

Received February 6, 2021, accepted March 4, 2021, date of publication March 15, 2021, date of current version April 2, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3065925

# Organizational Closeness Centralities of Workflow-Supported Performer-to-Activity Affiliation Networks

HYUN AHN AND KWANGHOON PIO KIM<sup>ID</sup>

Data and Process Engineering Research Laboratory, Division of Computer Science and Engineering, Contents Convergence Software Research Institute, Kyonggi University, Suwon 16227, South Korea

Corresponding author: Kwanghoon Pio Kim (kwang@kgu.ac.kr)

This work was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education under Grant NRF-2020R1A6A1A03040583.

**ABSTRACT** A workflow model specifies execution sequences of the associated activities and their affiliated relationships with roles, performers, invoked-applications, and relevant data. These affiliated relationships exhibit a series of valuable human-centered organizational knowledge and are utilized for exploring human resource's work patterns. This paper focuses not only on a specific type of affiliated relationships between performers and activities, in particular, which forms a performer-to-activity affiliation network, but also on a specific type of analysis techniques, which builds a closeness centrality measurement approach for quantifying the degrees of farnesses between performers as well as between activities. In other words, this paper investigates a series of formal approaches for building organizational closeness centrality measurement techniques on the specific type of affiliation networks. The investigation mainly deploys two types of algorithmic formalisms along with an operational example, which are measuring performer-centered organizational closeness centralities and activity-centered organizational closeness centralities, respectively. In order to validate the deployed algorithmic equations, the paper carries out a couple of operational experiments; One is on an ICN-based workflow package and the other is on a discovered workflow model mined from a dataset of workflow event logs. Summarily, this paper devises a series of algorithms and equations for measuring closeness centralities of activities, verify the devised algorithms and their related equations along with operational examples, and discuss the ultimate implications of these analysis techniques of organizational closeness centrality measurements as the performer-to-activity affiliation networking knowledge in workflow-supported organizations.

**INDEX TERMS** Workflow process, performer-activity affiliations, information control net, organizational social network, workflow-supported affiliation network, closeness centrality, workflow intelligence, human resource management.

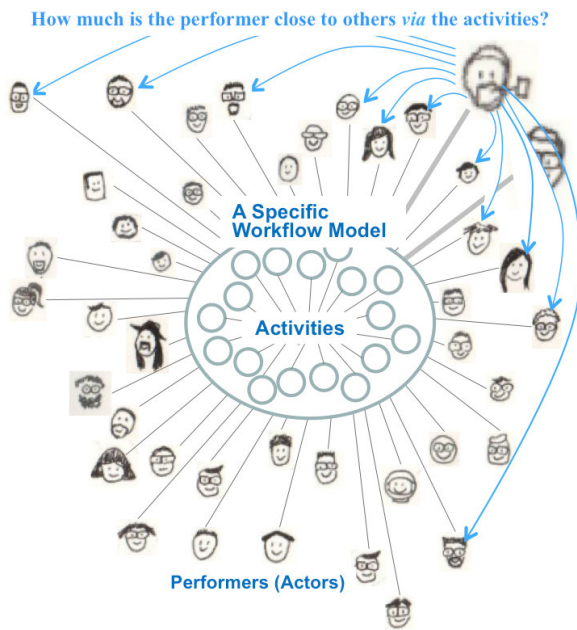
## I. INTRODUCTION

Recently, the workflow literature starts being interested in re-positioning workflow systems as a tool of business and organizational knowledge and intelligence. It begins from the strong belief that social relationships and collaborative behaviors among workflow-performers may affect the overall performance and being crowned with great successes in the real businesses and the working productivity as well. A typical outcome of those re-positioning works ought to be [1]–[8], in which the authors formalize a mechanism and its related algorithms to discover workflow-supported

affiliation networking knowledge from a conventional workflow model. Especially, the authors' research group had proposed the concept of workflow-supported affiliation networks and its closeness centrality analysis algorithm in [3] and [2], respectively. That is, in a workflow model the performers (or actors) are linked into activities through joint participation; conversely, the activities are connected to performers through joint involvement. A collection of these links and connections is formally and graphically represented by the concept of “*workflow-supported performer-to-activity affiliation network* [3].”

This paper focuses on quantitatively measuring the degrees of farnesses through the organizational closeness centralities of workflow-supported affiliation network models, each of

The associate editor coordinating the review of this manuscript and approving it for publication was Davide Aloini.



**FIGURE 1.** Degree of Farness: Conceptual Motivation of the Organizational Closeness Centrality.

which is formed by two key groups of the elements: a set of performers and a collection of activities. As shown in FIGURE 1, which is a conceptual definition of the organizational closeness centrality in a performer-to-activity affiliation network that is formed from a workflow model by the discovery algorithm proposed in [3], it is possibly done to uncover the relational structures (degree of work-intimacy) of workflow-performers through their joint involvement in activities, and to reveal the relational structures (degree of work-connectivity) of workflow-activities through their joint participation of common performers, as well, through measuring the degrees of farnesses. Thereafter, based upon the discovered affiliation network, this paper tries to quantify the degrees of farnesses through the organizational closeness centralities to measure how much the performers are close to others with reference to the workflow-supported activities. In more detail, it gives us the analytical performer-to-activity affiliation knowledge by finding the answers to the following questions:

- Degree of farness (work-intimacy) between workflow-performers: How quickly can a performer interact with others *via* very few intermediary activities in enacting workflow procedures?
- Degree of farness (work-connectivity) between workflow-activities: How closely can an activity connect to others *via* very few intermediary performers in enacting workflow procedures?

That is, this paper is basically concerned about organizational centralities [2], [7], [9]–[11] in a workflow-supported organization. The typical four out of the conventional centrality analysis techniques [12]–[14] are degree, closeness, betweenness, and eigenvector centralities, and the authors

are interested in the closeness centrality analysis technique, in particular. It is necessary to quantitatively measure the extents of closeness centralities between performers through activities as well as the extents of the closeness centralities between activities through performers in a specific workflow procedure. Also, the paper expatiates a mathematical representation of the affiliation relationships between workflow-performers and workflow-activities, *i.e.*, two-mode of dyadic graph and its bipartite matrix, and devises a series of algorithms for calculating the organizational closeness centralities of activities and performers on the corresponding bipartite matrix of a workflow-supported performer-to-activity affiliation network. Finally, the paper validates the devised algorithms through carrying out a couple of operational experiments on a group of synthetic workflow process models operated by 2 packages, 5 models, 50 activities and 16 performers and on a discovered workflow process model mined from a workflow event log dataset that was released from the Conformance Checking Challenge 2019 (CCC2019) that was a co-located event of the 1<sup>st</sup> international conference on process mining 2019.

In terms of the composition of the paper, the next section describes the literature survey result related to the topics of modeling, discovering, analyzing, and visualizing the workflow-supported organizational affiliation networks and social networks [15]. Two consecutive sections formally define the concept of workflow-supported organizational affiliation networks proposed in this paper and expatiate on its detailed formalisms not only for mathematically representing the workflow-supported performer-to-activity affiliation networks but also for measuring the closeness centralities of workflow-performers and workflow-activities, respectively. Finally, the final section carries out a couple of operational experiments to validate the practical applicability of the proposed organizational closeness centrality concept in workflow-supported organizations.

## II. RELATED WORKS

Recently, technology-supported social networks and organizational behavioral analytics issues [4]–[6], [8], [16]–[20] have been raised in the IT literature. Naturally, the workflow literature has just started transitioning into and focusing on social and collaborative work analyses in workflow-supported organizations, because workflow management systems are “human-centered systems [21],” where workflow procedures must be designed, deployed, and understood within their social and organizational contexts [22], [23], and they must be conveyed, facilitated, and navigated in large-scaled operational knowledge collections, at the same time. Individuals as employees in workflow-supported organizations have worked intra-organizationally [6] and inter-organizationally under the controls of workflow enactment systems. The research works from [5], [7], [8], [18], [24]–[26] suggested frameworks that enable to discover the human-centered organizational knowledge from a workflow-supported organization. From these

human-centered organizational contexts, it is possible for two concepts of organizational social networks among those individuals to be formed, in particular. One is the concept of workflow-supported social network [1], [7], [15], [27], the other is the concept of workflow-supported affiliation network [28]. The following are the descriptions of the pioneering works of these two concepts, respectively, which have been done in the workflow literature:

**Workflow-Supported Social Networks:** So far, the literature has delivered several research results about modeling, discovering, analyzing, and visualizing the workflow-supported social networks [15]. In terms of the discovery issue, there have been existing two main branches of research approaches in exploring organizational social networks from workflow-supported organizations; One is the discovery approach, and the other is the rediscovery approach. The latter is concerned with mining organizational social networks from enactment event logs of workflow procedures, which was firstly issued by Aalst [1]. The former is to discover organizational social networks through exploring the human-centered perspectives from workflow procedures themselves, which was issued at first by [7]. There are typical research and development outputs coping with the analysis and visualization issues in workflow-supported social networks.

Park *et al.* [10] built a theoretical approach for numerically analyzing closeness centrality measures among workflow-actors of workflow-supported social network models. The essential part of the proposed approach is a closeness centrality analysis equation and its algorithm that is able to efficiently compute the closeness centrality measures, and eventually the developed algorithm can be applied to analyzing the degree of work-intimacy among those workflow-actors who are allocated to perform the corresponding workflow procedure. Jeong, *et al.* [29] implemented a knowledge visualization framework and its system designated for the workflow-supported social networking knowledge, and the devised framework is pipelining from the XPDL<sup>1</sup>-formatted workflow model to the GraphML<sup>2</sup>-formatted workflow-supported social network. Especially, the framework adopted the open-sources information visualization toolkits, such as Prefuse, JFreeChart, and Log4j, in order to visualize the degree-centrality measurements to each of the workflow performers making up a workflow-supported social network. Ra and Kim [30] defined the workflow-supported social network as a work-sharing and collaborating network of workflow-actors performing workflow-related operations in an organization, and proposed an extended GraphML schema to visualize the degrees of closenesses (closeness among workflow-performers forming a workflow-supported social network, which is named to **ccWSSN-GraphML**). Won [8] in his Ph.D. dissertation deployed a statistical analysis approach and suggested its

interpretations as the organizational networking knowledge that can be extracted from the discovered human-centered network. Kim *et al.* [31] proposed a GraphML visualization framework that is able to visualize the degrees of workflow-performers' closeness centralities measured on workflow-supported organizational social networks. Additionally, Kim *et al.* [32] built a theoretical framework for quantitatively measuring and graphically representing the degrees of closeness centralization among workflow-performers, which comprises three procedural phases (discovery, analysis and quantitation phases) with a couple of functional transformations. Park *et al.* [15] formalized a theoretical framework coping with discovery phase and analysis phase, and conceive a series of formalisms and algorithms for representing, discovering, and analyzing the workflow-supported social network. As a theoretical basis, they used the conceptual methodology of information control nets that used to formally describe workflow procedures and business processes. The theoretical framework was expansively implemented in the name of a systematic framework that is able to automatically discover a workflow-supported social network from an XPDL-based workflow package, construct SocioMatrices from the discovered workflow-supported social network, analyze the SocioMatrices, and visualize the workload centrality measures of all the actors in the corresponding workflow-supported social network. Liu and Kumar [18] attempted to seek a sort of feasible linkage between the social network's features, such as human-resource count, strength, size, closeness, and density, and the workflow's operational performances in a workflow-supported organization. Note that the paper [18] renamed the workflow-supported social network, which goes by the name in this paper, as the socio-technical network, in which is enclosing people, process, and technology. In their research, the authors empirically identified the human-resource count feature as well as the closeness centrality feature as the most significant features of the network, which are strongly impacting on the operational performance of an underlying workflow procedure of the IT incident management domain. The closeness centrality feature out of the identified features and its empirical analysis result are closely related with the key issue of our research in this paper, and it ought to be a motivative feature so as for our research of this paper to be usefully and extensively applicable into the domain of those workflow performance evaluation, workflow redesign and reengineering, and managerial decision-making knowledge acquisition.

**Workflow-Supported Affiliation Networks:** It is important to remind that the human-centered affiliation relationships reveal how each of the individuals is associated with the essential entity-types of the organizational resources like activity, role, application, and relevant data. So far, three types of human-centered affiliation networks [28] were introduced in the literature, such as performer-to-activity affiliation network [2], [3], [33], Performer-to-Role affiliation network [34], and Performer-to-Application affiliation network [35]. Let's briefly introduce these pioneering works.

