26], published in 1948 and 1945, respectively. Sorensen-Dice is called F1-score and Dice similarity coefficient (DSC). DSC is calculated based on the Equation 4. The DSC is a relative similarity in the range of [0, 1].

$$DSC(X,Y) = 2|X \cap Y| / |X| + |Y| \tag{5}$$

The value of similarity between networks AN_L and AN_L' is one value of the set X. Therefore, X is a set of values which determines the conformity status for two networks. SF03 in figure 4 is implemented through Python and networkX library.

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762020000200107

5.2 Evaluation Strategy

Conformance checking is conducted by comparison of primary default models with extracted models from real event logs. Since in structural conformance checking (SF02) the comparison is based on the value of predefined metrics and concepts of social network analysis, the evaluation is not necessary. But of course, behavioural conformance checking requires evaluation step. Although we have predicted the expected activity network (AN_L') based on the social network, our problem is not network prediction. Thus, the evaluation of our model is not performed by dividing the dataset into two sets train and test.

The network AN_L' is formed with the hypothesis of the effectiveness of social relations on the formation of business interactions, along with we compare the real network AN_L to AN_L'. Consequently, it should be ensured about the results. Thus, the Friedman test is applied to validate the results. We use this test to compare the results when activity network is built full randomly (CCRL) and when created based on the network SN (CCTC) to determine that formation of relations in the activity network is influenced by social relationships structure. The random activity network is called AN_L''. This network is constructed according to algorithm 2 and its conformity with real activity network AN_L is investigated.

Definition 6: $AN_L'' = (V_{AN_L''}, E_{AN_L''})$ is defined as a static and undirected network. $V_{AN_L''}$ is a set of human resources, and we have $V_{AN_L''} = V_{AN_L}$. $E_{AN_L''} \subseteq V_{AN_L''} \times V_{AN_L''}$ which is a set of random relations among the human resources based on the Erdos-Renyi model [62].

ALGORITHM 2: Conformance checking of network AN_L'' with AN_L
NAME: Conformance checking based on the random links (CCRL)

Input: network AN_L, L
Output: AN_L'' , *conformance_rate[]*
Initialize $E_{AN_L''} = \{\}$, $V_{AN_L''} = V_{AN_L}$, $edge_counter = 0$
Begin
$E_{AN_L''}$= *Add_Edge_Randomly*($V_{AN_L''}, edge_counter, |E_{AN_L}|, L$)
conformance_rate[] = *Calculate_Conf_Rate*(AN_L, AN_L'')
return AN_L'', *conformance_rate[]*
End

Given that our modeling is based on multi-layered networks, for more accurate evaluation, the results of structural and behavioural conformance checking will be presented for modeling activity network based on the simple network too. Therefore, we consider the simple activity network as follows.

Definition 7: according to the definition of simple networks [99], $AN_S = (V_{AN_S}, E_{AN_S})$ is defined as an activity network (or workflow network) which is one-layered, static, undirected, and weighted. V_{AN_S} is a set of human resources and we have $V_{AN_S} \subset V_{SN}$, in other words, $V_{SN} \cap R = V_{AN_S}$ and $V_{SN} \cap V_{AN_S} \neq \phi$. $E_{AN_S} \subseteq V_{AN_S} \times V_{AN_S}$ is a set of relations based on joint activities between human resources. For example, there is an edge between the two vertices i and j when i and j have at least one same activity, such as A_n, in the various executions of business processes. Therefore, in this network, unlike the network AN_L, the type of flow exchanged between human resources does not matter. The number of joint activities between i and j is defined as the weight of their relation.

Similarly, the social network corresponding to human resources in AN_S will also be defined as definition 8.

Definition 8: given the previous definitions, $SN_S' = (V_{SN_S'}, E_{SN_S'})$ is defined as a static, undirected and unweighted social network so that SN_S' is a sub-graph of SN. $V_{SN_S'} \subset V_{SN}$ is a set of individuals that includes individuals in the network AN_S and their friends. Similarly, $E_{SN_S'} \subseteq V_{SN_S'} \times V_{SN_S'}$ is a set of friendship relationships between individuals.

All the notations employed in the research with their descriptions are listed in table 2.

Table 2: Research notations and descriptions

ID	Notation	Description	ID	Notation	Description
1	A	Set of activities	17	i	Label of information flow
2	AN_L	Layered activity network	18	L	Set of flow types
3	AN_L'	Expected activity network	19	p	Probability of referring work from one resource to another based on social relations

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762020000200107

Table 2: continuation					
4	$AN_L^{''}$	Randomly activity network	20	R	Set of resources
5	AN_S	Simple activity network	21	SN	Social network
6	C	Set of Case IDs	22	SN_L'	Social networks of V_{AN_L}
7	$EventLog$	Event logs	23	SN_S'	Social networks of V_{AN_S}
8	E_{AN_L}	Edges of AN_L	24	T	Set of timestamps
9	$E_{AN_L'}$	Edges of AN_L'	25	V_{AN_L}	Vertices of AN_L
10	$E_{AN_L^{''}}$	Edges of $AN_L^{''}$	26	$V_{AN_L'}$	Vertices of AN_L'
11	E_{AN_S}	Edges of AN_S	27	$V_{AN_L^{''}}$	Vertices of $AN_L^{''}$
12	E_{SN}	Edges of SN	28	V_{AN_S}	Vertices of AN_S
13	$E_{SN_L'}$	Edges of SN_L'	29	V_{SN}	Vertices of SN
14	E_{SN_S}	Edges of SN_S	30	$V_{SN_L'}$	Vertices of SN_L'
15	f	Label of financial flow	31	V_{SN_S}	Vertices of SN_S
16	g	Label of goods flow	32	X	Set of conformance rates

6 Experimental Results and Evaluation

This section presents dataset descriptions and the results of structural and behavioural conformance checking. Moreover, we evaluate the results by the statistical analysis.

6.1 Dataset

Our research dataset is simulated based on user behaviour modeling of an Iranian ECSE called Digikala during the period October 1, 2016, to 20 November 20, 2016. Business processes and user behaviour are modeled and simulated by the WoPeD tool with regard to real data of the Digikala. Generating a dataset by simulating based on actual behaviours and patterns has two advantages in comparison with real dataset: 1) The actual data includes a set of specific behaviours in the data collection time period. Consequently, not all possible behaviours are included or some behaviours are very low-frequent. So by changing the simulation settings, you can create more behaviours and examine them, and 2) When the dataset is limited to reality, the results are limited to the same specific case study. Even the size of the dataset can have a small volume. Hence the results cannot be reliable and the answers will not be decisive. But when the dataset is obtained by simulation based on modeling the user behaviour, it includes different possible cases and can be generated sufficiently; thus, the results will be more reliable and more certain. In many other studies [13], [30], [54], [55], [56], [59], [60], [71], user behaviour modeling and simulation techniques have been used to provide an appropriate dataset when there is no sufficiently adequate dataset.

The dataset includes two parts: 1) DS01, the dataset of the social network SN. The characteristics of DS01 are given in Table 3, and 2) DS02, the dataset of the event logs $EventLog$ that its features have expressed in definition 2.

Table 3: Characteristics of dataset DS01 (network SN)

$\|V_{SN}\|$	$\|E_{SN}\|$	**Degree distribution**	**Graph density**	**Number of connected components**	**Edges type**
1033	11425	Power-law	0.0214	1	unweighted, undirected

To ensure the experimental results, the dataset DS02 has two samples (DS02-s1, DS02-s2). Each sample includes four versions. Each version consists of the execution of 1000 various instances (or cases) of some electronic purchasing processes. The process involves the structure of the loop. The number of different main procedures is over 16. The number of event logs' records in DS02-s1 and DS02-s2 is 6216 and 6115, respectively.

The activity network AN_L based on the joint activities and various flows can be extracted for each version of DS02. For each activity network AN_L, a corresponding social network SN_L' can be selected from the social network SN. The naming of the networks is as follows; for instance, S1-AN_i-v1 and S2-SN_g'-v2. The first one is based on the information interactions, and it is extracted from the first version of the first sample. The second name belongs to the social network of the people who are involved in the second version processes of the second sample and are related to the goods interactions. Furthermore, S1-AN_s-v1 is the extracted activity network from the first version of the first sample which includes all the flows.

6.2 Results of Structural Conformance Checking

Since we are interested in checking the structural conformity between two networks SN_L' and AN_L, we examine and compare these two networks from different aspects: structural features of the network, homophily, social role (popularity) and link dependency. The results are the answer to the research questions, especially Q1.

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762020000200107

6.2.1 Network Structural Characteristics

Table 4 indicates the structural features of various networks. Concerning vertex type, all the networks are one-mode. Additionally, regarding edge type, social networks SN_L' and SN_S' are unweighted and undirected. Whereas the activity networks AN_L and AN_S are weighted and undirected. For instance, figures 6 to 11 expose the graph of social networks and activity networks (based on the version v1 of sample S1) for each various network modeling method. Since the resources used in the layer f are not human resources, in this study, the networks SN_L' and SN_S' do not exist in the layer f. Therefore, networks SN_L' and SN_S' include the interactions of information and goods.

The networks SN_L' and SN_S' involves the people who have a role in networks AN_L and AN_S (the human resources in networks AN_L or AN_S, and their friends); therefore the number of vertices in networks SN_L' and SN_S' is more than networks AN_L and AN_S (table 4). According the table 4 and figures 6 to 11, the density of all the networks SN_L' and SN_S' is less than activity networks AN_L and AN_S. It means that the density of relationships in activity networks is higher since the people who have performed the same tasks are connected to each other. On the other hand, the tendency of people to connect with each other in activity networks is more than social networks and the clustering coefficient makes it clear. Thus, most of the components of activity networks are the complete graph and sometimes their GCC value is equal to 1.

Regarding the scenario of extracting activity networks and their un-connectivity, each component in networks AN_L and AN_S is representing one role in the business. Due to the human resources did not have multi-roles and the roles did not have joint activity, the graph is disconnected and contains separated components. The larger component in networks AN_i and AN_S is a component of the people who have had a role in the system as a customer. The number of components in networks AN_S is equivalent to the total number of components in the networks AN_i and AN_g. If the components of the activity network are connected, the community detection algorithms lead to the identification of roles. Furthermore, if resources are multi-roles, the overlapped community detection methods are employed to identify roles.