<sup>1</sup>XPDL stands for XML Process Definition Language.

<sup>2</sup>GraphML stands for Graph Markup Language

Kim [3] firstly issued the workflow-supported performer-to-activity affiliation network as a special type of organizational social network knowledge acquired from deploying workflow technologies. In the paper, the author theoretically derived a series of concepts and algorithms not only for representing and discovering those knowledge but also for analyzing the discovered knowledge. These theoretical concepts and related algorithms were developed under the methodology of information control net workflow models, and showed that the discovered knowledge eventually represents involvements and participation relationships between a group of performers and a group of activities in workflow models. Based upon the concept and its related algorithms, the paper [36] implemented a knowledge discovery system designated only for the workflow-supported performer-to-activity affiliation networking knowledge. Battsetseg *et al.* [2] proposed a theoretical formalism to analyze a workflow-supported performer-to-activity affiliation network by measuring the organizational closeness centralities of performers as well as the organizational closeness centralities of activities. That is, the paper devised a series of algorithms for analyzing the closeness centralities of activities and performers, and described the ultimate implications of these analysis results as activity-performer affiliation knowledge in workflow-supported organizations. Additionally, the paper [2] was selected as an outstanding paper award and recommended to a journal paper published in [33]. Note that these are going to be fully extended and completed into the proposed theoretical formalism through this paper. Kim *et al.* [34] formalized the workflow-supported performer-to-role affiliation network. In the paper, the authors formally defined the workflow-supported performer-to-role affiliation networking knowledge through a series of theoretical formalisms and practical implementation for modeling, discovering, and visualizing workflow performer-to-role affiliation networking knowledge, and practical details as workflow performer-to-role affiliation knowledge representation, discovery, and visualization techniques. In particular, the paper summarily described the implications of the proposed affiliation networking knowledge as business process intelligence, and how much it is worth to discover and visualize the knowledge in workflow-driven organizations and enterprises that produce massively parallel interactions and large-scaled operational data collections through deploying and enacting massively parallel and large-scale workflow models. Kim *et al.* [35] arranged a way of discovering the performer-to-application affiliation knowledge reflecting the affiliated relationships between performers and invoked applications in a workflow process, as the most recent pioneering work in terms of discovering the special types of human-centered affiliation knowledge. Especially, this work ought to be a valuable trial in terms of assessing the utilizations and values of invoked-applications in workflow-supported organizations. Reijers *et al.* [37] pioneered the human-centered resource management issue in a workflow-supported

organization, which can be interpreted by a conceptual species of the workflow-supported affiliation networks. Through this research, they showed that the high degree of geographical closenesses among workflow-performers be led to the positive effect on workflow-supported organizational performance by conducting a case study of distributed teamworks on a workflow process model. Similarly, Kumar *et al.* [38] developed an optimal model of the human-centered resource assignments in a workflow-supported organization, and it showed to capture the compatibility between workflow-performers while assigning tasks in a workflow model to a group of performers in order to improve the quality and increase the throughput in enacting the instances of a corresponding workflow model.

Conclusively, these pioneering works, which are concerned about the human-centered affiliation knowledge, are the outputs in the stage of initiative research works, which is the discovery phase. The next stage ought to be the analysis phase. The paper of [2] was just a half-finished step forward to the analysis phase shifting from the discovery phase. This paper tries to complete the step opening the door to the analysis phase by extensively investigating formal approaches of closeness centrality measurements on workflow-supported performer-to-activity affiliation networks. In particular, P. Busch and his colleague in [19], [39] raised the logical necessity of the conceptual triangulation of workflow management, social network analysis, and knowledge management, which ought to be one of the circumstantial evidences of the theoretical importance of this paper. In terms of specialized network analysis techniques, additionally, a group of recent publications [24], [25], [40]–[45] proposed and fulfilled their own special networks and analysis, in which they fulfilled those analyses of betweenness and closeness centralities with graph theory, network optimization, and social networks like as we do the similar way in this paper.

### III. WORKFLOW-SUPPORTED PERFORMER-TO-ACTIVITY AFFILIATION NETWORKS

As a knowledge representation method for the workflow-supported performer-to-activity affiliation knowledge, the paper [3], which is authored by the authors' research group, defined a graphical and formal representation model, which is the performer-to-activity affiliation network model, and its primary goal was to achieve the following dual objectives:

- to uncover the relational structures of workflow-performers through their joint involvement in activities, and
- to reveal the relational structures of workflow-activities through their joint participation of common performers.

Additionally, those relational structures can be weighed to measure the extent of their strengths by assigning a value to each of relations between nodal types. Therefore, there are two types of activity-to-performer (or performer-to-activity) affiliations such as binary performer-to-activity affiliation type and weighted performer-to-activity affiliation



type. In the binary performer-to-activity affiliation type, its value (0 or 1) implies a binary relationship of involvement (or participation), while values in the weighted activity-performer affiliation type may represent various implications according to their application domains; typical examples of values might be stochastic (or probabilistic) values, strengths, and frequencies. This section simply introduces the concept of workflow-supported performer-to-activity affiliation network according to the binary performer-to-activity affiliation type and deploy the concept along with an operational example.

**A. REPRESENTATIONS**

The affiliation knowledge representation can be graphically depicted by a bipartite graph, which is defined as a workflow affiliation graph in this paper. So, the graphical model of a performer-to-activity affiliation network consists of two types of graphical nodes (a set of performers shaped in hexagon and a set of workflow activities shaped in circle) and a set of undirected edges between two nodal types, which means that the workflow affiliation graph is an undirected bipartite graph. In other words, a performer-to-activity affiliation graph is built by undirected lines connecting from performers aligned on one side of the diagram to workflow activities aligned on the other side, and it does not permit lines neither among performers nor among the workflow activities. Eventually, a performer-to-activity affiliation graph with the  $g$  number of performers and the  $h$  number of workflow activities can be transformed into the mathematical representation of an adjacency matrix with 2-dimension of  $g \times h$ .

**1) THE FORMAL REPRESENTATION**

The workflow-supported performer-to-activity affiliation network consists of two different types of nodes (a set of performers and a set of activities) and a set of edges connecting only between the different nodal types. The formal representation is defined as the following [Definition 1] as introduced in [3].

*Definition 1 (Performer-to-Activity Affiliation Network Model):* The workflow-supported performer-to-activity affiliation network model is formally defined as  $\Lambda = (\sigma, \psi, S)$ , over a set  $C$  of performers (actors), a set  $A$  of activities, a set  $V$  of weight-values, a set  $E_p \subseteq (C \times A)$  of edges (pairs of performers and activities), and a set  $E_a \subseteq (A \times C)$  of edges (pairs of activities and performers), where,  $\wp(A)$  represents a power set of the activity set,  $A$ :

- $S$  is a finite set of work-sharing actors or groups of some external performer-to-activity affiliation network models;
- $\sigma = \sigma_p \cup \sigma_v$  /\* Involvement Knowledge \*/  
 where,  $\sigma_p : C \rightarrow \wp(A)$  is a single-valued mapping function from a performer to its set of involved activities;  $\sigma_v : E_p \rightarrow V$  is a single-valued mapping function from an edge ( $\in E_p$ ) to its weight-value;
- $\psi = \psi_a \cup \psi_v$  /\* Participation Knowledge \*/  
 where,  $\psi_a : A \rightarrow \wp(C)$  is a single-valued mapping function from an activity to a set of participated performers;

and  $\psi_v : E_a \rightarrow V$  is a single-valued function from an edge ( $\in E_a$ ) to its weight-value;

**2) THE MATHEMATICAL REPRESENTATION: ADJACENCY MATRIX**

Eventually, the performer-to-activity affiliation network model has to be transformed into a mathematical representation form in order to be analyzed as quantitative organizational knowledge. As if the performer-to-activity affiliation network model is graphically represented by a bipartite affiliation graph, at the same time it is mathematically represented by an adjacency affiliation matrix, too. The adjacency affiliation matrix can be realized by either an involvement adjacency matrix or a participation adjacency matrix. That is, a performer-to-activity affiliation network is mathematically transformed into a performer-to-activity adjacency matrix that records the presence and absence of  $g$  performers at  $h$  workflow activities; thus its dimensions are  $g$  rows and  $h$  columns, respectively. If a certain performer  $\phi_i$  attends a workflow activity  $\alpha_j$ , then the entry in the  $i^{th}$  and  $j^{th}$  cell in the matrix equals to 1; otherwise the entry is 0. Denoting a binary performer-to-activity adjacency matrix as  $Z$ , its  $x_{i,j}$  values meet these conditions:

$$x_{i,j} = \begin{cases} 1 & \text{if performer, } \phi_i, \text{ is affiliated with activity, } \alpha_j \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

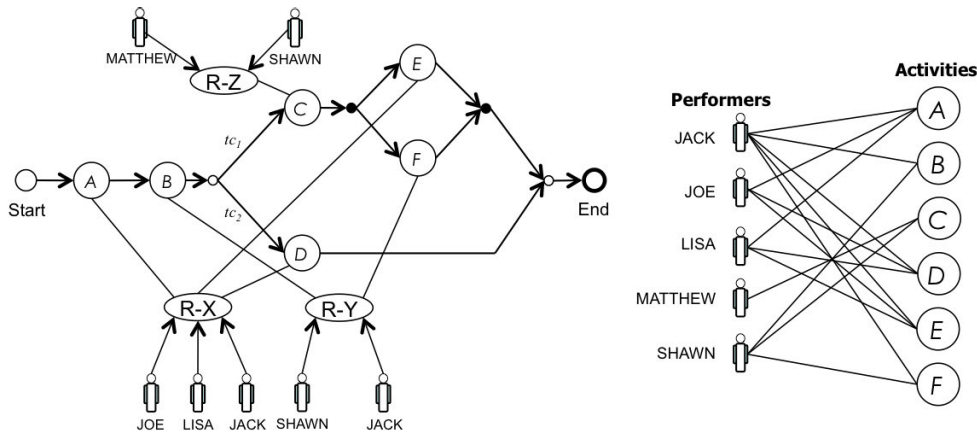
- The row total, also called row marginals, ( $\bar{D}_r$ ), of activity-performer adjacency matrix  $Z$  sum to the number of workflow activities that each performer will attend, which implies the involvement relations between activities and performers in a corresponding workflow model.

$$\bar{D}_r = \left[ \sum_{j=1}^h x_{i,j} \right]_{i=1}^g \quad (2)$$

- The column marginals, ( $\bar{D}_c$ ), indicate the number of performers who will attend each workflow activity's enactment, which implies the participation relations between performers and activities in a corresponding workflow model.

$$\bar{D}_c = \left[ \sum_{i=1}^g x_{i,j} \right]_{j=1}^h \quad (3)$$

The paper [3] developed an algorithm, which is the binary adjacency affiliation matrix generation algorithm, that automatically transforms an activity-to-performer affiliation network into two types of adjacency matrix. Due to the page limitation, it won't be described the details of the algorithm. Note that the algorithm distinguishes the involvement adjacency matrix ( $Z_p$ ) from the participation adjacency matrix ( $Z_a$ ), which easily calculate and represent the row marginals ( $\bar{D}_r$ ) and the column marginals ( $\bar{D}_c$ ), respectively.



**FIGURE 2.** Graphical Representations of an Information Control Net of the Product-Order Workflow Procedure and its Performer-to-Activity Affiliation Network.

3) KNOWLEDGE ANALYTICS: BIPARTITE MATRIX

Based upon the types of adjacency matrices ( $Z_p, Z_a$ ), it is possible to analyze a variety of knowledge analytics, such as centrality analysis equations [12], mean rates analysis [2], density measurements [28], centrality measurements [11], [14], [46], and so on, issued from the conventional social network literature. In general, an affiliation network is a bipartite graph, as described in the previous section, in which non-directed lines connect performers aligned on one side of the diagram to the workflow activities aligned on the other side. So, in order for an affiliation network to be reasonably analyzed, it needs to be mathematically represented in a bipartite matrix containing both sets of performers and activities in the rows and columns; assuming that an affiliation network has  $g$  performers and  $h$  activities, then its bipartite matrix has dimensions  $(g + h) \times (g + h)$ . Consequently, using the involvement adjacency matrix ( $Z_p$ ) and the participation adjacency matrix ( $Z_a$ ) forms an affiliation bipartite matrix,  $X^{P,A}$ , which can be schematically represented as the following equation:

$$X^{P,A} = \begin{bmatrix} \mathbf{0} & Z_p \\ Z_a & \mathbf{0} \end{bmatrix} \quad (4)$$

By applying the affiliation bipartite matrix, it is possible to measure various social networking knowledge analytical properties on activity-to-performer affiliation networking knowledge, such as density [2] and centrality [2], [46]. Density measures on an affiliation bipartite matrix give us important basic knowledge, like the average number of activities jointly attended by the pairs of performers, the average number of performers simultaneously assigned to the pairs of activities, and so on. There are, also, four types of centrality [46] possibly measuring on an affiliation bipartite matrix: degree, closeness, betweenness, and eigenvector centrality. Our focus aims at the property of closeness centrality in this paper. The details of these centrality properties won't be described anymore, because of beyond the scope of the paper.

**TABLE 1.** Binary Involvement Adjacency Matrix of FIGURE 2.

$Z_p$	$\alpha_A$	$\alpha_B$	$\alpha_C$	$\alpha_D$	$\alpha_E$	$\alpha_F$
$\phi_{jack}$	1	1	0	1	1	1
$\phi_{joe}$	1	0	0	1	1	0
$\phi_{lisa}$	1	0	0	1	1	0
$\phi_{matthew}$	0	0	1	0	0	0
$\phi_{shawn}$	0	1	1	0	0	1

**B. AN OPERATIONAL EXAMPLE**

FIGURE 2 depicts the graphical representations of an information control net (ICN) [47] model and its affiliation network model as an input and an output of the knowledge discovering algorithm [28], [35], which are the product-order workflow procedure [48], [49] and its performer-to-activity affiliation network, respectively. In terms of discovering a workflow-supported performer-to-activity affiliation network, it is needed to have not a full description of the original ICN-based workflow model [47], but a partial description only showing the activities' precedence( $\delta$ ), roles's assignment( $\epsilon$ ), performers's assignment( $\pi$ ), and the transition conditions( $\kappa$ ) that are directly related with the social perspective's point of view. At last, the discovered performer-to-activity affiliation network model represents the involvement knowledge( $\sigma$ ) as well as the participation knowledge( $\psi$ ) between the activities and the performers as a bipartite affiliation graph. Note that the formal representations of the product-order workflow procedure won't be given in this paper, and you can find out them in [48], [49].

It is necessary to apply the algorithm [3] to the previous example of the activity-to-performer affiliation network; the input of the algorithm is the formal representation property sets of the activity-to-performer affiliation network model, and its output is two types of binary activity-to-performer adjacency matrices,  $Z_p$  and  $Z_a$ , which correspond to the involvement adjacency matrix of TABLE 1 and the participation adjacency matrix of TABLE 2, respectively. Based on the adjacency matrix table, it is able to calculate the row

TABLE 2. Binary Participation Adjacency Matrix of FIGURE 2.

$Z_a$	$\phi_{jack}$	$\phi_{joe}$	$\phi_{lisa}$	$\phi_{matthew}$	$\phi_{shawn}$
$\alpha_A$	1	1	1	0	0
$\alpha_B$	1	0	0	0	1
$\alpha_C$	0	0	0	1	1
$\alpha_D$	1	1	1	0	0
$\alpha_E$	1	1	1	0	0
$\alpha_F$	1	0	0	0	1

marginals ( $\bar{D}_r$ ) as well as the column marginals ( $\bar{D}_c$ ), as the following equations:

$$\bar{D}_r = \left[ \sum_{j=1}^6 x_{i,j} \right]_{i=1}^5 = [5, 3, 3, 1, 3] \quad (5)$$

$$\bar{D}_c = \left[ \sum_{i=1}^5 x_{i,j} \right]_{j=1}^6 = [3, 2, 2, 3, 3, 2] \quad (6)$$

As you see in TABLE 1, the rows of workflow-performers show that performer<sub>1</sub> ( $\phi_{jack}$ ) is involved in 5 activities, performer<sub>2</sub> ( $\phi_{joe}$ ) and performer<sub>3</sub> ( $\phi_{lisa}$ ) are involved in 3 activities, and performer<sub>4</sub> ( $\phi_{matthew}$ ) and performer<sub>5</sub> ( $\phi_{shawn}$ ) are involved in 1 activity and 3 activities, respectively. In the same context, the columns of workflow-activities show that the organization allots 3 performers to each of activity<sub>1</sub> ( $\alpha_A$ ), activity<sub>4</sub> ( $\alpha_D$ ), and activity<sub>5</sub> ( $\alpha_E$ ), and it allots 2 performers to each of activity<sub>2</sub> ( $\alpha_B$ ), activity<sub>3</sub> ( $\alpha_C$ ), and activity<sub>F</sub> ( $\alpha_A$ ). Also, TABLE 2 represents the inverse of TABLE 1 as you see.

In interpreting the equation (5), each value of  $\bar{D}_r$  implies the number of activities, in which the corresponding performer is involved; so, it is easily infer that, for instance, the performer  $\phi_{jack}$  is involved in three workflow activities' enactments. As a result, because each of the entries of  $D_r$  is the number of workflow activities affiliated by each performer, the average number of activities being involved by a single performer (a mean value of the performer's affiliated involvement rate) is  $\frac{15}{5} = 3.0$  activities yielded from summing all the entries and dividing by the number of performers ( $g$ ). Also, in terms of interpretation of the equation (6), each value of  $\bar{D}_c$  reveals the number of performers, who are participating to the corresponding workflow activity; so, the workflow activity  $\alpha_D$  is enacted by a total of three performers. Likewise, because each of the entries of  $D_c$  is the number of performers participating to the corresponding workflow activities' enactments, the average number of performers being participated to a single activity (a mean value of the activity's affiliated participation rate) is  $\frac{15}{6} = 2.5$  performers yielded from summing all the entries and dividing by the number of activities ( $h$ ).

### C. IMPLICATIONS OF MEASURING THE ORGANIZATIONAL CLOSENESS CENTRALITY

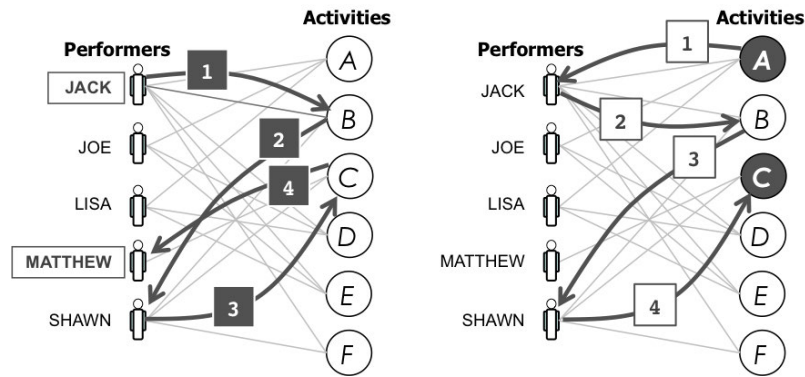
Like the information control net, almost all the workflow models commonly employ the five essential entity types (activity, role, performer, repository and application

entity-types) to represent organizational works and their procedural collaborations. Through the associative relationships between performer entity type and others, it is able to obtain the human-centered organizational knowledge such as behavioral, social, informational, collaborative, and historical knowledge. Therefore, it is possible to interpret the workflow management system as "people system" that must be designed, deployed, and understood within their social and organizational contexts. The people system ought to be able to support any formations of collaborative activities among people, which eventually build up the human-centered collaborative knowledge as the most influential and meaningful organizational knowledge. The authors have sought the most reasonable metric units for evaluating the degrees of collaborations among people in a workflow-supported organization, and it was able to find out the two types of organizational social networks at last, which are the workflow-supported social network and the workflow-supported affiliation network as described in the previous section.

In particular, the workflow-supported affiliation network ought to be a very useful means to formally and graphically represent the human-centered resource allocation and management patterns in a workflow-supported organization. This paper is mainly concerned with the performer-to-activity affiliation patterns within a specific workflow model, and their behavioral knowledge that can be obtained by using the typical metric of social networks, **Centrality**<sup>3</sup> [13]. The most widely used centrality measures are degree, closeness, betweenness, and eigenvalue. This paper actively adopts the closeness centrality metric in quantifying the degree of collaboration among people allotted into a group of workflow models. In other words, a specific workflow model is defined by a group of activities and their temporal enactment sequences, and it is associated with a group of performers taking charge of enacting its activities. In the information control net, the associations between activities and performers are defined through a group of roles.<sup>4</sup> The activity-to-performer association is not a direct association but a transitive association, and so it is formed through activity-to-role associations and role-to-performer associations. The activity-to-performer associations can be transitively obtained from

<sup>3</sup>Centrality, where a prominent actor has high involvement in many relations, regardless of whether sending and receiving ties, in a social network.

<sup>4</sup>Note that **Role** is a named designator for one or more participants which conveniently acts as the basis for partitioning of work skills, access controls, execution controls, and authority / responsibility, and **Performer** is a person that can fulfill roles to execute, to be responsible for, or to be associated with activities of an information control net.



**FIGURE 3.** Measuring the Degrees of Farnesses (Closeness Distances) *via* Activities and *via* Performers.

the activity-to-role associations and the role-to-performer associations. It is sure that the activity-performer association is divided by two directed associations (activity-to-performer association and performer-to-activity association) and both of them are many-to-many relationships, which imply that not only more than one performer can participate in enacting an activity, but also a performer is able to participate in enacting one or more activities. From analyzing these activity-to-performer associations, it is able to finally produce a workflow-supported performer-to-activity affiliation network.

This paper is particularly interested in adopting the novel concept of organizational closeness centrality<sup>5</sup> to measure the degree of farness (e.g. the collaboration degree of work-intimacy) among performers as well as the degree of farness (e.g. the share degree of work-connectivity) among activities along with the human-centered resource allocation patterns and management strategies. The semantic significance of closeness distance (degree of farness) in terms of the work-intimacy metric refers to how quickly a performer can interact with others via intermediary activities where the performers are jointly participating to. At the same time, the semantic significance of closeness distance (degree of farness) in terms of the work-connectivity metric refers to how closely an activity can connect to others via performers who are jointly involved in the same activities. FIGURE 3 illustrates the degree of farness (closeness distance) measurements in the workflow-supported performer-to-activity affiliation network exemplified in the previous section. The left-hand side shows a procedure of the degree of farness (closeness distance) calculation (result = 4) between the two performers, ‘JACK’ and ‘MATHEW,’ via the two activities, ‘B’ and ‘C,’ while the right-hand side is to draw a procedure of the degree of farness (closeness distance) calculation (result = 4) between the two activities, ‘A’ and ‘C’ via the

<sup>5</sup>The basic concept of closeness centrality [12], [13] in an affiliation network was originally developed to reflect how near an actor is to the other actors through events as well as to reflect how much an event is connected to the other events through actors.

two performers of ‘JACK’ and ‘SHAWN.’ In consequence of those consecutive calculations of all the performers, it is possible to draw the answers to the following questions from measuring the derees of farnesses through the organizational closeness centralities on a workflow-supported performer-to-activity affiliation network:

- *The collaboration degree of work-intimacy:* How quickly can a performer interact with others *via* very few intermediary activities in enacting workflow procedures?
- *The share degree of work-connectivity:* How closely can an activity connect to others *via* very few intermediary performers in enacting workflow procedures?

Conclusively, the answers to the questions is able to convey a very valuable and meaningful insight to the corresponding workflow-supported organization. It is assured that the primary rationale of the organizational closeness centrality ought to be on the questions and the answers. A series of theoretical deployments on the organizational closeness centralities will be expatiated in the next consecutive sections. Finally, it is possible to cover from all the theoretical details from the mathematical extensions of the organizational closeness centrality formulas to the algorithmic implementations to be realized as a workflow-supported organizational intelligent system supporting to measure the individual levels as well as the group levels of the organizational closeness centralities in a workflow-supported organization.

**IV. ORGANIZATIONAL CLOSNESS CENTRALITIES**

As described in the previous section, the workflow-supported performer-to-activity affiliation network is not only in a graphical form of bipartite graph with undirected lines connecting workflow-performers aligned on one side of the diagram to workflow-activities aligned on the other side, but also in a mathematical form of adjacency matrix with workflow-performer rows and workflow-activity columns. Based upon the formal graph and mathematical matrix, it is able to possibly analyze a variety of qualitative analytics



TABLE 3. Binary Performer-to-Activity Affiliation Bipartite Matrix of FIGURE 2.