In order for summarizing, the degree distribution charts of social networks and activity networks with respect to different network modeling method for the version v1 of the first dataset sample S1 are demonstrated in figures 10 and 11. Degree distribution of vertices in social networks follows the power-law distribution while activity networks do not follow this distribution.

6.2.2 Homophily

In the activity networks, people are divided into two main groups: 1) the people who are human resources in the commercial processes, and 2) the people as customers who carry out some tasks through social commerce systems in the commercial processes. To check the link homophily, we examine two situations: 1) based on all the people in the activity network and social network and 2) based on the people are only human resources. Figures 12 to 14 and figures 15 to 17 respectively indicate the results.

Figures 12 to 14 (a) and figures 15 to 17 (a) indicate the probability of establishing a relation between two vertices in the activity network based on the number of their mutual friends in the social network. Figures 12 to 14 (b) and figures 15 to 17 (b) demonstrate the change of the probability of the forming a relationship between two nodes based on the number of mutual friends in the social networks and relative to the previous state in the chart. For instance, in figures 12 to 14, (a) AN_i, the linear charts for various versions of the dataset have overlap and low differences. In general, with the increase in the number of common neighbors between two individuals in network SN_L', the connection probability of these two individuals will increases, although the growth is at a low rate. Changes of connection probability are disclosed clearly in figures 12 to 14, (b) AN_i.

When people are only human resources, the results are different. For example, figures 15 to 17, (a) AN_i indicates the probability of building relationships among human resources does not have a regular trend with the increase in the number of mutual neighbors. For the value set of {2, 3, 4}, the probability of linking is relatively more than the other values in figures 15 to 17, (a) AN_i. Although the likelihood of developing a relationship for 9 common neighbors is high, the number of connections for this value is very small (e.g., the highest number is 2, which is 2/6 = 0.33). Changes of connection probability in the activity networks based on the information flow are presented in figures 12 to 14, (b) AN_i for the number of common friends.

Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 15 / ISSUE 2 / MAY 2020 / 93-121
© 2020 Universidad de Talca - Chile
This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762020000200107

Table 4: Structural characteristics of different activity networks and corresponded social networks

Modeling Method	Dataset	$\|V\|$	$\|E\|$	D	$\#CC$	GCC
Multi-layered network	S1-AN_i-v1	441	55664	0.5737	5	0.8799
	S1-SN_i'- v1	866	7121	0.0190	2	0.3595
	S2-AN_i- v1	464	58294	0.5427	5	0.8766
	S2-SN_i'- v1	946	8379	0.0187	2	0.4866
	S1-AN_i- v2	444	55677	0.5661	5	0.8807
	S1-SN_i'- v2	879	7320	0.0190	2	0.3873
	S2-AN_i- v2	465	58300	0.5404	5	0.8768
	S2-SN_i'- v2	948	8429	0.0188	2	0.4994
	S1-AN_i- v3	444	55683	0.5662	5	0.8820
	S1-SN_i'- v3	875	7290	0.0191	2	0.3830
	S2-AN_i- v3	468	58316	0.5336	5	0.8775
	S2-SN_i'- v3	952	8552	0.0190	2	0.4899
	S1-AN_i- v4	446	55692	0.5612	5	0.8821
	S1-SN_i'- v4	875	7289	0.0191	2	0.3839
	S2-AN_i- v4	467	58306	0.5358	5	0.8764
	S2-SN_i'- v4	952	8455	0.0187	2	0.4878
	S1-AN_g-v1	34	61	0.1087	8	1.0
	S1-SN_g'- v1	560	1347	0.0086	1	0.3565
	S2-AN_g- v1	34	66	0.1176	7	0.9248
	S2-SN_g'- v1	452	1182	0.116	1	0.3077
	S1-AN_g- v2	42	93	0.1080	8	1.0
	S1-SN_g'- v2	487	1221	0.0103	1	0.2763
	S2-AN_g- v2	34	63	0.1123	7	0.9832
	S2-SN_g'- v2	446	1141	0.0115	2	0.2233
	S1-AN_g- v3	39	82	0.1107	8	1.0
	S1-SN_g'- v3	506	1300	0.0102	1	0.2890
	S2-AN_g- v3	38	78	0.1110	8	1.0
	S2-SN_g'- v3	488	1298	0.0109	1	0.2665
	S1-AN_g- v4	39	81	0.1093	8	1.0
	S1-SN_g'- v4	513	1309	0.0010	1	0.2886
	S2-AN_g- v4	38	82	0.1166	7	0.9860
	S2-SN_g'- v4	507	1425	0.0111	2	0.2580
Simple network	S1-AN_S- v1	475	55725	0.4950	13	0.8885
	S1-SN'$_S$- v1	920	7922	0.0187	2	0.4254
	S2-AN_S- v1	498	58360	0.4716	12	0.8799
	S2-SN'$_S$- v1	970	8999	0.0191	2	0.4988
	S1-AN_S- v2	486	55770	0.4732	13	0.8910
	S1-SN'$_S$- v2	923	8032	0.0189	2	0.4252
	S2-AN_S- v2	499	58363	0.4697	12	0.8841
	S2-SN'$_S$- v2	970	9005	0.0192	2	0.5006
	S1-AN_S- v3	483	55765	0.4791	13	0.8915
	S1-SN'$_S$- v3	922	8038	0.0189	2	0.4300
	S2-AN_S- v3	506	58394	0.4570	13	0.8867
	S2-SN'$_S$- v3	975	9203	0.0194	2	0.5054
	S1-AN_S- v4	485	55773	0.4752	13	0.8916
	S1-SN'$_S$- v4	923	8047	0.0189	2	0.4306
	S2-AN_S- v4	505	58388	0.4588	12	0.8846
	S2-SN'$_S$- v4	975	9175	0.0193	2	0.5060

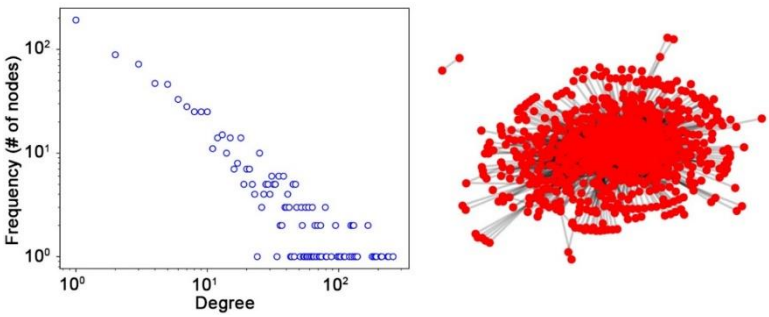


Figure 6: Graph and degree distribution chart of social network S1-SN_i'-v1

Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 15 / ISSUE 2 / MAY 2020 / 93-121
© 2020 Universidad de Talca - Chile

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762020000200107

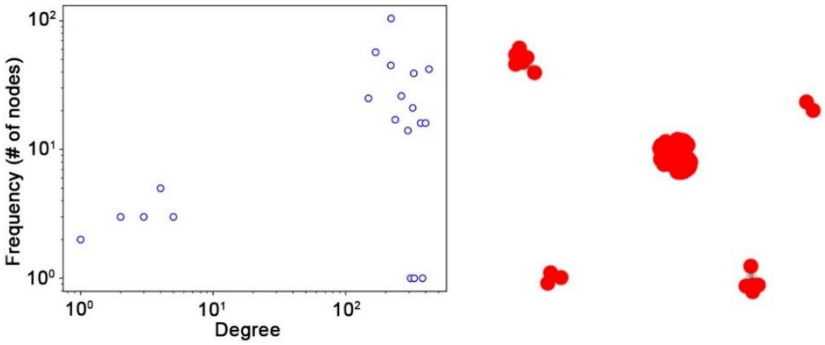


Figure 7: Graph and degree distribution chart of activity network S1-AN_i-v1

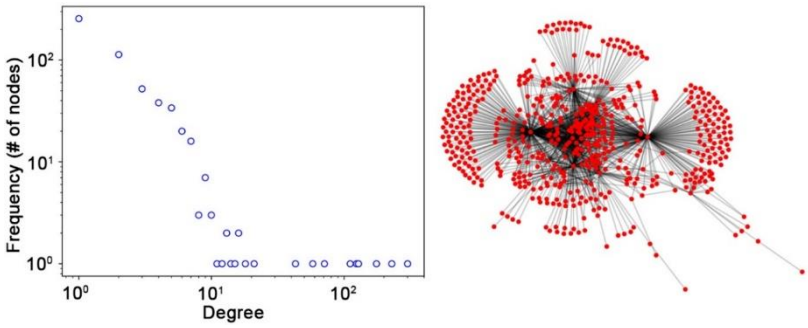


Figure 8: Graph and degree distribution chart of social network S1-SN_g'-v1

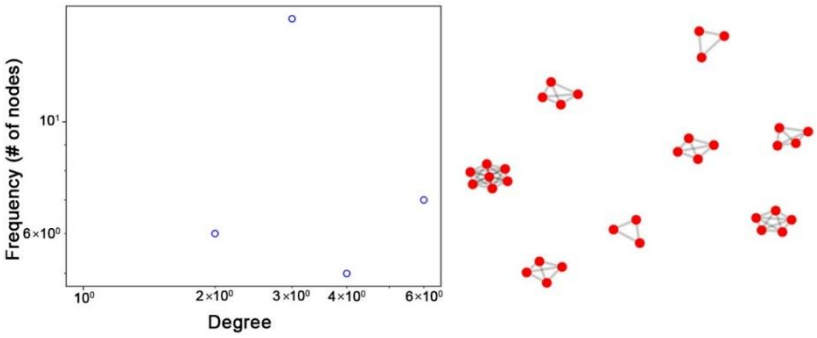


Figure 9: Graph and degree distribution chart of activity network S1-AN_g-v1

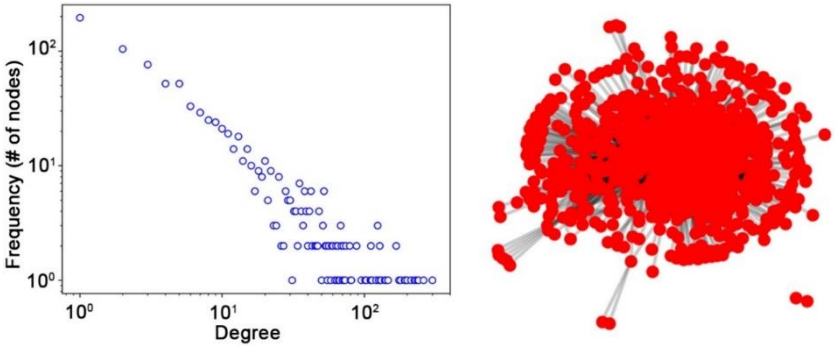


Figure 10: Graph and degree distribution chart of social network S1-SN'_S-v1

Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 15 / ISSUE 2 / MAY 2020 / 93-121
© 2020 Universidad de Talca - Chile
This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762020000200107