$\mathbf{X}^{P,A}$	$\phi_{jack}$	$\phi_{joe}$	$\phi_{lisa}$	$\phi_{matt.}$	$\phi_{sha.}$	$\alpha_A$	$\alpha_B$	$\alpha_C$	$\alpha_D$	$\alpha_E$	$\alpha_F$
$\phi_{jack}$	—	0	0	0	0	1	1	0	1	1	1
$\phi_{joe}$	0	—	0	0	0	1	0	0	1	1	0
$\phi_{lisa}$	0	0	—	0	0	1	0	0	1	1	0
$\phi_{matthew}$	0	0	0	—	0	0	0	1	0	0	0
$\phi_{shawn}$	0	0	0	0	—	0	1	1	0	0	1
$\alpha_A$	1	1	1	0	0	—	0	0	0	0	0
$\alpha_B$	1	0	0	0	1	0	—	0	0	0	0
$\alpha_C$	0	0	0	1	1	0	0	—	0	0	0
$\alpha_D$	1	1	1	0	0	0	0	0	—	0	0
$\alpha_E$	1	1	1	0	0	0	0	0	0	—	0
$\alpha_F$	1	0	0	0	1	0	0	0	0	0	—

and quantitative measures, such as structural equivalence analysis [6], [13], correspondence analysis [46], mean rates analysis [13], density measurements [13], centrality measurements [12], and so on. In this paper, our focus concentrates upon the centrality measurements of the workflow-supported affiliation network model. More precisely speaking, it is necessary to investigate an algorithmic formalism for analyzing organizational closeness centrality measurements, of workflow-supported performer-to-activity affiliation networks.

K. Faust in [12] defined a series of well-described equations that can be applied to calculating the organizational closeness centralities based upon the bipartite affiliation matrix of a workflow-supported performer-to-activity affiliation network. In terms of developing a series of algorithms based upon the organizational closeness centrality measurement equations, it is needed to consider the relationship between the organizational closeness centrality of a performer and the organizational closeness centrality of the activities to which the performer belongs, and the relationship between the organizational closeness centrality of an activity and the organizational closeness centrality of its performers. This section, therefore, investigates the organizational closeness centralities according to the performer-centered affiliation aspect and the activity-centered affiliation aspect, separately.

**A. THE PERFORMER-TO-ACTIVITY AFFILIATION BIPARTITE MATRIX**

In order to measure the extent of organizational closeness centralities on a workflow-supported performer-to-activity affiliation network, it is needed to arrange a definite bipartite matrix, which is called performer-to-activity affiliation bipartite matrix, properly obtained from both the involvement adjacency matrix and the participation adjacency matrix of the corresponding affiliation network. As described in the previous section, the performer-to-activity affiliation bipartite matrix is arranged with sets of workflow-performers and workflow-activities in the rows and columns, respectively. It is assuming again that an affiliation graph is made up of  $g$  workflow-performers and  $h$  workflow-activities,

then its performer-to-activity affiliation bipartite matrix has the elements of  $(g + h) \times (g + h)$  in two dimensions. Likewise, a performer-to-activity affiliation bipartite matrix,  $\mathbf{X}^{P,A}$ , is obtained from the involvement adjacency matrix ( $\mathbf{Z}_p$ ) and the participation adjacency matrix ( $\mathbf{Z}_a$ ), which is graphically represented in the equation (7). For an example, the performer-to-activity affiliation graph is introduced in FIGURE 2; its eventual performer-to-activity affiliation bipartite matrix and its involvement and participation relationships are shown in TABLE 3.

$$\mathbf{X}^{P,A} = \begin{bmatrix} \mathbf{0} & \mathbf{Z}_p \\ \mathbf{Z}_a & \mathbf{0} \end{bmatrix} \tag{7}$$

In principle, the organizational closeness centrality measures are based upon the distances from a node  $n_i$  to other nodes in a performer-to-activity affiliation graph, and the distances are calculated on the performer-to-activity affiliation bipartite matrix, just like the equation (7) and TABLE 3. Because the affiliation graph is a bipartite graph, which implies that performers are only adjacent to activities and that activities are only adjacent to performers, this paper has to consider two types of the relationships as stated in [12]; one is the relationship between the organizational closeness centrality of a performer and the organizational closeness centralities of the activities to which the performer belongs, the other is the relationship between the organizational closeness centrality of an activity and the organizational closeness centrality of its performers. Therefore, the next section investigates firstly the distances from a performer in the affiliation bipartite matrix as a function of the geodesic distances to the activities in which it is involved, and the distances from an activity as a function of the geodesic distances to the performers to which it participates, in second.

**B. ORGANIZATIONAL CLOSNESS CENTRALITIES ON PERFORMERS**

Basically, the meaning of closeness centrality index in a social network implies the ‘Farness’ from a node to other nodes in the social network, and the core function of which is to calculate the average geodesic (shorted path) distance between a node and all other nodes in the graph. As described in the previous section, the performer-to-activity affiliation

**Algorithm 1** The Geodesic Distances Algorithm From a Performer

---

```

1: Given Global A Binary Affiliation Bipartite Matrix,  $X^{P,A}[g + h, g + h]$ ; A Set of Performers,  $P$ ; A Set of Activities,  $A$ ;
2: procedure PcGeodesicDistance
3:   Input A Performer (From),  $n_i$ ; Either a Performer or an Activity (To),  $n_j$ ;
4:   Output A Performer-Centered Geodesic Distance Measure,  $G^{P,A}[n_i, n_j]$ ;
5:   Local An Activity Distance Vector,  $G_k[1..h]$ , initialized by maximum;
6:   Local A Performer Distance Vector,  $H_k[1..g]$ , initialized by maximum;
7:   if ( $n_j \in P \wedge n_i \neq n_j$ ) then
8:     for ( $\forall m_k \in A_k$  adjacent to  $n_i$ ) do
9:        $G_k[m_k] \leftarrow gDistance(n_i, m_k, n_j)$ ;
10:    end for
11:     $G^{P,A}[n_i, n_j] \leftarrow 1 + \text{minimum}\left(G_k[i]\right)_{i=1}^h$ ;
12:  else if ( $n_j \in A$ ) then
13:    for ( $\forall m_k \in A_k$  adjacent to  $n_i$ ) do
14:      if ( $m_k = n_j$ ) then
15:         $G_k[m_k] \leftarrow 0$ ;
16:      else if ( $m_k \neq n_j$ ) then
17:         $P_s \leftarrow$  all performers who are adjacent to  $m_k$ ;  $P_s \leftarrow P_s - n_i$ ;
18:        for ( $\forall n_s \in P_s$ ) do
19:           $H_k[n_s] \leftarrow hDistance(m_k, n_s, n_j)$ ;
20:        end for
21:         $G_k[m_k] \leftarrow 1 + \text{minimum}\left(H_k[i]\right)_{i=1}^g$ ;
22:        initialize  $H_k[1..g]$  by maximum;
23:      end if
24:    end for
25:  end if
26:   $G^{P,A}[n_i, n_j] \leftarrow 1 + \text{minimum}\left(G_k[i]\right)_{i=1}^h$ ;
27:  return  $G^{P,A}[n_i, n_j]$ ;
28: end procedure

```

---

network is a special type of the workflow-supported social network, and it is represented by a bipartite affiliation graph with relationships (or connections) between performers and activities in a group of workflow procedures. Thus, calculating the organizational closeness centralities of performers in a bipartite affiliation graph<sup>6</sup> begins with a function of the geodesic distances between the performers and the activities. The geodesic distance from a performer node  $n_i$  to any node  $n_j$  (performer) is which is choosing a minimal distance value for every activity node  $m_k$  adjacent to  $n_i$ . The geodesic distance

<sup>6</sup>Note that the bipartite affiliation graph must be a connected graph. If the graph is disconnected, then organizational closeness centrality is undefined, since some pairs of nodes are unreachable and the distance between them is infinite.

from a performer node  $n_i$  to any node  $m_j$  (activity) is which is choosing a minimal distance value for every activity node  $m_k$  adjacent to  $n_i$  with choosing a minimal distance value for every performer node  $n_s$  adjacent to  $m_k$ . From these geodesic distances of every pair of performer-to-activity and performer-to-performer in a bipartite affiliation graph, it is able to eventually build up a performer-centered geodesic distance matrix,  $G^{P,A}$ .

#### 1) ALGORITHM FOR THE GEODESIC DISTANCE FROM A PERFORMER

By extensively applying the equations of (8) and (9), as shown at the bottom of the page, it is possible to calculate

$$d(n_i, n_j) = 1 + \text{minimum}\left(d(m_k, n_j)\right)_{k=1}^a \quad (n_i \neq n_j \wedge a \leq h) \quad (8)$$

$$d(n_i, m_j) = \begin{cases} 1 + \text{minimum}\left(d(m_k, m_j)\right)_{k=1}^a & (a \leq h) \\ d(m_k, m_j) = 1 + \text{minimum}\left(d(n_s, m_j)\right)_{s=1}^b & (b \leq g \wedge m_k \neq m_j) \end{cases} \quad (9)$$

---

```

1: procedure gDistance
2:   Input The Initiation Performer,  $n_{ini}$ ; The Adjacent Activity,  $m_k$ ; The Destination Performer,  $n_j$ ;
3:   Local An Activity Distance Vector,  $G_k[1..h]$ , initialized by maximum;
4:   if ( $X^{P,A}[m_k, n_j] = 0$ ) then ▷ no direct tie between  $m_k$  and  $n_j$ .
5:      $P_s \leftarrow$  all performers who are adjacent to  $m_k$ ;  $P_s \leftarrow P_s - n_{ini}$ ;
6:     for ( $\forall n_s \in P_s$ ) do
7:        $A_s \leftarrow$  all activities that are adjacent to  $n_s$ ;  $A_s \leftarrow A_s - m_k$ ;
8:       for ( $\forall m_s \in A_s$ ) do
9:          $G_k[m_s] \leftarrow 2 + gDistance(n_s, m_s, n_j)$ ;
10:      end for
11:    end for
12:   else if ( $X^{P,A}[m_k, n_j] = 1$ ) then ▷ direct tie between  $m_k$  and  $n_j$ .
13:      $G_k[m_k] \leftarrow 1$ ;
14:   end if
15:   return  $\text{minimum} \left( G_k[i] \right)_{i=1}^h$ ;
16: end procedure

```

---

```

17: procedure hDistance
18:   Input The Initiation Activity,  $m_{ini}$ ; The Adjacent Performer,  $n_s$ ; The Destination Activity,  $m_j$ ;
19:   Local A Performer Distance Vector,  $H_k[1..g]$ , initialized by maximum;
20:   if ( $X^{P,A}[n_s, m_j] = 0$ ) then ▷ no direct tie between  $n_s$  and  $m_j$ .
21:      $A_k \leftarrow$  all activities that are adjacent to  $n_s$ ;  $A_k \leftarrow A_k - m_{ini}$ ;
22:     for ( $\forall m_k \in A_k$ ) do
23:        $P_k \leftarrow$  all performers who are adjacent to  $m_k$ ;  $P_k \leftarrow P_k - n_s$ ;
24:       for ( $\forall n_k \in P_k$ ) do
25:          $H_k[n_s] \leftarrow 2 + hDistance(m_k, n_k, m_j)$ ;
26:       end for
27:     end for
28:   else if ( $X^{P,A}[n_s, m_j] = 1$ ) then ▷ direct tie between  $n_s$  and  $m_j$ .
29:      $H_k[n_s] \leftarrow 1$ ;
30:   end if
31:   return  $\text{minimum} \left( H_k[i] \right)_{i=1}^g$ ;
32: end procedure

```

---

the organizational closeness centralities of performers for a workflow-supported performer-to-activity affiliation network. The essential part of those equations must be the functions of calculating the geodesic distance from a performer node,  $n_i$ , to another performer node,  $n_j$ , and the geodesic distance from a performer node,  $n_i$ , to an activity node,  $m_j$ , respectively. This subsection describes about devising an algorithm with recursive functions, to algorithmically implement the essential equations of (8) and (9). Assume that the algorithm operates on a given performer-to-activity affiliation adjacency matrix,  $\mathbf{X}^{P,A}$ , representing the corresponding workflow-supported performer-to-activity affiliation network, and its functional procedure name is ‘PcGeodesicDistance()’ using two recursive functions, ‘gDistance()’ and ‘hDistance()’, which are corresponding to the two essential equations of (8) and (9), respectively. The output of the algorithm is the geodesic distance of a performer,  $n_i$ , to either a performer or an activity,  $n_j$ , and it is saved on the performer-centered geodesic distance matrix,  $\mathbf{G}^{P,A}$ , as a value of the cell,  $\mathbf{G}^{P,A}[n_i, n_j]$ . The time complexity

of the algorithm is  $\mathbf{O}(\mathbf{N})$ , where  $\mathbf{N} = g + h - 1$ , and  $g$  is the number of performers and  $h$  is the number of activities in a corresponding workflow-supported performer-to-activity affiliation network.

As a result, the algorithm returns a geodesic distance only from a single performer to either another performer or an activity, and so the algorithm iteratively operates for  $g \times (g + h - 1)$  times in order to calculate all of the cells on the performer-centered geodesic distance matrix,  $\mathbf{G}^{P,A}$ . For an operational example, the algorithm applies to the affiliation adjacency matrix, TABLE 3, of the workflow-supported performer-to-activity affiliation network discovered from the product-order workflow procedure in FIGURE 2, and it was able to produce the complete performer-centered geodesic distance matrix, as shown in TABLE 4.

## 2) PERFORMER-CENTERED ORGANIZATIONAL CLOSENESS CENTRALITY MEASUREMENT EQUATIONS

According to the performer-centered geodesic distance matrix, the equations are developed for measuring

TABLE 4. Performer-Centered Geodesic Distance Matrix,  $\mathbf{G}^{P,A}$ , from TABLE 3.

$\mathbf{G}^{P,A}$	$\phi_{jack}$	$\phi_{joe}$	$\phi_{lisa}$	$\phi_{matt.}$	$\phi_{sha.}$	$\alpha_A$	$\alpha_B$	$\alpha_C$	$\alpha_D$	$\alpha_E$	$\alpha_F$
$\phi_{jack}$	—	2	2	4	2	1	1	3	1	1	1
$\phi_{joe}$	2	—	2	4	4	1	5	3	1	1	1
$\phi_{lisa}$	2	2	—	6	4	1	3	5	1	1	3
$\phi_{matthew}$	4	6	6	—	2	5	3	1	5	5	3
$\phi_{shawn}$	2	4	4	2	—	3	1	1	3	3	1

organizational closeness centralities among performers and between performers and activities in a bipartite affiliation network. The essential part of the equations is on summing all of the geodesic distances on each row corresponding to each of the performers, and which is the following equation (10),  $G_{DP}(n_i)(1 \leq i \leq g)$ .

$$\left[ G_{DP}(n_i) \right]_{i=1}^g = \left[ \sum_{j=1}^{g+h} G^{P,A}[n_i, n_j] \ (i \neq j) \right]_{i=1}^g \quad (10)$$

Based on the summation function of geodesic distances and the conventional closeness centrality equations [12], the organizational closeness centrality measurement equations are redefined as the following equations, (11) and (12), as shown at the bottom of the next page, which are the index and the standardized index of the organizational closeness centrality of a performer ( $n_i$ ), respectively, in a workflow-supported performer-to-activity affiliation network. That is, the index of organizational closeness centrality of a performer is computed as the inverse of the average geodesic distance from a performer to all of other nodes (both performers and activities). In particular, the standardized index is needed to control the size of the performer-to-activity affiliation network by multiplying the index and  $(2 \times (g-1) + h) \cdot (g+h-1)^{-1}$  together. For an instance, when a performer is close to all other activities and an activity is close to all performers,  $OC_{CP}(n_i) = (g+h-1) \cdot (2 \times (g-1) + h)^{-1}$ , which implies that the index varies with network size. Thus, through multiplying the index by  $(2 \times (g-1) + h) \cdot (g+h-1)^{-1}$ , the standardized index gives always the measure between 0 and 1, which makes for us to allow meaningful comparisons of performers across different sizes of performer-to-activity affiliation networks.

C. ORGANIZATIONAL CLOSENESS CENTRALITY ON ACTIVITIES

So far, the previous subsection has arranged the formalism for measuring performer-centered organizational closeness centralities on a workflow-supported performer-to-activity affiliation network, whereas this subsection investigates the organizational closeness centrality from the workflow-activities' points of view. Basically, the workflow-supported performer-to-activity affiliation network is a bipartite graph depicting two dimensional relationships between performers and activities by a sets of performer-to-activity pairs and activity-performer pairs. Thus, it is necessary to calculate the organizational closeness centralities of activities, again. Likewise, it begins with two functions of the geodesic distances

between activities and other activities as well as between activities and performers, respectively. First, the one function of geodesic distance measuring from an activity node  $m_i$  to any other activity node  $m_j$  is which is choosing a minimal distance value from every performer node  $n_k$  adjacent to  $m_i$ . Next, the other function of geodesic distance measuring from an activity node  $m_i$  to any performer node  $n_j$  is which is choosing a minimal distance value from every performer node  $n_k$  adjacent to  $m_i$  with choosing a minimal distance value from every activity node  $m_s$  adjacent to  $n_k$ . From these geodesic distances of every pair of activity-to-activity and activity-to-performer in a bipartite affiliation graph, it is able to eventually build up an activity-centered geodesic distance matrix,  $\mathbf{G}^{A,P}$ . Note that  $m$  and  $n$  represent an activity node and a performer node, respectively, in a corresponding workflow-supported performer-to-activity affiliation network.

1) ALGORITHM FOR THE GEODESIC DISTANCE FROM AN ACTIVITY

This subsection develops an algorithm for implementing the above equations of (13) and (14), as shown at the bottom of the next page, by revising the algorithm developed in the previous subsection, which is able to calculate the organizational closeness centralities from a performers' point of view. Likewise, the essential part of those equations must be the functions of calculating the geodesic distance from an activity node,  $m_i$ , to another activity node,  $m_j$ , and the geodesic distance from an activity node,  $m_i$ , to a performer node,  $n_j$ , respectively. Assume that the algorithm also operates on a given performer-to-activity affiliation adjacency matrix,  $\mathbf{X}^{P,A}$ , representing the corresponding workflow-supported performer-to-activity affiliation network, and its functional procedure name is 'AcGeodesicDistance()' using two recursive functions, 'gDistance()' and 'hDistance()', too. The output of the algorithm is the geodesic distance measure of an activity,  $m_i$ , to either a performer or an activity,  $m_j$ , and it is saved on the activity-centered geodesic distance matrix,  $\mathbf{G}^{P,A}$ , as a value of the cell,  $\mathbf{G}^{A,P}[m_i, m_j]$ . The time complexity of the algorithm is  $\mathbf{O}(\mathbf{N})$ , where  $\mathbf{N} = g + h - 1$ , and  $g$  is the number of performers and  $h$  is the number of activities in a corresponding workflow-supported performer-to-activity affiliation network.

2) ACTIVITY-CENTERED ORGANIZATIONAL CLOSENESS CENTRALITY MEASUREMENT EQUATIONS

Based upon the geodesic distance functions of (13) and (14), it is necessary to make an expression for summing all



**TABLE 5. Performer-Centered Organizational Closeness Centrality Measures in FIGURE 2.**

	$G_{DP}(n_i)$	$OC_{CP}(n_i)$	$OC_{CP}^S(n_i)$
$\phi_{jack}$	18	$10 \cdot 18^{-1} = 0.555$	$14 \cdot 18^{-1} = 0.7777$
$\phi_{joe}$	24	$10 \cdot 24^{-1} = 0.416$	$14 \cdot 24^{-1} = 0.5833$
$\phi_{lisa}$	28	$10 \cdot 28^{-1} = 0.357$	$14 \cdot 28^{-1} = 0.5000$
$\phi_{matthew}$	40	$10 \cdot 40^{-1} = 0.250$	$14 \cdot 40^{-1} = 0.3500$
$\phi_{shawn}$	24	$10 \cdot 24^{-1} = 0.416$	$14 \cdot 24^{-1} = 0.5833$

**TABLE 6. Activity-Centered Geodesic Distance Matrix,  $G^{A,P}$ , of FIGURE 2.**

$G^{A,P}$	$\phi_{jack}$	$\phi_{joe}$	$\phi_{lisa}$	$\phi_{matt.}$	$\phi_{sha.}$	$\alpha_A$	$\alpha_B$	$\alpha_C$	$\alpha_D$	$\alpha_E$	$\alpha_F$
$\alpha_A$	1	1	1	5	3	—	2	4	2	2	2
$\alpha_B$	1	3	3	3	1	2	—	2	2	2	2
$\alpha_C$	3	5	5	1	1	4	2	—	4	4	4
$\alpha_D$	1	1	1	5	3	2	2	4	—	2	2
$\alpha_E$	1	1	1	5	3	2	2	4	2	—	2
$\alpha_F$	1	3	3	3	1	2	2	2	2	2	—

**TABLE 7. Activity-Centered Organizational Closeness Centrality Measures in FIGURE 2.**

	$G_{DA}(m_i)$	$OC_{CA}(m_i)$	$OC_{CA}^S(m_i)$
$\alpha_A$	22	$10 \cdot 22^{-1} = 0.454$	$15 \cdot 22^{-1} = 0.6818$
$\alpha_B$	21	$10 \cdot 21^{-1} = 0.476$	$15 \cdot 21^{-1} = 0.7142$
$\alpha_C$	28	$10 \cdot 28^{-1} = 0.357$	$15 \cdot 28^{-1} = 0.5357$
$\alpha_D$	23	$10 \cdot 23^{-1} = 0.434$	$15 \cdot 23^{-1} = 0.6521$
$\alpha_E$	23	$10 \cdot 23^{-1} = 0.434$	$15 \cdot 23^{-1} = 0.6521$
$\alpha_F$	21	$10 \cdot 21^{-1} = 0.416$	$15 \cdot 21^{-1} = 0.7142$

of the geodesic distances on each of rows, which is deserved for each of the activities, and the expression is just the following equation (15), as shown at the bottom of the page. Also, the organizational closeness central-

ity measurement equations are redefined as the following expressions, (16) and (17), as shown at the bottom of page 15, which are the index and the normalized index of the organizational closeness centrality of activities, respectively,

- The Index of Performer-Centered Organizational Closeness Centrality

$$\left[OC_{CP}(n_i)\right]_{i=1}^g = \left[(g + h - 1) \cdot \left(G_{DP}(n_i)\right)^{-1}\right]_{i=1}^g \tag{11}$$

- The Standardized Index of Performer-Centered Organizational Closeness Centrality

$$\left[OC_{CP}^S(n_i)\right]_{i=1}^g = \frac{(2 \times (g - 1) + h)}{(g + h - 1)} \cdot \left[OC_{CP}(n_i)\right]_{i=1}^g \tag{12}$$

$$d(m_i, m_j) = 1 + \mathbf{minimum}\left(d(n_k, m_j)\right)_{k=1}^a \quad (m_i \neq m_j \wedge a \leq g) \tag{13}$$

$$d(m_i, n_j) = \begin{cases} 1 + \mathbf{minimum}\left(d(n_k, n_j)\right)_{k=1}^a & (a \leq g) \\ d(n_k, n_j) = 1 + \mathbf{minimum}\left(d(m_s, n_j)\right)_{s=1}^b & (b \leq h \wedge n_k \neq n_j) \end{cases} \tag{14}$$

$$\left[G_{DA}(m_i)\right]_{i=1}^h = \left[\sum_{j=1}^h d(m_i, m_j) + \sum_{j=1}^g d(m_i, n_j)\right]_{i=1}^h \quad (m_i \neq m_j) \tag{15}$$