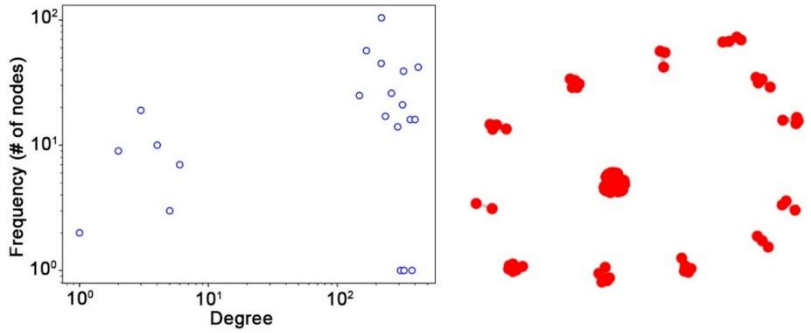


Figure 11: Graph and degree distribution chart of activity network S1-AN_S-v1

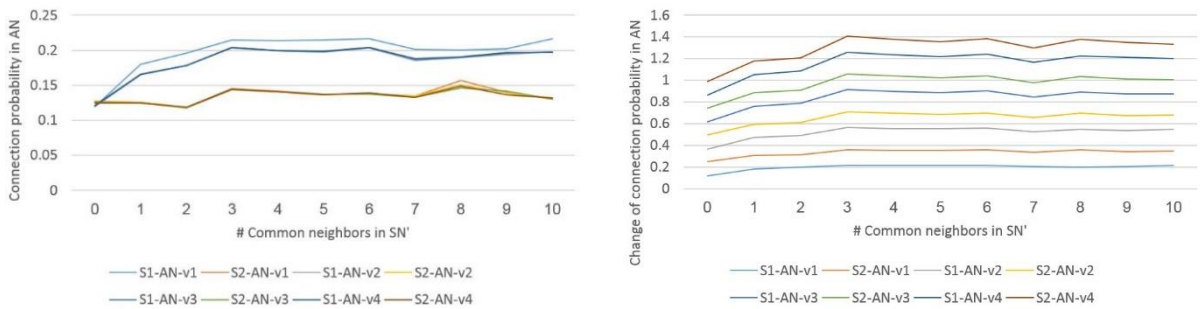
6.2.3 Popularity

The popularity status of connected individuals is presented in figure 9. As an instance, the figure 9, (a) AN_i shows that two people who are ordinary in network SN_i' are connected to each other in network AN_i with the probability of 0.14 (14%). Regarding this study, the likelihood that the two popular nodes in network AN_i are connected is more than the time when one of them is popular. Probability values and its change pattern in the activity network AN_g based on the goods flow is different from AN_i. Due to the number of information flows in compare with goods flow are higher, the probability pattern in figure 9, (c) AN_S is similar to the pattern of activity network AN_i. The different behaviour of the goods flow based activity network is concealed among all the other flows.

Figure 10, (a) AN_i indicates the same examination for the individuals who are human resources. Like link homophily, the behaviour of human resources is different from all other people and does not follow a specific pattern. Except for two cases S1-v3 and S2-v3 in information flow, the connection probability of two popular nodes is less than the other cases. Of course, the number of links for these two exceptions is low. Figure 10, (b) AN_i proves this reality. As an example, in version S1-v3, among all the links of human resource in network AN_i, just 5% belongs to the nodes which both of them are popular. Almost, most of the human resources' relations in the network AN_i belong to the nodes which are ordinary. As well, in figure 10, we observe the different behaviour of the human resources in the activity network which is based on the goods flow. According to figure 10, (b) AN_g, the possibility of the link between two human resources which one of them is a popular person, is more than the other statuses.

6.2.4 Link Dependency

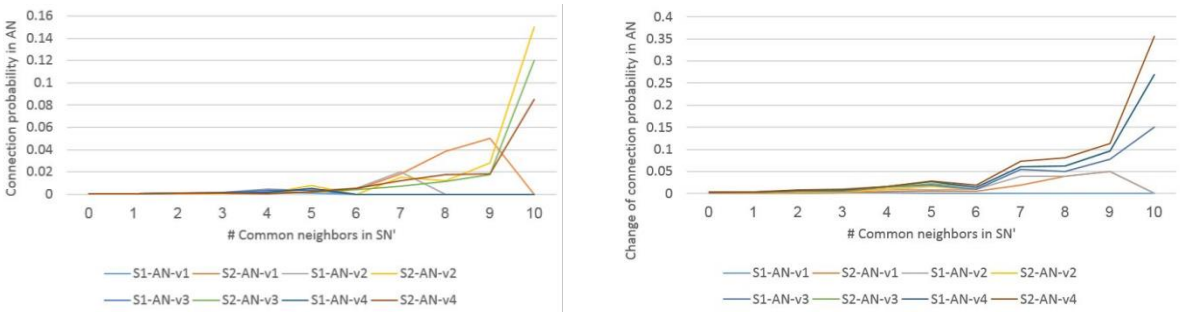
Is referring a case of work from one human resource to another one depending on their social relations? We studied it for different flows in event logs, and the results have been reported in table 5. Various versions of the dataset have different social link dependency. Version 4 (S1-v4 and S2-v4) has the most link dependency. Hence, it is interpreted that with the probability of p, human resource r_i refers work to human resource r_j if r_i and r_j are connected in the social network. The value of p for the version S1-v3 based on the information flows is equivalent to 0.42.



(a) AN_i - All resources (b) AN_i - All resources (changes)

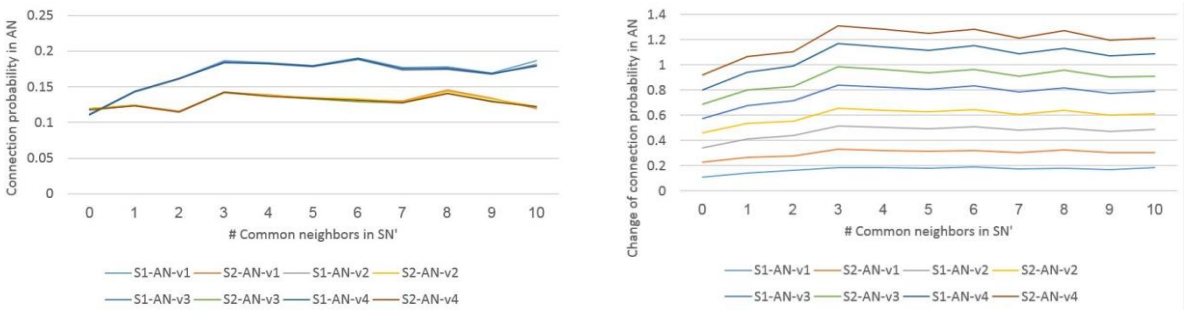
Figure 12: Homophily of link for all the people in the activity network AN_i based on the corresponded social network

Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 15 / ISSUE 2 / MAY 2020 / 93-121
© 2020 Universidad de Talca - Chile
This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762020000200107



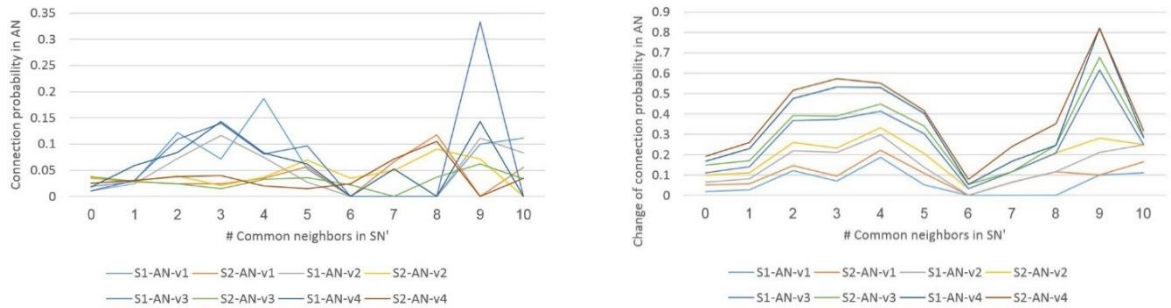
(a) AN_g - All resources (b) AN_g - All resources (changes)

Figure 13: Homophily of link for all the people in the activity network AN_g based on the corresponded social network



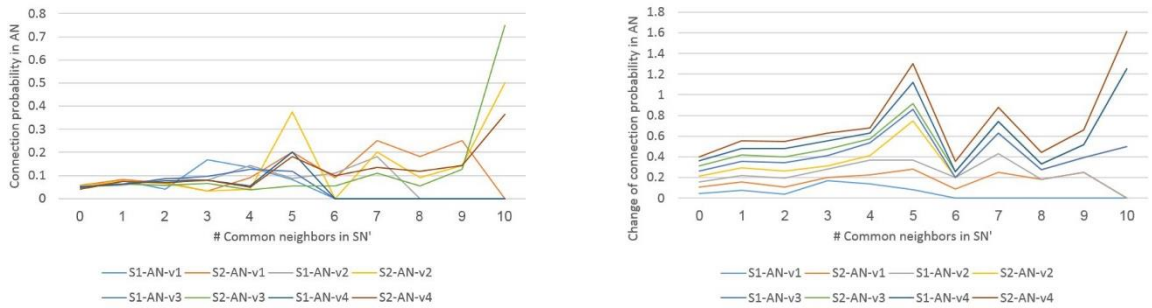
(a) AN_S - All resources (b) AN_S - All resources (changes)