**Algorithm 2** The Geodesic Distances Algorithm From an Activity

---

```

1: Given Global A Binary Affiliation Bipartite Matrix,  $X^{P,A}[g + h, g + h]$ ; A Set of Performers,  $P$ ; A Set of Activities,  $A$ ;
2: procedure AcGeodesicDistance
3:   Input An Activity (From),  $m_i$ ; Either an Activity or a Performer (To),  $m_j$ ;
4:   Output An Activity-Centered Geodesic Distance Measure,  $G^{A,P}[m_i, m_j]$ ;
5:   Local A Performer Distance Vector,  $H_k[1..g]$ , initialized by maximum;
6:   Local An Activity Distance Vector,  $G_k[1..h]$ , initialized by maximum;
7:   if ( $m_j \in A \wedge m_i \neq m_j$ ) then
8:     for ( $\forall n_k \in P_k$  adjacent to  $m_i$ ) do
9:        $H_k[n_k] \leftarrow hDistance(m_i, n_k, m_j)$ ;
10:    end for
11:     $G^{A,P}[m_i, m_j] \leftarrow 1 + \text{minimum}\left(H_k[i]\right)_{i=1}^g$ ;
12:   else if ( $m_j \in P$ ) then
13:     for ( $\forall n_k \in P_k$  adjacent to  $m_i$ ) do
14:       if ( $n_k = m_j$ ) then
15:          $H_k[n_k] \leftarrow 0$ ;
16:       else if ( $n_k \neq m_j$ ) then
17:          $A_s \leftarrow$  all activities that are adjacent to  $n_k$ ;  $A_s \leftarrow A_s - m_i$ ;
18:         for ( $\forall m_s \in A_s$  adjacent to  $n_k$ ) do
19:            $G_k[m_s] \leftarrow gDistance(n_k, m_s, m_j)$ ;
20:         end for
21:          $H_k[n_k] \leftarrow 1 + \text{minimum}\left(G_k[i]\right)_{i=1}^h$ ;
22:         initialize  $G_k[1..h]$  by maximum;
23:       end if
24:     end for
25:   end if
26:    $G^{A,P}[m_i, m_j] \leftarrow 1 + \text{minimum}\left(H_k[i]\right)_{i=1}^g$ ;
27:   return  $G^{A,P}[m_i, m_j]$ ;
28: end procedure

```

---

```

1: procedure gDistance
2:   Input The Initiation Performer,  $n_{ini}$ ; The Adjacent Activity,  $m_k$ ; The Destination Performer,  $n_j$ ;
3:   Output The Minimum Geodesic Distance between  $m_k$  and  $n_j$ ;
4:   Local An Activity Distance Vector,  $G_k[1..h]$ , initialized by maximum;
5:   if ( $X^{P,A}[m_k, n_j] = 0$ ) then ▷ no direct tie between  $m_k$  and  $n_j$ .
6:      $P_s \leftarrow$  all performers who are adjacent to  $m_k$ ;  $P_s \leftarrow P_s - n_{ini}$ ;
7:     for ( $\forall n_s \in P_s$ ) do
8:        $A_s \leftarrow$  all activities that are adjacent to  $n_s$ ;  $A_s \leftarrow A_s - m_k$ ;
9:       for ( $\forall m_s \in A_s$ ) do
10:         $G_k[m_s] \leftarrow 2 + gDistance(n_s, m_s, n_j)$ ;
11:       end for
12:     end for
13:   else if ( $X^{P,A}[m_k, n_j] = 1$ ) then ▷ direct tie between  $m_k$  and  $n_j$ .
14:      $G_k[m_k] \leftarrow 1$ ;
15:   end if
16:   return  $\text{minimum}\left(G_k[i]\right)_{i=1}^h$ ;
17: end procedure

```

---

in a workflow-supported performer-to-activity affiliation network.

That is, the index of organizational closeness centrality of an activity is computed as the inverse of the average geodesic distance from an activity to all of other nodes

(both performers and activities), as defined in the case of the performer-centered organizational closeness centrality. In particular, the normalized index is needed to control the size of the performer-to-activity affiliation network by multiplying the index and  $(2 \times (h - 1) + g) \cdot (g + h - 1)^{-1}$

---

```

18: procedure hDistance
19:   Input The Initiation Activity,  $m_{ini}$ ; The Adjacent Performer,  $n_s$ ; The Destination Activity,  $m_j$ ;
20:   Output The Minimum Geodesic Distance between  $n_s$  and  $m_j$ ;
21:   Local A Performer Distance Vector,  $H_k[1..g]$ , initialized by maximum;
22:   if ( $X^{P,A}[n_s, m_j] = 0$ ) then ▷ no direct tie between  $n_s$  and  $m_j$ .
23:      $A_k \leftarrow$  all activities that are adjacent to  $n_s$ ;  $A_k \leftarrow A_k - m_{ini}$ ;
24:     for ( $\forall m_k \in A_k$ ) do
25:        $P_k \leftarrow$  all performers who are adjacent to  $m_k$ ;  $P_k \leftarrow P_k - n_s$ ;
26:       for ( $\forall n_k \in P_k$ ) do
27:          $H_k[n_s] \leftarrow 2 + hDistance(m_k, n_k, m_j)$ ;
28:       end for
29:     end for
30:   else if ( $X^{P,A}[n_s, m_j] = 1$ ) then ▷ direct tie between  $n_s$  and  $m_j$ .
31:      $H_k[n_s] \leftarrow 1$ ;
32:   end if
33:   return minimum $\left(H_k[i]\right)_{i=1}^g$ ;
34: end procedure

```

---

together. For an instance, when an activity is close to all of the performers and a performer is close to all of the activities, the index is  $OC_{CA}(n_i) = (g + h - 1) \cdot (2 \times (h - 1) + g)^{-1}$ , which implies that the index varies with network size. Thus, through multiplying the index by  $(2 \times (h - 1) + g) \cdot (g + h - 1)^{-1}$ , it is necessary to normalize the measure between 0 and 1, which makes for us to allow meaningful comparisons of workflow-activities across different sizes of performer-to-activity affiliation networks.

#### D. DISCUSSION

So far, this section has described the details of the algorithmic formalisms, as the investigation results, for the organizational closeness centrality measurements in a workflow-supported performer-to-activity affiliation network. In particular, because the affiliation network is a bipartite graph, which implies that performers are only adjacent to activities and that the activities are only adjacent to the performers, it is necessary to consider the two types of the closeness relationships in terms of developing the algorithmic formalisms. One is the performer-centered organizational closeness centrality, the other is the activity-centered organizational closeness centrality. The combination of these two types of the closeness relationships shows that there is a clear relationship between the organizational closeness centralities of performers and the organizational closeness centralities

of activities in a workflow-supported performer-to-activity affiliation network. That is, the organizational closeness centrality of a performer is a function of the minimum distances to its activities, and the organizational closeness centrality of an activity is a function of the minimum distances to its performers. In conducting the investigation, several issues and assumptions are taken into account for measuring the organizational closeness centralities as the followings:

- **Connectedness.** It is assumed that the workflow-supported performer-to-activity affiliation network must be a connected network in order for the devised algorithms to operate correctly. If the affiliation network is disconnected, then its organizational closeness centrality equations are undefined, since some pairs of nodes are unreachable and the geodesic distances between them are infinite or undefined, which implies that the denominators of the equations become infinite or undefined, too. At the same time, it is needed to further investigate how to treat the disconnected workflow-supported performer-to-activity affiliation networks in measuring their organizational closeness centralities, as one of the future research works.
- **The Indexes and Network Sizes.** The indexes of the organizational closeness centralities can be calculated through taking either the inverse of the sum of geodesic distances [13] or the inverse of the average

- 
- The Index of Organizational Closeness Centrality of Activities

$$\left[OC_{CA}(m_i)\right]_{i=1}^h = \left[(g + h - 1) \cdot \left(G_{DA}(m_i)\right)^{-1}\right]_{i=1}^h \quad (16)$$

- The Normalized Index of Organizational Closeness Centrality of Activities

$$\left[OC_{CA}^S(m_i)\right]_{i=1}^h = \frac{(2 \times (h - 1) + g)}{(g + h - 1)} \cdot \left[OC_{CA}(m_i)\right]_{i=1}^h \quad (17)$$

geodesic distance [12]. the latter is taken in this investigation, because the indexes measured from the former are numerically too small to be difficult to interpret as meaningful implications. The indexes' value is between 0 and 1 because of taking the inverse function, and its magnitude is completely dependent on the size of a corresponding performer-to-activity affiliation network. For an instance, the minimal connected performer-to-activity affiliation network with  $g = 1$  and  $h = 1$  ought to be measured as the largest index, 1, and the bigger (the higher numbers of  $g$  and  $h$ ) the network's size is, then the smaller the index of the network's organizational closeness centrality is. With the indexes, it is unable to fulfill the meaningful comparisons among performers across different sizes of workflow-supported performer-to-activity affiliation networks. In consequence, this is the reason why the standardized indexes of the organizational closeness centralities have to be developed.

- **Group Organizational Closeness Centralization.**

In principle, it is necessary to measure the extent of group organizational closeness centralization that gives a dispersion measure indicating the hierarchy of organizational closeness centralities within a workflow-supported performer-to-activity affiliation network. Specifically, the group organizational closeness centralization is allowing us to measure the extent to which performers in a given performer-to-activity affiliation network differ in their organizational closeness centralities. Therefore, it is necessary to develop the indexes of group organizational closeness centralizations on both performers and activities as the followings: Note that  $OC_{CP}^S(n^*)$  and  $OC_{CA}^S(m^*)$  denote the observed-largest standardized indexes of the organizational closeness centralities of performers and activities, respectively, and  $OC_{CP}^S(n_i)$  and  $OC_{CP}^S(m_i)$  are the standardized indexes of the  $g + h - 1$  other performers and activities, respectively.  $OC_{CP}^S(n^+)$  and  $OC_{CA}^S(m^+)$  denote the theoretically maximum possible standardized indexes of the organizational closeness centralities of performers and activities, respectively. As results, in the equation of (18), as shown at the bottom of the next page, the numerator sums the observed differences between the largest performer centrality and all the others, and the denominator is the theoretically maximum possible sum of those differences. Likewise, the numerator of the equation (23), as shown at the bottom of page 23, sums the observed differences between the largest activity centrality and all the others, and the denominator is the theoretically maximum possible sum of those differences. In the equations of (19) and (21), as shown at the bottom of the next page, the denominator of the maximum possible sums occurs in the *mutual-star* bipartite graph with satisfying  $g = h$  in a workflow-supported performer-to-activity affiliation network, not only where one performer is involved in all the activities, but all of the activities are participated only to the performer,

but also where one activity is participated to all the performers, but all of the performers are involved only in the activity, simultaneously. The differences in the organizational closeness centrality measures between the most central performer (or activity) and all the other performers and activities would be the theoretically maximum. So, devising the denominators for the cases of  $g > h$  and  $g < h$ , would be left as a future research work.

- **Weighted Affiliation: Valued Affiliation Bipartite Matrix.** By applying the affiliation bipartite matrix, it is possible to measure various analytical properties on activity-to-performer affiliation networking knowledge. It is also necessary to take the valued affiliations into consideration. That is, the activity-to-performer affiliation network, as exemplified in the previous subsection, shows that five performers are involved in six activities with involvement's weight-value ( $= 1$ ), and that five performers participate in six activities with participation's weight value ( $= 1$ ), as well. In terms of interpreting the weight-values, the involvement's weight-values on the performer-activity edges imply designed (or planned) work-involvements, while the participation's weight-values on the activity-performer edges represent "show (or no-show)" implying designed (or planned) work-participations. At this moment, it is necessary to remind the workflow-supported affiliation network rediscovery issue concerning about fulfilled work-affiliation (work-involvement/participation) relations that can be rediscovered from workflow execution logs. The weight-values of these fulfilled work-affiliation relations imply "frequencies or the number of times" counted from the corresponding workflow instances. This is why it is needed to differentiate the binary affiliation bipartite matrix from the valued affiliation bipartite matrix, and it is assured that the weighted affiliations represented in a valued affiliation bipartite matrix have meaningful implications in workflow-supported organizations.

## V. OPERATIONAL EXPERIMENTS

So far, the concept of workflow-supported performer-to-activity affiliation networks has been proposed and its related mathematical formalisms have been deployed for organizational closeness centrality measurements on the affiliations between performers and activities. This section tries to validate the practical implementation of these proposed concept and algorithms through a couple of operational experiments. One is for validating the organizational closeness centrality measurements on the binary workflow-supported performer-to-activity affiliation network through being fulfilled upon a group of synthetic workflow process models that are arranged in two workflow process packages, five workflow process models, fifty activities and sixteen performers as shown in FIGURE 4 and TABLE 8, and the other is for validating the organizational closeness centrality measurements on the weighted workflow-supported performer-to-activity



**TABLE 8. Activity Arrangements and Performer Involvements in the Workflow Packages.**

W. Package	Workflow Model	Activity	Performer
HR-Dept-Pkg <sub>1</sub>	Hiring-Workflow-Model	$\alpha_1 \sim \alpha_{16}$ (16 Activities)	Jeff, Ed, Christiaan, Emily, Adam, Cynthia,
	Performance-Management-Workflow-Model	$\alpha_{17} \sim \alpha_{25}$ (9 Activities)	Joylette, Amanda, Nathaniel, Bryan, Tamara, Ashley, Ryan, Alan, Chris, Holly
HR-Dept-Pkg <sub>2</sub>	Employee-Training-Workflow-Model	$\alpha_{26} \sim \alpha_{36}$ (11 Activities)	Jeff, Ed, Christiaan, Emily, Adam, Cynthia, Joylette, Amanda,
	Department-Management-Workflow-Model	$\alpha_{37} \sim \alpha_{44}$ (8 Activities)	Nathaniel, Bryan, Tamara, Ashley, Ryan, Alan, Chris, Holly
	Salary-Negotiation-Workflow-Model	$\alpha_{45} \sim \alpha_{50}$ (6 Activities)	

affiliation network through being fulfilled upon the discovered ICN-based workflow process model mined from a dataset of workflow event logs, which was released for the Conformance Checking Challenge 2019 (CCC2019) that was a co-located event of the 1<sup>st</sup> international conference on process mining 2019.

**A. EXPERIMENT FOR BINARY AFFILIATIONS BETWEEN PERFORMERS AND ACTIVITIES**

The workflow process models used in the operational experiment is assumed to be arranged for the department of human resources in an imaginary workflow-supported organization. As shown in the figure, the department is operated by two packages of workflow processes, HR-Dept-Pkg<sub>1</sub> and HR-Dept-Pkg<sub>2</sub>, each of which is grouped by two workflow process models, Hiring-Workflow and Performance-Management-Workflow, and three workflow process models, Employee-Training-Workflow, Department-Management-Workflow, Salary-Negotiation-Workflow, respectively, and is

graphically defined by the ICN-based workflow process modeling system developed by the authors’ research group. All of these workflow process models are eventually represented in a textual form of XPDL-based workflow process models. Assume that all the 50 activities in the models are properly arranged into the five workflow process models according to the rule of the information control net methodology, and they are selectively assigned by the 16 performers as shown in FIGURE 4 and TABLE 8.

**1) DISCOVERY OF BINARY PERFORMER-TO-ACTIVITY AFFILIATIONS**

For the sake of the operational experiment, the authors’ research group implemented a discovery and analysis system that is able to discover a workflow-supported performer-to-activity affiliation network model from the XPDL-based workflow process packages and models prepared for the operational experiment, that is able to also visualize the discovered affiliation network model, and that is able to

– The Index of Group Organizational Closeness Centrality on Performers

$$OC_{GP} = \frac{\sum_{i=1}^{g+h-1} [OC_{CP}^S(n^*) - OC_{CP}^S(n_i)]}{\mathbf{maximum} \sum_{i=1}^{g+h-1} [OC_{CP}^S(n^+) - OC_{CP}^S(n_i)]} \tag{18}$$

$$OC_{GP} = \frac{\sum_{i=1}^{g+h-1} [OC_{CP}^S(n^*) - OC_{CP}^S(n_i)]}{(g + h - 2) \cdot \frac{2 \cdot (h-1)}{2 \cdot (g-1) + 3 \cdot (h-1) + 1}}, (g = h) \tag{19}$$

– The Index of Group Organizational Closeness Centrality on Activities

$$OC_{GA} = \frac{\sum_{i=1}^{g+h-1} [OC_{CA}^S(m^*) - OC_{CA}^S(m_i)]}{\mathbf{maximum} \sum_{i=1}^{g+h-1} [OC_{CA}^S(m^+) - OC_{CA}^S(m_i)]} \tag{20}$$

$$OC_{GP} = \frac{\sum_{i=1}^{g+h-1} [OC_{CA}^S(m^*) - OC_{CA}^S(m_i)]}{(g + h - 2) \cdot \frac{2 \cdot (g-1)}{2 \cdot (h-1) + 3 \cdot (g-1) + 1}}, (g = h) \tag{21}$$

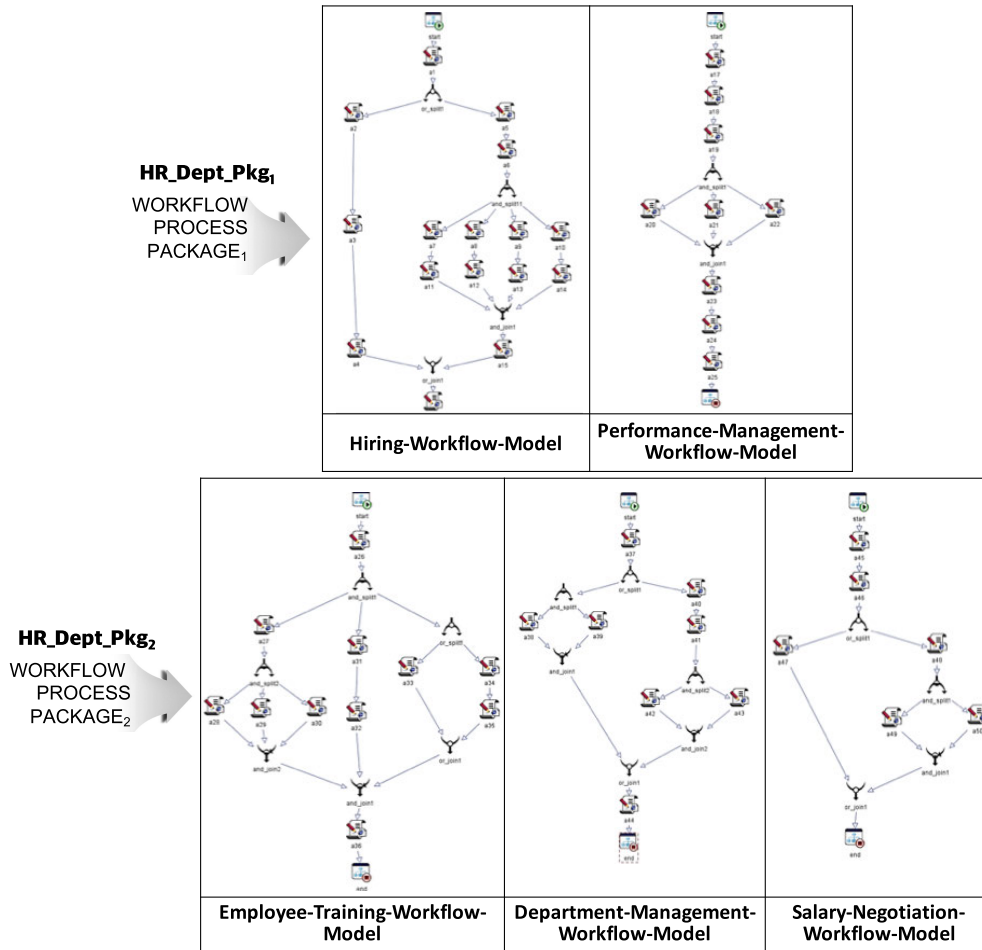


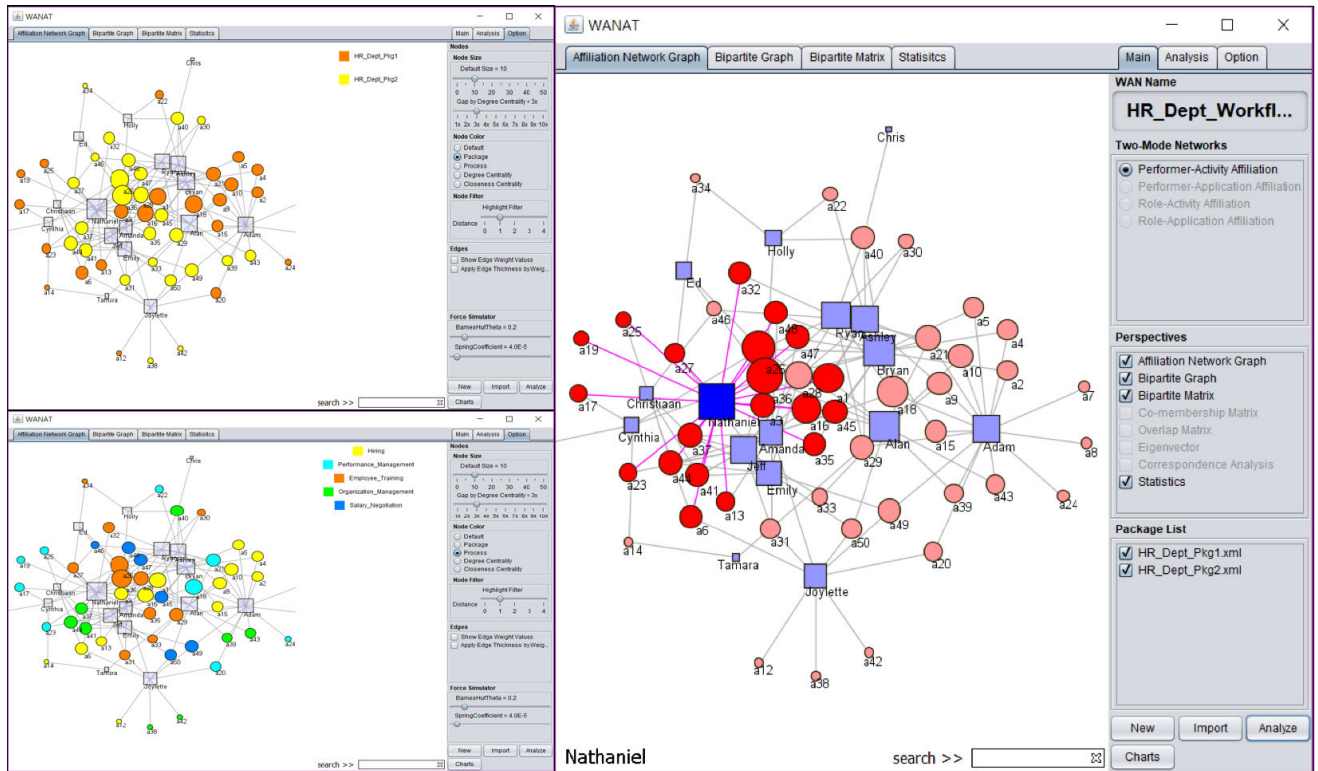
FIGURE 4. The ICN-Based Workflow Package and its Models for the Operational Experiment.

TABLE 9. The Performer-to-Activity Affiliations Discovered from the Workflow Packages.

Performer	Assigned Activities
$\phi_{jeff}$	$\alpha_1, \alpha_{16}, \alpha_{23}, \alpha_{28}, \alpha_{31}, \alpha_{33}, \alpha_{36}, \alpha_{37}, \alpha_{41}, \alpha_{44}$
$\phi_{ed}$	$\alpha_{26}, \alpha_{27}, \alpha_{34}, \alpha_{36}$
$\phi_{christiaan}$	$\alpha_6, \alpha_{25}, \alpha_{37}, \alpha_{41}, \alpha_{44}, \alpha_{46}$
$\phi_{emily}$	$\alpha_{27}, \alpha_{28}, \alpha_{37}, \alpha_{41}, \alpha_{44}, \alpha_{49}, \alpha_{50}$
$\phi_{adam}$	$\alpha_2, \alpha_4, \alpha_5, \alpha_7, \alpha_8, \alpha_9, \alpha_{10}, \alpha_{15}, \alpha_{18}, \alpha_{20}, \alpha_{21}, \alpha_{24}, \alpha_{39}, \alpha_{43}, \alpha_{49}$
$\phi_{cynthia}$	$\alpha_{11}, \alpha_{14}, \alpha_{17}, \alpha_{26}, \alpha_{36}$
$\phi_{joylette}$	$\alpha_3, \alpha_6, \alpha_{12}, \alpha_{20}, \alpha_{31}, \alpha_{38}, \alpha_{42}, \alpha_{49}, \alpha_{50}$
$\phi_{amanda}$	$\alpha_{13}, \alpha_{18}, \alpha_{31}, \alpha_{37}, \alpha_{41}, \alpha_{44}, \alpha_{45}, \alpha_{47}, \alpha_{48}$
$\phi_{nathaniel}$	$\alpha_1, \alpha_3, \alpha_6, \alpha_{13}, \alpha_{16}, \alpha_{17}, \alpha_{19}, \alpha_{23}, \alpha_{25}, \alpha_{26}, \alpha_{27}, \alpha_{32}, \alpha_{35}, \alpha_{36}, \alpha_{37}, \alpha_{41}, \alpha_{44}, \alpha_{45}, \alpha_{47}, \alpha_{48}$
$\phi_{bryan}$	$\alpha_2, \alpha_4, \alpha_5, \alpha_{15}, \alpha_{18}, \alpha_{21}, \alpha_{26}, \alpha_{28}, \alpha_{35}, \alpha_{36}, \alpha_{40}, \alpha_{45}, \alpha_{47}, \alpha_{48}$
$\phi_{tamara}$	$\alpha_{14}, \alpha_{29}$
$\phi_{ashley}$	$\alpha_1, \alpha_9, \alpha_{10}, \alpha_{16}, \alpha_{18}, \alpha_{22}, \alpha_{28}, \alpha_{29}, \alpha_{30}, \alpha_{40}, \alpha_{46}$
$\phi_{ryan}$	$\alpha_{18}, \alpha_{21}, \alpha_{26}, \alpha_{28}, \alpha_{29}, \alpha_{30}, \alpha_{32}, \alpha_{40}$
$\phi_{alan}$	$\alpha_1, \alpha_9, \alpha_{10}, \alpha_{16}, \alpha_{21}, \alpha_{26}, \alpha_{33}, \alpha_{36}, \alpha_{39}, \alpha_{43}, \alpha_{50}$
$\phi_{chris}$	$\alpha_{40}$
$\phi_{holly}$	$\alpha_3, \alpha_{22}, \alpha_{34}, \alpha_{40}$

finally measure the organizational closeness centralities on performers as well as on activities based upon the algorithms proposed in this paper. FIGURE 5 and TABLE 9 show the captured screens of three different types of visualizing the

affiliation network model and the performer-to-activity affiliation relationships discovered from the operational experiment, respectively, through the implemented system. The upper screen of the left-hand side of the figure is to visualize