Figure 14: Homophily of link for all the people in the activity network AN_S based on the corresponded social network



(a) AN_i - Human resources (b) AN_i - Human resources (changes)

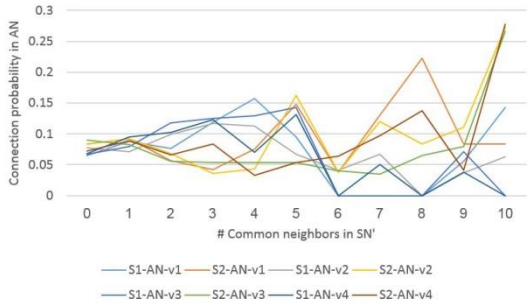
Figure 15: Homophily of link for all HRs in the activity network AN_i based on the corresponded social network



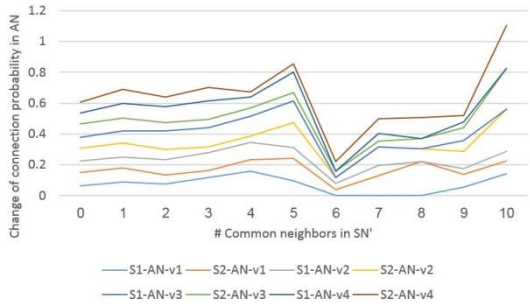
(a) AN_g - Human resources (b) AN_g - Human resources (changes)

Figure 16: Homophily of link for all HRs in the activity network AN_g based on the corresponded social network

Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 15 / ISSUE 2 / MAY 2020 / 93-121
© 2020 Universidad de Talca - Chile
This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762020000200107

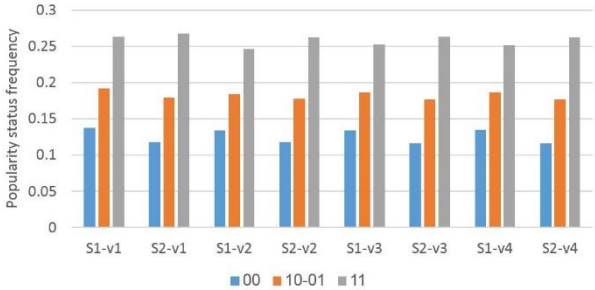


(a) AN_S - Human resources

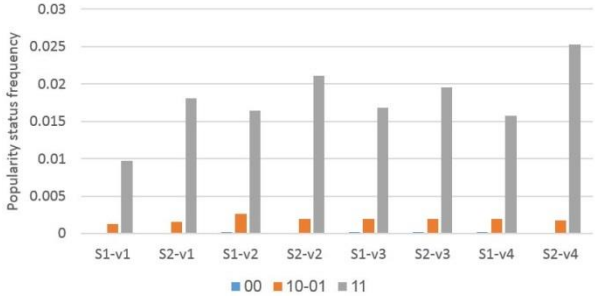


(b) AN_S - Human resources (changes)

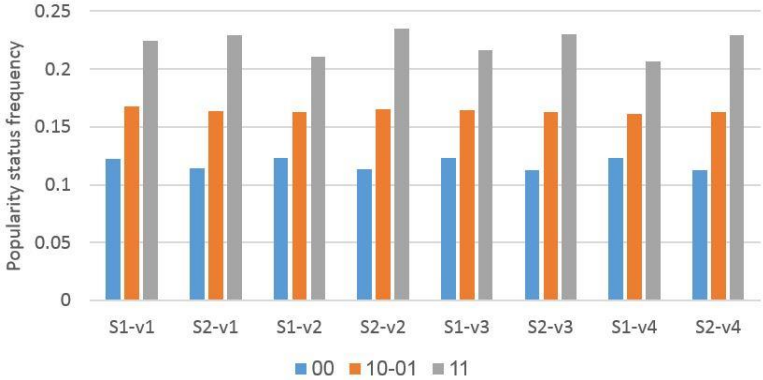
Figure 17: Homophily of link for all HRs in the activity network AN_S based on the corresponded social network



(a) AN_i

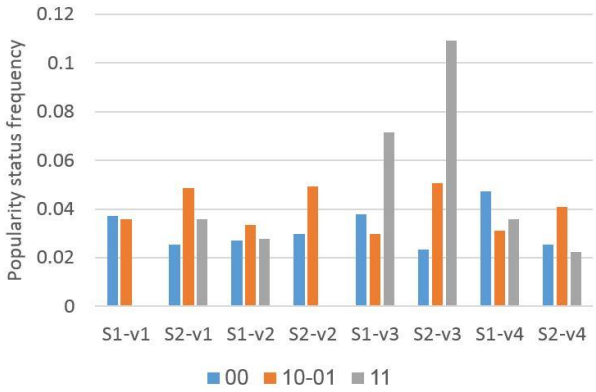


(b) AN_g

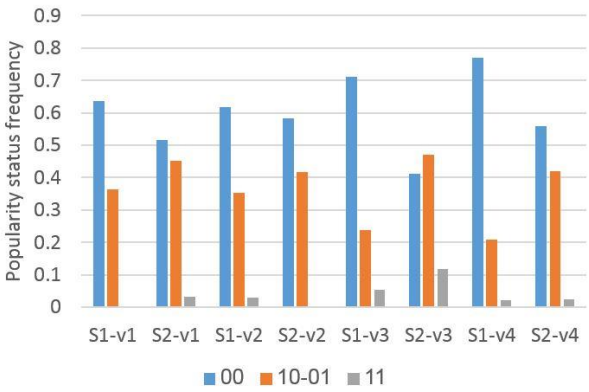


(c) AN_S

Figure 18: Popularity status of connected people in activity networks, relative to the corresponded social networks



(a) AN_i



(b) AN_i

Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 15 / ISSUE 2 / MAY 2020 / 93-121
© 2020 Universidad de Talca - Chile
This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762020000200107

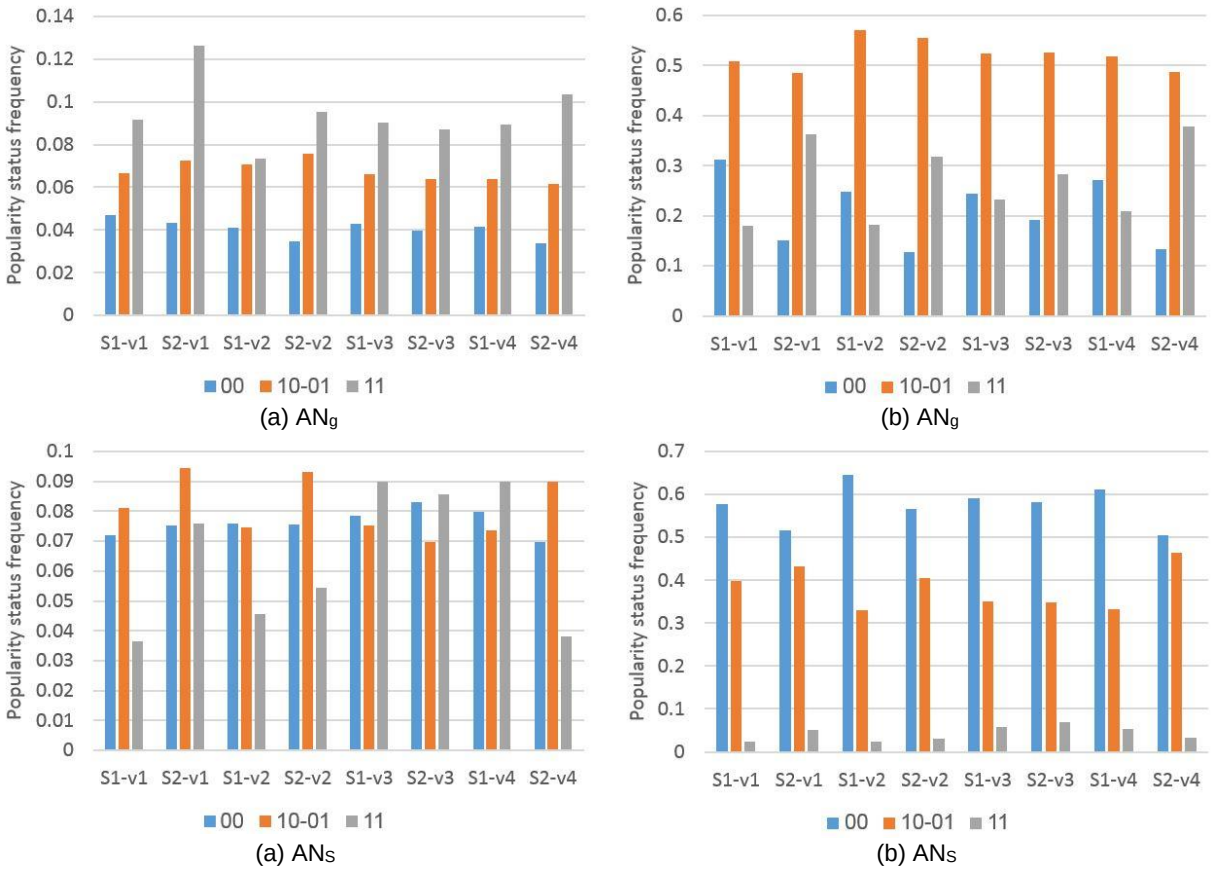


Figure 19: (a) Popularity status of connected HRs in the activity networks, relative to the corresponded social networks (b) Popularity frequency of connected HRs in the activity networks

Table 5: Link dependencies based on different flows for different versions of the dataset

Dataset	Link dependency for informational flow	Link dependency for goods flow	Link dependency regardless of flow type
S1-v1	0.08	0.093	0.086
S1-v2	0.24	0.14	0.17
S1-v3	0.42	0.26	0.30
S1-v4	0.49	0.35	0.39
S2-v1	0.04	0.12	0.09
S2-v2	0.11	0.21	0.17
S2-v3	0.25	0.35	0.30
S2-v4	0.41	0.54	0.48

6.2.5 Evaluation of Structural Conformance Checking

The set X for each activity network of human resources are listed in table 6. Each row is the values of set X for one activity network and its corresponding social network. Table 6 consists of values for modeling the activity network based on the multi-layered network and simple network. As can be noticed, different modeling leads to different values of set X. Multi-layered network modeling is more precise and represents phenomenally the reality in structural conformance checking. The difference of activity network based on the informational and goods flows are obvious in multi-layered network modeling, while it is not clear in simple network modeling. Due to the table 6, the answers of the first and second research questions are:

- The answer of Q1: Yes; According to the results of previous sections and table 6, the characteristics of networks are different. The networks of every flow type have various features.

- The answer of Q2: Yes; The reality is more revealed when the flows are separated. Thus, the results are more reliable by multi-layered network modeling.

Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 15 / ISSUE 2 / MAY 2020 / 93-121
© 2020 Universidad de Talca - Chile
This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762020000200107

6.3 Results of Behavioural Conformance Checking

Algorithms 1 and 2 are dependent on the stochastic theory. Hence, the results of each implementation of these algorithms may not be the same. Thus, the behavioural conformance checking results for an average of 1000 different implementations on the datasets S1 and S2 are reported in table 7. Algorithm 1 has two extensions, that is, CCTC1, and CCTC2. The edges between two popular nodes are ignored in CCTC1, while in CCTC2, they are not removed.

Regarding table 7, the conformance rate for network AN_L ($L = \{i, g\}$) with network AN_L' is in most cases more than the conformity of network AN_L with network AN_L''. Thus, it is concluded the relations in the activity network are affected by the social network's structure. Moreover, the conformance rate based on the CCTC2 is greater than the CCTC1 for the layers of goods and information. Especially, the value of the similarity index is greatest for CCTC2 in all cases at network AN_g. It is justified by structural conformance checking results in section 5.1.

Table 6: The set X based on the structural conformance checker values for human resources

Method	Dataset	L	$\|V\|$	$\|E\|$	D	$\#CC$	GCC	Homophily	Popularity	Link Dependency
Multi-layered network modeling	S1- v1	*i*	No	No	No	No	No	No	No	0.08
	S1- v2		No	No	No	No	No	No	No	0.24
	S1- v3		No	No	No	No	No	No	No	0.42
	S1- v4		No	No	No	No	No	No	No	0.49
	S2- v1		No	No	No	No	No	No	No	0.04
	S2- v2		No	No	No	No	No	No	No	0.11
	S2- v3		No	No	No	No	No	No	Yes	0.25
	S2- v4		No	No	No	No	No	No	No	0.41
	S1- v1	*g*	No	No	No	No	No	No	Yes	0.093
	S1- v2		No	No	No	No	No	No	Yes	0.14
	S1- v3		No	No	No	No	No	No	Yes	0.26
	S1- v4		No	No	No	No	No	No	Yes	0.35
	S2- v1		No	No	No	No	No	No	Yes	0.12
	S2- v2		No	No	No	No	No	No	Yes	0.21
	S2- v3		No	No	No	No	No	No	Yes	0.35
	S2- v4		No	No	No	No	No	No	Yes	0.54
Simple network modeling	S1- v1	*i* and *g* (no difference)	No	No	No	No	No	No	No	0.086
	S1- v2		No	No	No	No	No	No	No	0.17
	S1- v3		No	No	No	No	No	No	No	0.30
	S1- v4		No	No	No	No	No	No	No	0.39
	S2- v1		No	No	No	No	No	No	No	0.09
	S2- v2		No	No	No	No	No	No	No	0.17
	S2- v3		No	No	No	No	No	No	No	0.30
	S2- v4		No	No	No	No	No	No	No	0.48

Owing to the results of structural conformance checking, it was predictable that the actual behaviour of human resources would be hidden if the different flows as the network layers were not separated from each other. Hence, the CCTC1 and CCTC2 algorithms did not obtain better conformance rate in comparison with CCRL when the activity network is modeled based on the simple network. Thus, multi-layered network modeling in the context of s-commerce systems is more accurate and suitable for mining the comercial processes and organizations. Every row of each comparison in table 7 is equivalent one value of set X. It is represented that CCTC2 has the best values.

6.3.1 Evaluation of Behavioural Conformance Checking

The Friedman test is a non-parametric statistical hypothesis test. It is applied to investigate whether there is a meaningful statistical difference among the results, and to compare and validate the performance of algorithms [65]. Due to stochastic and random characteristics of CCTC1, CCTC2, and CCRL algorithms, this statistical test is required. Therefore, this test is performed between the CCTC algorithms and the other one (CCRL). The null hypothesis in the Friedman test (H_0) states the equality of medians among algorithms, whereas the alternative hypothesis (H_1) shows the difference. Significance level (α) illustrates the probability of the null hypothesis rejection while it is true. If P-value were less than the significance level, H_0 would be rejected. Further details on nonparametric statistical tests are found in the study of Derrac [25].

Calculation of ranking is the first step in the Friedman test. Average ranking is applied to rank each algorithm, and equal values are given to the average ranks [65]. Table 8 shows the ranks for each of the employed algorithms. Due to the ranks for both Jaccard and DSC indexes are equal, we reported both of them in one table. As can be observed, the algorithm CCTC2 has provided the first rank for the activity networks based on the information and goods flows.

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762020000200107

Besides, algorithm CCTC1 has obtained the second rank for information flow based activity network. The algorithm CCRL has the first rank in the activity network based on the simple network modeling. In other words, when there is no distinction among flows, CCRL has gained the better rank.

The P-value and Chi-square values of Friedman's test are listed in table 9. From the Chi-square distribution table, the critical value for $(3-1)=2$ degree of freedom with a 0.05 significance level is 5.99. Regarding table 9, Chi-square values for all datasets and activity networks are higher than the critical value. It means that H_0 is rejected while H_1 is accepted in all cases. Furthermore, the P-value is very small for all experiments, which affirms the rejection of H_0. It demonstrates a significant difference in the behaviour of the algorithms. Table 9 confirms that there is a statistically significant difference among the algorithms CCTC1, CCTC2, and CCRL. Moreover, our proposed algorithm results are not obtained by chance, since the gained P-value is less than 0.05 in all cases, except for versions s2-v1 for the networks AN_i. Due to the probability value is very low ($p=0.04$) in version *S2-v1* for AN_i, the Friedman test is rejected. Whereas the probability of linking two nodes per equation 1 is 0.0016, which is very low. Therefore, our study hypothesis is proved.

Table 7: Similarity for an average of 1000 implementations

Method	**Dataset**	***L***	***p***	**CCRL (AN_L vs. AN_L'')**		**CCTC1 (AN_L vs. AN_L')**		**CCTC2 (AN_L vs. AN_L')**	
				Jaccard index	DSC index	Jaccard index	DSC index	Jaccard index	DSC index
Multi-layered network modeing	*S1-v1*	*i*	0.08	0.0204	0.039	0.0232	0.0443	0.0202	0.0387
	S1-v2		0.24	0.0159	0.0308	0.0187	0.0362	0.0194	0.0377
	S1-v3		0.42	0.0206	0.0401	0.0259	0.0502	0.0281	0.0543
	S1-v4		0.49	0.0254	0.0492	0.0238	0.0462	0.0316	0.0609
	S2-v1		0.04	0.0162	0.0319	0.0166	0.0325	0.0181	0.0356
	S2-v2		0.11	0.0171	0.0334	0.0183	0.0356	0.0167	0.0326
	S2-v3		0.25	0.0212	0.0412	0.0199	0.0388	0.0275	0.0531
	S2-v4		0.41	0.0151	0.0292	0.01	0.0194	0.0104	0.0201
	S1-v1	*g*	0.093	0.0324	0.0623	0.028	0.054	0.0335	0.0642
	S1-v2		0.14	0.0276	0.0537	0.025	0.049	0.0277	0.0538
	S1-v3		0.26	0.0284	0.055	0.0231	0.0451	0.03	0.0581
	S1-v4		0.35	0.0264	0.0512	0.0165	0.0321	0.0256	0.0497
	S2-v1		0.12	0.0315	0.0611	0.0023	0.0044	0.0315	0.0611
	S2-v2		0.21	0.0254	0.0494	0.007	0.0139	0.0239	0.0466
	S2-v3		0.35	0.0235	0.0456	0.0099	0.0199	0.0196	0.0383
	S2-v4		0.54	0.03	0.058	0.0097	0.0192	0.0335	0.0647
Simple network modeling	*S1-v1*	*i* and *g* (no difference)	0.086	0.0429	0.0817	0.0468	0.0888	0.044	0.0837
	S1-v2		0.17	0.039	0.0747	0.0394	0.0754	0.0373	0.0716
	S1-v3		0.30	0.0428	0.082	0.04	0.0767	0.0404	0.0775
	S1-v4		0.39	0.0434	0.083	0.0337	0.065	0.0359	0.0692
	S2-v1		0.09	0.042	0.0803	0.044	0.0841	0.041	0.0787
	S2-v2		0.17	0.0368	0.0707	0.0391	0.0749	0.0364	0.07
	S2-v3		0.30	0.0387	0.0744	0.0357	0.0687	0.0374	0.072
	S2-v4		0.48	0.0418	0.08	0.0361	0.07	0.0341	0.0658

Table 10 displays the cases in which the CCTC2 is better, worse, or equal to other algorithms. The algorithm CCTC2 is performed better than other ones in most of the cases, except for network AN_S. Since there is no difference among flows in network AN_S, the reality of human resources' behaviour could not be observed. Therefore, it infers that the CCTC2 has not gained better performance for the activity networks which is modeled based on the simple network.

Table 8: Friedman's ranking based on the Jaccard and DSC indexes for algorithms CCTC and CCRL

L / **Dataset**	**Multi-layered network modeling**						**Simple network modeling**		
	I			*g*			*i* and *g* (no difference)		
	CCRL	CCTC1	CCTC2	CCRL	CCTC1	CCTC2	CCRL	CCTC1	CCTC2
S1-v1	2	1	3	2	3	1	3	1	2
S1-v2	3	2	1	2	3	1	2	1	3
S1-v3	3	2	1	2	3	1	1	3	2
S1-v4	2	3	1	1	3	2	1	3	2
S2-v1	3	2	1	1.5	3	1.5	2	1	3
S2-v2	2	1	3	1	3	2	2	1	3
S2-v3	2	3	1	1	3	2	1	3	2
S2-v4	1	3	2	2	3	1	1	2	3
Average rank	2.25	2.125	1.625	1.5625	3	1.4375	1.625	1.875	2.5
Overall rank	3	2	1	2	3	1	1	2	3

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762020000200107

Table 9: Friedman's test results for different versions of datasets and activity networks

Method	L	Dataset	p	Chi-square value	P-value
Multi-layered network modeling	*i*	*S1-v1*	0.08	13.2833	0.0013
		S1-v2	0.24	22.2107	1.5032E-05
		S1-v3	0.42	83.0979	9.0264E-19
		S1-v4	0.49	117.8519	2.5632E-26
		*S2-v1**	0.04	5.9663	0.0506
		S2-v2	0.11	6.7355	0.0345
		S2-v3	0.25	118.9734	1.463E-26
		S2-v4	0.41	88.4692	6.1541E-20
	g	*S1-v1*	0.093	48.8602	2.4556E-11
		S1-v2	0.14	27.1794	1.2533E-06
		S1-v3	0.26	133.2546	1.1592E-29
		S1-v4	0.35	361.4774	3.2075E-79
		S2-v1	0.12	1478.8596	0.0
		S2-v2	0.21	986.1177	7.3663E-215
		S2-v3	0.35	680.7374	1.5127E-148
		S2-v4	0.54	1327.6226	5.1336E-289
Simple network modeling	*i* and *g* (no difference)	*S1-v1*	0.086	34.8976	2.643E-08
		S1-v2	0.17	15.7048	0.0004
		S1-v3	0.30	19.6576	5.3878E-05
		S1-v4	0.39	295.8422	5.7368E-65
		S2-v1	0.09	25.8546	2.4308E-06
		S2-v2	0.17	24.9685	3.7858E-06
		S2-v3	0.30	32.6788	8.0145E-08
		S2-v4	0.48	191.6073	2.4717E-42

Table 10: Pairwise comparison between algorithm CCTC2 and other algorithms

Method	L	Comparison	Better	Worse	Equal
Multi-layered network modeling	*i*	CCTC2 vs. CCRL	5	3	0
		CCTC2 vs. CCTC1	6	2	0
	g	CCTC2 vs. CCRL	4	3	1
		CCTC2 vs. CCTC1	8	0	0
Simple network modeling	*i* and *g* (no difference)	CCTC2 vs. CCRL	0	8	0
		CCTC2 vs. CCTC1	3	5	0

7 Discussion

Conforming to the structural conformance checking results in the network structural features section, it proves that networks SN_L' and AN_L (for all types of flows) are different from each other. The network SN_L' has the degree distribution of power-law, while network AN does not follow this distribution and involves the components that most of them are fully connected. Each community (in our study, each component) represents one role in business.

What is important in the investigation of link homophily and popularity in all activity networks is the different behaviour of human resources relative to all the people. Moreover, although the probability of referring work between two connected human sources in the network SN_L' is dissimilar in different versions of the dataset; this difference and increasing trend are not seen in link homophily and popularity charts. The reason is that the probability p is based on the number of the referrers. Thus, if two human resources in the network AN_L have carried out several joint activities, there is a weighted edge between them and the weight is not considered in the charts. The other remarkable point is the different behaviour pattern in networks AN_i, AN_g and AN_S compared to each other. If we did not separate the various flows and did not observe them separately as a multi-layered network, this significant difference was not highlighted. Different types of flows (i.e., interaction types) should not be blindly merged. In fact, we gain this advantage through multi-layered network modeling and flow type distinction. We introduce the adding of flow type feature to event logs for the first time.

Despite these contradictions and differences, how could the conformity of the two networks SN_L' and AN_L be investigated for different flows? We solved this problem by extending the triadic closure concept in social networks theory. Actually, a behavioural conformance checking was accomplished. We assumed that social relations influence commercial interactions among human resources and activity network. Regarding this assumption, we proposed a model for forming the activity network based on social relations. Next, the similarity between real activity network and activity network based on our proposed model was investigated. The similarity of the generated activity network based on the triadic closure with the actual activity network is greater for each of the information and goods flows. In

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762020000200107

conclusion, according to the Friedman test, business interactions among human resources are affected by the structure of social relations and do not form by chance or randomly. Therefore, our assumption is right. When the multi-layered network modeling is not employed, the actual behaviour of human resources is hidden in data; as a result, the similarity of the random activity network with the actual activity network increases.

Although the frequency of edges between two popular nodes is not significant, the results of CCTC1 and CCTC2 are different. The edges between two popular nodes are not ignored in CCTC2, while in CCTC1, they are removed. Particularly in the goods flows, this difference is meaningful. Hence, retaining relations between two popular nodes is preferable, though their number is low.

8 Conclusion

The proof of the effect of social relations on business interactions formation among human resources in the context of social commerce was the main objective of this research. Following this goal, we assumed that if the formation of business interactions is based on social relations, then is there a meaningful relationship between the network of social relations and the activity network (the extracted network from business interactions among human resources)? To this end, conformance checking was the motivation of our research.

Hence, we tackled the issue of conformance checking in the context of social commerce systems that do not only include information flows. We formally defined the conformance checking problem based on graph theory and multi-layered networks. We also examined conformance from two structural and behavioural aspects. It has been shown that the separation of informational, goods and financial flows in the context of social commerce systems has led to the discovery of hidden behaviours in the system. If the flows are not separated (i.e., there are no distinctions among flows), the insight obtained from the system will not be accurate and the results will be incorrect. We also used social theories and the way a triadic closure relation is formed to check the conformance between the activity network and social network, moreover to examine the impact of the social network on the activity network. The results showed interesting phenomena.

8.1 Implications

The conformance checking compares the real execution of a business process with the original process model [16], [79], [92]. In fact, in the process model discovery approach, the process model discovered from the event log is compared with the initial model. A process model is a network of tasks with logical relations. In this research, we examined the conformance checking in the organizational mining approach. The initial model is the structure of the relations among individuals (i.e., human resources). The structure of the individuals' links can be official or unofficial. An official structure is one of a variety of organizational structures; while unofficial structures are based on social relationships, email, telephone calls, and so on. A network structure can extract from every official/unofficial structure among human resources through a well-defined scenario. Similarly, the model discovered from the event logs in the organizational mining approach can be the organizational structure extracted from the event logs. It can also be the activity networks extracted from the event logs. Therefore, in organizational mining, the networks contain human resource relations. In this study, the conformance checking of the social relations' network with the activity network of human resources was considered.

In recent years, due to the expansion of social networking applications, especially in a variety of businesses (e.g., social commerce), the study of social networks and activity networks is becoming increasingly important simultaneously. Due to the diverse flows in the social commerce systems and IoT-based systems in the future, the proposed framework can be of considerable use for studying users (i.e., human resources) behaviour. It is worth noting that the type of each exchanged flow among business resources should be specified and recorded in event logs. The type of exchanged flow among resources as contextual and semantic information is essential.

Among the important applications of the proposed framework are evaluating the performance of the organizational resources and identifying the individuals effective in the formation of business relationships and cooperation among individuals. Hereupon, studying the effect of unofficial relationships on the development of business relationships and the identification of the productive structure on the execution of a process rather than a formal or predefined structure are some of the applications of the proposed framework. Consequently owing to the influence of social relations on business interactions, the prediction and management of business markets are possible. It conveys a significant implication for business process managers, researchers, and practitioners interested in social commerce.

This study presents several managerial implications. First, the findings show that not only social relations affected on the customer behaviour, but also affected on the formation of the cooperation among human business resources. Therefore, improvement of social relations among human resources can influence the business process performance. Second, discovering the social structure that affects the activity network of the human resources can affect human resource management and business profitability. At last, according to the recent research by Chang [18] and Chau [19], the identification of individuals and social relations that influence the business process can facilitate and accelerate the prevention and detection of fraud and smuggling.

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762020000200107

8.2 Future Works

Applying network modeling, in particular, multi-layered networks and employing social network analysis methods, especially the link prediction, exposes a novel research direction in conformance checking and organizational mining. There are many potential future research directions for this work. First, if the human resources can play several roles in the process, then what features will have the activity network based on the joint activities? Second, if the activity network is extracted based on the handover of work scenario, then a directed network will be created. How will the conformance checking of the directed activity network be carried out with the undirected social network? Third, how do other social relations (e.g., messaging network, e-mail network, call network, etc.) form the business interactions (i.e., activity network)? Which one has more effect on the activity network? Finally, theorizing on why and how people build business interactions with each other based on social relationships in different kinds of commercial social networks is an intriguing direction for further research. Can the activity network be constructed from social relations? Furthermore, although there was no proper dataset during the research time, we intend to investigate real dataset to gain more trusted evaluations in the future. Eventually, we hope that the findings of this research can help researchers to identify new research questions and may open the door for future research.

References

[1] R. Accorsi and T. Stocker, On the exploitation of process mining for security audits: The conformance checking case, in Proceedings SAC ’12 of the 27th Annual ACM Symposium on Applied Computing, Trento, Italy, 2012, pp. 1709-1716.
[2] A. Adel Fares Gadelrab Mohamed, Process Mining Aapplication Considering the Organizational Perspective Using Social Network Analysis. Porto: Universidade do Porto, 2016.
[3] S. Alizadeh and A. Norani, ICMA: A new efficient algorithm for process model discovery, Applied Intelligence, vol. 48, no. 11, pp. 4497-4514, 2018.
[4] C. C. Alves, Social Network Analysis for Business Process Discovery, Lisbon, Technical University of Lisbon, 2010.
[5] A. Appice, Towards mining the organizational structure of a dynamic event scenario, Journal of Intelligent Information Systems, vol. 50, no. 1, pp. 165-193, 2018.
[6] A. Appice and D. Malerba, A co-training strategy for multiple view clustering in process mining, IEEE Transactions on Services Computing, vol. 9, no. 6, pp. 832-845, 2015.
[7] H. Ariouat, K. Barkaoui and J. Akoka, Improving process models discovery using AXOR clustering algorithm, presented at the International Conference on Information Science and Applications, Pattaya, in Lecture Notes in Electrical Engineering, vol. 339, Springer, February, 2015, pp. 623-629.
[8] L. Backstrom, P. Boldi, M. Rosa, J. Ugander, and S. Vigna, Four degrees of separation, in Proceedings of the 3rd Annual ACM Web Science Conference on - WebSci ’12, Evanston, Illinois, 2012, pp. 33-42.
[9] A. B. Badiru, Handbook of Industrial and Systems Engineering, Second Edition. Bosa Roca: CRC Press, Taylor & Francis Group, 2014.
[10] Y. Baghdadi, A framework for social commerce design, Information Systems, vol. 60, pp. 95-113, 2016.
[11] Y. Baghdadi, From e-commerce to social commerce: A framework to guide enabling cloud computing, Journal of Theoretical and Applied Electronic Commerce Research, vol. 8, no. 3, pp. 12-38, 2013.
[12] F. Battiston, J. Iacovacci, V. Nicosia, G. Bianconi, and V. Latora, Emergence of multiplex communities in collaboration networks, Plos one, vol. 11, no. 1, 2016.
[13] F. Bezerra and J. Wainer, Algorithms for anomaly detection of traces in logs of process aware information systems, Information Systems, vol. 38, no. 1, pp. 33-44, 2013.
[14] G. Bianconi, R. K. Darst, J. Iacovacci, and S. Fortunato, Triadic closure as a basic generating mechanism of communities in complex networks, Physical Review E, vol. 90, no. 4, 2014.
[15] H. Bossel, S. Klaczko-Ryndziun, and N. Müller, Eds., Systems theory in the social sciences: stochastic and control systems, pattern recognition, fuzzy analysis, simulation, behavioral models. Basel: Birkhäuser, 1976.
[16] A. Burattin, S. J. van Zelst, A. Armas-Cervantes, B. F. van Dongen, and J. Carmona, Online conformance checking using behavioural patterns, in Business Process Management, vol. 11080 (M. Weske, M. Montali, I. Weber, and J. vom Brocke, Eds.). Cham: Springer International Publishing, 2018, pp. 250-267.
[17] M. Cafarella, I. F. Ilyas, M. Kornacker, T. Kraska, and C. Re, Dark data: Are we solving the right problems?, in Proceedings the 2016 IEEE 32nd International Conference on Data Engineering (ICDE), Helsinki, Finland, 2016, pp. 1444-1445.
[18] Y.-C. Chang, K.-T. Lai, S.-C. T. Chou, and M.-S. Chen, Mining the networks of telecommunication fraud groups using social network analysis, in Proceedings of the 2017 IEEE/ACM ASONAM, Sydney, Australia, 2017, pp. 1128-1131.
[19] D. H. Chau and C. Faloutsos, Fraud Detection Using Social Network Analysis, a Case Study, in Encyclopedia of Social Network Analysis and Mining. New York, NY; Springer, 2014.
[20] A. Corbellini, S. Schiaffino and D. Godoy, Intelligent analysis of user interactions in a collaborative software engineering context, in Advances in New Technologies, Interactive Interfaces and Communicability, vol. 7547 (F. Cipolla-Ficarra, K. Veltman, D. Verber, M. Cipolla-Ficarra, and F. Kammüller, Eds.). Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 114-123.

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762020000200107

[21] T. D. Costa Teves, Social Commerce: Conceptual Model and Customer Perception, Master. Italy: Poltecnico Di Milano, 2013.
[22] G. Crispin, Stochastic Methods: A Handbook for the Natural and Social Sciences. Berlin Heidelberg: Springer-Verlag 2009.
[23] J. Cui, F. Wang and J. Zhai. (2010) Citation networks as a multi-layer graph: Link prediction and importance ranking. Snap Standford. [Online]. Available: http://snap.stanford.edu/class/cs224w-2010/proj2010/05_Project Report.pdf
[24] R. G. Curty and P. Zhang, Website features that gave rise to social commerce: a historical analysis, Electronic Commerce Research and Applications, vol. 12, no. 4, pp. 260-279, 2013.
[25] J. Derrac, S. García, D. Molina and F. Herrera, A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms, Swarm and Evolutionary Computation, vol. 1, no. 1, pp. 3-18, 2011.
[26] L. R. Dice, Measures of the amount of ecologic association between species, Ecology, vol. 26, no. 3, pp. 297-302, 1945.
[27] D. Easley and J. Kleinberg, Networks, Crowds, and Markets. Cambridge University Press, 2010.
[28] Z. Ebadi Abouzar, L. Esmaeili and S. A. Hashemi G., Centrality measures analysis in overlapped communities: An empirical study, presented at the 8th International Symposium on Telecommunications, Tehran, September, 2016.
[29] L. Esmaeili and S. A. Hashemi, A Systematic review on social commerce, Journal of Strategic Marketing, vol. 27, no. 4, pp. 317-355, 2019.
[30] L. Esmaeili and S. A. Hashemi, Rural intelligent public transportation system design: Applying the design for re-engineering of transportation e-commerce system in Iran, International Journal of Information Technologies and Systems Approach, vol. 8, no. 1, pp. 1-27, 2015.
[31] D. R. Ferreira and C. Alves, Discovering user communities in large event logs, presented at International Conference on Business Process Management, Clermont-Ferrand, in Lecture Notes in Business Information Processing, vol. 99, Springer, 29 August - 2 September, 2012, pp. 123-134.
[32] S. A. Golder and S. Yardi, Structural predictors of tie formation in twitter: Transitivity and mutuality, Proceedings of the 2nd IEEE International Conference on Social Computing, Minneapolis, August, 2010, pp. 88-95.
[33] K. Goldstein, Prepare for Social Commerce, Direct Marketing News, 2006. [Online]. Available: http://www.dmnews.com/prepare-for-social-commerce/article/93985.#/
[34] J. Gorner, J. Zhang and R. Cohen, Improving trust modeling through the limit of advisor network size and use of referrals, Electronic Commerce Research and Applications, vol. 12, pp. 112-123, 2013.
[35] J. A. Grochow and M. Kellis, Network Motif Discovery Using Subgraph Enumeration and Symmetry-Breaking, in Research in Computational Molecular Biology, vol. 4453, T. Speed and H. Huang, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 92-106.
[36] N. Hajli, A research framework for social commerce adoption, Information Management & Computer Security, vol. 21, no. 3, pp. 144-54, 2012.
[37] H. Han and S. Trimi, Social Commerce Design: A Framework and Application, Journal of theoretical and applied electronic commerce research, vol. 12, no. 3, pp. 50-68, 2017.
[38] D. J. Hruschka and J. Henrich, Friendship, cliquishness, and the emergence of cooperation, Journal of Theoretical Biology, vol. 239, no. 1, pp. 1-15, 2006.
[39] Z. Huang and M. Benyoucef, From e-commerce to social commerce: A close look at design features, Electronic Commerce Research and Applications, vol. 12, pp. 246-259, 2013.
[40] H. Huang, Y. Dong, J. Tang, H. Yang, N. V. Chawla, and X. Fu, Will Triadic Closure Strengthen Ties in Social Networks?, ACM Transactions on Knowledge Discovery from Data, vol. 12, no. 3, pp. 1-25, 2018.
[41] IBM. (2009) Social commerce defined. Digital Intelligence Today. [Online]. Available: http://digitalintelligence today.com/documents/IBM2009.pd f
[42] P. Jaccard, The distribution of the flora in the alpine zone, New Phytologist, vol. 11, no. 2, pp. 37-50, 1912.
[43] H. Jeong, H. Kim, and K. P. Kim, Betweenness Centralization Analysis Formalisms on Workflow-Supported Org-Social Networks, presented at the 16th International Conference on Advanced Communication Technology, 2014, pp. 1168-1172.
[44] J. Jiang et al., Understanding latent interactions in online social networks, ACM Transactions on the Web, vol. 7, no. 4, pp. 1-39, 2013.
[45] Y. Kavurucu, A comparative study on network motif discovery algorithms, International Journal of Data Mining and Bioinformatics, vol. 11, no. 2, p. 180, 2015.
[46] T. Kawamoto, A stochastic model of tweet diffusion on the Twitter network, Physica A: Statistical Mechanics and its Applications, vol. 392, no. 16, pp. 3470-3475, Aug. 2013.
[47] P. Kazienko, K. Musial, E. Kukla, T. Kajdanowicz, and P. Bródka, Multidimensional Social Network: Model and Analysis, in Lecture Notes in Computer Science, Lecture Notes in Artificial Intelligence, Springer, 2011, pp. 378-387.
[48] E. Khadangi, A. Bagheri, and A. Zarean, Empirical analysis of structural properties, macroscopic and microscopic evolution of various Facebook activity networks, Quality & Quantity, vol. 52, no. 1, pp. 249-275, 2018.
[49] S. Khakabimamaghani, I. Sharafuddin, N. Dichter, I. Koch, and A. Masoudi-Nejad, QuateXelero: An Accelerated Exact Network Motif Detection Algorithm, PLoS ONE, vol. 8, no. 7, p. e68073, Jul. 2013.
[50] P. F. Lazarsfeld and R. K. Merton, Friendship as a social process: A substantive and methodological analysis, Freedom and control in modern society, vol. 18, pp. 18-66, 1954.

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762020000200107

[51] W. L. J. Lee, H. M. W. Verbeek, J. Munoz-Gama, W. M. P. van der Aalst, and M. Sepúlveda, Recomposing conformance: Closing the circle on decomposed alignment-based conformance checking in process mining, Information Sciences, vol. 466, pp. 55-91, 2018.
[52] T.-P. Liang and E. Turban, Introduction to the special issue - Social Commerce: A Research Framework for Social Commerce, International Journal of Electronic Commerce, vol. 16, no. 2, pp. 5-13, Winter -2012 2011.
[53] R. S. Mans, M. H. Schonenberg, M. Song, W. M. P. van der Aalst, and P. J. M. Bakker, Application of Process Mining in Healthcare - A Case Study in a Dutch Hospital, in Communications in Computer and Information Science, 2008, vol. 25, pp. 425-438.
[54] S. Mardani, M. K. Akbari, and S. Sharifian, Fraud detection in Process Aware Information systems using MapReduce, in 2014 6th Conference on Information and Knowledge Technology (IKT), Shahrood, Iran, 2014, pp. 88-91.
[55] S. Mardani and H. R. Shahriari, A new method for occupational fraud detection in process aware information systems, in 2013 10th International ISC Conference on Information Security and Cryptology (ISCISC), Yazd, Iran, 2013, pp. 1-5.
[56] K. Mark and L. Csaba, Analyzing Customer Behaviour Model Graph (CBMG) using Markov Chains, in Intelligent Engineering Systems, 2007 International Conference on, Hotel Griff, 2007, pp. 71-76.
[57] P. Marsden, Commerce gets social: How your networks are driving what you buy, Digital Intelligence Today, 06-Jan-2011. [Online]. Available: http://digitalintelligencetoday.com/speed-summary-wired-feb-2011-cover-story-on-social-commerce./
[58] M. A. May and L. W. Doob, Competition and cooperation, vol. 25. Social science research council, 1937.
[59] D. A. Menascé, V. A. . Almeida, R. Fonseca, and M. A. Mendes, Business-oriented resource management policies for e-commerce servers, Performance Evaluation, vol. 42, no. 2-3, pp. 223-239, Sep. 2000.
[60] D. A. Menascé, V. A. F. Almeida, R. Fonseca, and M. A. Mendes, A methodology for workload characterization of E-commerce sites, in Proceedings of the 1st ACM conference on Electronic commerce - EC '99, Denver, Colorado, United States, 1999, pp. 119-128.
[61] P. Mikalef, M. Giannakos, and A. Pateli, Shopping and Word-of-Mouth Intentions on Social Media, Journal of theoretical and applied electronic commerce research, vol. 8, no. 1, pp. 5-6, 2013.
[62] R. Milo, N. Kashtan, S. Itzkovitz, M. E. J. Newman, and U. Alon, On the uniform generation of random graphs with prescribed degree sequences, arXiv:cond-mat/0312028, Dec. 2003.
[63] S. Molinillo, F. Liébana-Cabanillas, and R. Anaya-Sánchez, A Social Commerce Intention Model for Traditional E-Commerce Sites, Journal of Theoretical and Applied Electronic Commerce Research, vol. 13, no. 2, pp. 80-93, 2018.
[64] S. J. Mousavirad and H. Ebrahimpour-Komleh, Human mental search: a new population-based metaheuristic optimization algorithm, Applied Intelligence, vol. 47, no. 3, pp. 850-887, Oct. 2017.
[65] S. J. Mousavirad and H. Ebrahimpour-Komleh, Multilevel image thresholding using entropy of histogram and recently developed population-based metaheuristic algorithms, Evolutionary Intelligence, vol. 10, no. 1-2, pp. 45-75, 2017.
[66] S. Omidi, F. Schreiber, and A. Masoudi-Nejad, MODA: an efficient algorithm for network motif discovery in biological networks, Genes Genet. Syst., vol. 84, no. 5, pp. 385-395, 2009.
[67] S. Oskamp and D. Perlman, Effects of friendship and disliking on cooperation in a mixed-motive game, Journal of Conflict Resolution, vol. 10, no. 2, pp. 221-226, 1966.
[68] S. M. Rahman and M. S. Raisinghani, Electronic Commerce: Opportunity and Challenges. USA: Idea Group Publishing, 2000.
[69] P. Ribeiro and F. Silva, g-tries: an efficient data structure for discovering network motifs, in Proceedings of the 2010 ACM Symposium on Applied Computing - SAC '10, Sierre, Switzerland, 2010, p. 1559.
[70] A. Rozinat and W. M. P. van der Aalst, Conformance Checking of Processes Based on Monitoring Real Behaviour, Information Systems, vol. 33, no. 1, pp. 64-95, 2008.
[71] G. Ruffo, R. Schifanella, M. Sereno, and R. Politi, WALTy: a user behaviour tailored tool for evaluating web application performance, in Third IEEE International Symposium on Network Computing and Applications, 2004. (NCA 2004). Boston, MA, USA, 2004, pp. 77-86.
[72] A. Shahmohammadi, E. Khadangi, and A. Bagheri, Presenting new collaborative link prediction methods for activity recommendation in Facebook, Neurocomputing, vol. 210, pp. 217-226, 2016.
[73] E. Shekat, M. Rahgozar, and M. Asadpour, Structural link prediction based on ant colony approach in social networks, Physica A, vol. 419, pp. 80-94, 2015.
[74] M. Shukla, S. Manjunath, R. Saxena, S. Mondal, and S. Lodha, POSTER: WinOver Enterprise Dark Data, in CCS '15 Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, Denver, Colorado, USA, 2015, pp. 1674-1676.
[75] M. Song and W. M. P. van der Aalst, Towards comprehensive support for organizational mining, Decision Support Systems, vol. 46, no. 1, pp. 300-317, 2008.
[76] T. Sorenson, A Method of Establishing Groups of Equal Amplitudes in Plant Sociology Based on Similarity of Species Content and Its Application to Analyses of the Vegetation on Danish Commons, Kongelige Danske Videnskabernes Selskab, Biologiske Skrifter, vol. 5, pp. 1-34, 1948.
[77] W. D. Sunindyo, T. Moser, D. Winkler, and S. Biffl, Process analysis and organizational mining in production automation systems engineering, 2010.
[78] P.-N. Tan, M. Steinbach, and V. Kumar, Introduction to Data Mining, Second. Pearson Addison-Wesley, 2005.
[79] N. Tax, X. Lu, N. Sidorova, D. Fahland, and W. M. P. van der Aalst, The imprecisions of precision measures in process mining, Information Processing Letters, vol. 135, pp. 1-8, 2018.

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762020000200107

[80] Trivedi and K. Gokulnath, Research on dark data analysis to reduce data complexity in big data, International education and research journal, vol. 3, no. 5, pp. 361-362, 2017.
[81] S. Uddin and M. J. Jacobson, Dynamics of email communications among university students throughout a semester, Computers & Education, vol. 64, pp. 95-103, 2013.
[82] A. Vajapeyajula, P. Radhakrishnan, and V. Varma, Survey of Social Commerce Research, in Mining Intelligence and Knowledge Exploration, vol. 9468, Springer International Publishing, 2016, pp. 493-503.
[83] W. M. P. van der Aalst, Process mining: data science in action, Second. Springer, 2016.
[84] W. M. P. van der Aalst, No knowledge without processes: process mining as a tool to find out what people and organizations really do, in KEOD, 2014.
[85] W. M. P. van der Aalst, Process Mining: Discovery, Conformance and Enhancement of Business Processes. Springer-Verlag Berlin Heidelberg, 2011.
[86] W. M. P. van der Aalst, Business alignment: using process mining as a tool for Delta analysis and conformance testing, Requirements Engineering, vol. 10, no. 3, pp. 198-211, 2005.
[87] W. M. P. van der Aalst et al., Process Mining Manifesto, in F. Daniel et al. (Eds.): BPM 2011 Workshops, Part I, LNBIP, 2012, vol. 99, pp. 169-194.
[88] W. M. P. van der Aalst and A. K. A. de Medeiros, Process Mining and Security: Detecting Anomalous Process Executions and Checking Process Conformance, Electronic Notes in Theoretical Computer Science, vol. 121, no. 4, pp. 3-21, 2005.
[89] W. M. P., van der Aalst, M. Dumas, C. Ouyang, A. Rozinat, and E. Verbeek, Conformance checking of service behaviour, ACM Transactions on Internet Technology, vol. 8, no. 3, 2008.
[90] W. M. P. van der Aalst, H. A. Reijers, and M. Song, Discovering Social Networks from Event Logs, Computer Supported Cooperative Work, vol. 14, no. 6, pp. 549-593, 2005.
[91] W. M. P. van der Aalst and M. Song, Mining social networks: Uncovering interaction patterns in business processes, in Business Process Management, 2004, pp. 244-260.
[92] S. J. van Zelst, A. Bolt, M. Hassani, B. F. van Dongen, and W. M. P. van der Aalst, Online conformance checking: relating event streams to process models using prefix-alignments, International Journal of Data Science and Analytics, 2017.
[93] W. Wang, X. Bai, F. Xia, T. M. Bekele, X. Su, and A. Tolba, From triadic closure to conference closure: the role of academic conferences in promoting scientific collaborations, Scientometrics, vol. 113, no. 1, pp. 177-193, 2017.
[94] P. Wang, B. Xu, Y. Wu, and X. Zhou, Link prediction in social networks: the state-of-the-art, Science China Information Sciences, vol. 58, no. 1, pp. 1-38, 2015.
[95] S. Wasserman and K. Faust, social network analysis: methods and application. Cambridge University Press, 1994.
[96] L. Wen, J. Wang, W. M. P. van der Aalst, B. Huang, and J. Sun, Mining Process Models with Prime Invisible Tasks, Data & Knowledge Engineering, vol. 69, no. 10, pp. 999-1021, 2010.
[97] L. Yao, L. Wang, L. Pan, and K. Yao, Link Prediction Based on Common-Neighbors for Dynamic Social Network, in Procedia Computer Science, 2016, vol. 83, pp. 82-89.
[98] S. Yeon Yoon, Empirical Investigation of Web 2.0 Technologies for Social Commerce and Implementation of Social App Prototypes, Master, University of Ottawa, Canada, 2013.
[99] M. Zhang, Social Network Analysis: History, Concepts, and Research, in Handbook of Social Network Technologies and Applications, Springer, Boston, MA, 2010, pp. 3-21.
[100] W. Zhao and X. Zhao, Process Mining from the Organizational Perspective, in 17th International Symposium, ISMIS, vol. 5722, J. Rauch, Z. W. Raś, P. Berka, and T. Elomaa, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014, pp. 701-708.
[101] M. Zignani, S. Gaito, G. P. Rossi, Z. Zhao, H. Zheng, and B. Y. Zhao, Link and Triadic Closure Delay : Temporal Metrics for Social Network Dynamics, presented at the 8th International AAAI Conference on Weblogs and Social Media, 2014, pp. 564-573.
[102] L. Zhou, P. Zhang, and H.-D. Zimmermann, Social commerce research: An integrated view, Electronic Commerce Research and Applications, vol. 12, pp. 61-68, 2013.