**FIGURE 5. Three Types of Visualization for the Workflow-Supported Performer-to-Activity Affiliation Network Model Captured from the Implemented System.**

the affiliation network model coloring the performers and the activities in the same workflow package, whereas the lower screen is to visualize the model coloring the performers and the activities in the same workflow model. And the right-hand side of the figure is to visualize the performer-to-activity affiliation network model with holding all the performers and all the activities in their same colors as well as in the shapes of rectangular and circle, respectively, and it also visualizes the performer (=Nathaniel) colored in blue and its affiliated activities (=  $\alpha_1, \alpha_3, \alpha_6, \alpha_{13}, \alpha_{16}, \alpha_{17}, \alpha_{19}, \alpha_{23}, \alpha_{25}, \alpha_{26}, \alpha_{27}, \alpha_{32}, \alpha_{35}, \alpha_{36}, \alpha_{37}, \alpha_{41}, \alpha_{44}, \alpha_{45}, \alpha_{47}, \alpha_{48}$ ) colored in red, in particular. Actually, the three screens of the figure comes from the table of the performer-to-activity assignment relationships that the implemented system automatically discovered from the XPDL-based workflow models of the operational experiment. Through these screens of the figure and the related table, it is possibly show that the concept of workflow-supported performer-to-activity affiliation networks is tangible and automatically realizable into a software system.

**2) MEASUREMENTS OF THE ORGANIZATIONAL CLOSNESS CENTRALITIES**

For the model of workflow-supported performer-to-activity affiliation networks introduced in the previous subsection, an operational experiment is carried out for measuring the organizational closeness centralities on all the performers

(e.g. 16 performers) as well as all the activities (e.g. 50 activities). For the operational experiment, the implemented system is supplemented with appending the organizational closeness centrality measurement and visualization functions as followings:

- 1) *Performer-Centered Organizational Closeness Centrality Measurement on the Operational Experiment:* First of all, it is made up for the implemented system by applying the performer-centered organizational closeness centrality measurement equations to build the performer-centered geodesic distance matrix, and produced the indexes as well as their standardized indexes of organizational closeness centralities of the performers. The upper captured-screen of FIGURE 6 shows the organizational closeness centralities on all the performers in the screen snapshot captured from and produced by the implemented system, and TABLE 10 is to list all the indexes as well as standardized indexes of the organizational closeness centralities on all the performers involved in the models of the operational experiment.
  - The measures in the table shows that the organizational closeness centrality index and its standardized index of the performer  $\phi_{nathaniel}$ , for an example, are 0.4276 and 0.5263, respectively, and the measures are the highest indexes among all the performers involved in the five workflow models of the operational experiment. Through these indexes, it is able

TABLE 10. The Performer-Centered Organizational Closeness Centrality Measures in FIGURE 5.

	$G_{DP}(n_i)$	$OC_{CP}(n_i)$	$OC_{CP}^S(n_i)$
$\phi_{jack}$	174	$65 \cdot 174^{-1} = 0.3736$	$174 \cdot 80 \cdot 65^{-1} = 0.4598$
$\phi_{ed}$	202	$65 \cdot 202^{-1} = 0.3218$	$202 \cdot 80 \cdot 24^{-1} = 0.3960$
$\phi_{christiaan}$	212	$65 \cdot 212^{-1} = 0.3066$	$212 \cdot 80 \cdot 65^{-1} = 0.3774$
$\phi_{emily}$	178	$65 \cdot 178^{-1} = 0.3652$	$178 \cdot 80 \cdot 65^{-1} = 0.4494$
$\phi_{adam}$	174	$65 \cdot 174^{-1} = 0.3736$	$174 \cdot 80 \cdot 65^{-1} = 0.4598$
$\phi_{cynthia}$	206	$65 \cdot 206^{-1} = 0.3155$	$206 \cdot 80 \cdot 65^{-1} = 0.3883$
$\phi_{joylette}$	182	$65 \cdot 182^{-1} = 0.3571$	$182 \cdot 80 \cdot 65^{-1} = 0.4396$
$\phi_{amanda}$	180	$65 \cdot 180^{-1} = 0.3611$	$180 \cdot 80 \cdot 65^{-1} = 0.4444$
$\phi_{nathaniel}$	152	$65 \cdot 152^{-1} = 0.4276$	$152 \cdot 80 \cdot 65^{-1} = 0.5263$
$\phi_{bryan}$	164	$65 \cdot 164^{-1} = 0.3963$	$164 \cdot 80 \cdot 65^{-1} = 0.4878$
$\phi_{tamara}$	240	$65 \cdot 240^{-1} = 0.2708$	$240 \cdot 80 \cdot 65^{-1} = 0.3333$
$\phi_{ashley}$	172	$65 \cdot 172^{-1} = 0.3779$	$172 \cdot 80 \cdot 65^{-1} = 0.4651$
$\phi_{ryan}$	172	$65 \cdot 172^{-1} = 0.3779$	$172 \cdot 80 \cdot 65^{-1} = 0.4651$
$\phi_{alan}$	168	$65 \cdot 168^{-1} = 0.3869$	$168 \cdot 80 \cdot 65^{-1} = 0.4762$
$\phi_{chris}$	250	$65 \cdot 250^{-1} = 0.2600$	$250 \cdot 80 \cdot 65^{-1} = 0.3200$
$\phi_{holly}$	204	$65 \cdot 204^{-1} = 0.3186$	$204 \cdot 80 \cdot 65^{-1} = 0.3922$

TABLE 11. Activity-Centered Organizational Closeness Centrality Measures in FIGURE 4.

Activity	$G_{DA}(m_i)$	$OC_{CA}(m_i)$	$OC_{CA}^S(m_i)$
$\alpha_1$	168	$65 \cdot 168^{-1} = 0.3869$	$114 \cdot 168 \cdot 65^{-1} = 0.6786$
$\alpha_3$	182	$65 \cdot 182^{-1} = 0.3571$	$114 \cdot 182 \cdot 65^{-1} = 0.6264$
$\alpha_5$	196	$65 \cdot 196^{-1} = 0.3316$	$114 \cdot 196 \cdot 65^{-1} = 0.5816$
$\alpha_7$	238	$65 \cdot 238^{-1} = 0.2731$	$114 \cdot 238 \cdot 65^{-1} = 0.4790$
$\alpha_9$	186	$65 \cdot 186^{-1} = 0.3495$	$114 \cdot 186 \cdot 65^{-1} = 0.6129$
$\alpha_{11}$	270	$65 \cdot 270^{-1} = 0.2407$	$114 \cdot 270 \cdot 65^{-1} = 0.4222$
$\alpha_{13}$	202	$65 \cdot 202^{-1} = 0.3218$	$114 \cdot 202 \cdot 65^{-1} = 0.5644$
$\alpha_{15}$	196	$65 \cdot 196^{-1} = 0.3316$	$114 \cdot 196 \cdot 65^{-1} = 0.5816$
$\alpha_{17}$	208	$65 \cdot 208^{-1} = 0.3125$	$114 \cdot 208 \cdot 65^{-1} = 0.5481$
$\alpha_{19}$	216	$65 \cdot 216^{-1} = 0.3009$	$114 \cdot 216 \cdot 65^{-1} = 0.5278$
$\alpha_{21}$	180	$65 \cdot 180^{-1} = 0.3611$	$114 \cdot 180 \cdot 65^{-1} = 0.6333$
$\alpha_{23}$	208	$65 \cdot 208^{-1} = 0.3125$	$114 \cdot 208 \cdot 65^{-1} = 0.5481$
$\alpha_{25}$	212	$65 \cdot 212^{-1} = 0.3066$	$114 \cdot 212 \cdot 65^{-1} = 0.5377$
$\alpha_{26}$	156	$65 \cdot 156^{-1} = 0.4167$	$114 \cdot 156 \cdot 65^{-1} = 0.7308$
$\alpha_{28}$	172	$65 \cdot 172^{-1} = 0.3779$	$114 \cdot 172 \cdot 65^{-1} = 0.6628$
$\alpha_{30}$	220	$65 \cdot 220^{-1} = 0.2955$	$114 \cdot 220 \cdot 65^{-1} = 0.5182$
$\alpha_{32}$	190	$65 \cdot 190^{-1} = 0.3421$	$114 \cdot 190 \cdot 65^{-1} = 0.6000$
$\alpha_{34}$	242	$65 \cdot 242^{-1} = 0.2686$	$114 \cdot 242 \cdot 65^{-1} = 0.4711$
$\alpha_{36}$	152	$65 \cdot 152^{-1} = 0.4276$	$114 \cdot 152 \cdot 65^{-1} = 0.7500$
$\alpha_{38}$	246	$65 \cdot 246^{-1} = 0.2642$	$114 \cdot 246 \cdot 65^{-1} = 0.4634$
$\alpha_{40}$	186	$65 \cdot 186^{-1} = 0.3495$	$114 \cdot 186 \cdot 65^{-1} = 0.6129$
$\alpha_{42}$	246	$65 \cdot 246^{-1} = 0.2642$	$114 \cdot 246 \cdot 65^{-1} = 0.4634$
$\alpha_{44}$	186	$65 \cdot 186^{-1} = 0.3495$	$114 \cdot 186 \cdot 65^{-1} = 0.6129$
$\alpha_{46}$	216	$65 \cdot 216^{-1} = 0.3009$	$114 \cdot 216 \cdot 65^{-1} = 0.5278$
$\alpha_{48}$	184	$65 \cdot 184^{-1} = 0.3533$	$114 \cdot 184 \cdot 65^{-1} = 0.6196$
$\alpha_{50}$	196	$65 \cdot 196^{-1} = 0.3316$	$114 \cdot 196 \cdot 65^{-1} = 0.5816$

to capture the human-centered organizational knowledge that the performer  $\phi_{nathaniel}$  is not only interacting with others by communicating most directly or through most few intermediaries but also participating most directly to the workflow-activities in performing the instances of the corresponding workflow models. Conclusively speaking, it is analytically predicted that the performer  $\phi_{nathaniel}$  ought to be the most interactive employee with others as well as the most knowledgeable employee on business activities in fulfilling the five workflow models of the operational experiment.

2) Activity-Centered Organizational Closeness Centrality Measurement on the Operational Experiment: Next, it is

made up for the implemented system again for applying the organizational closeness centrality measurement equations to build the activity-centered geodesic distance matrix and to produce the indexes as well as their normalized indexes of organizational closeness centralities of all the activities. The lower captured-screen of FIGURE 6 shows the screen snapshot captured from the implemented system, on which you can see the measured organizational closeness centralities of all the activities. TABLE 11 is to list the evenly selected out of all the measured indexes and standardized indexes of the organizational closeness centralities of all the activities. – The measures in the tables show that the organizational closeness centrality index and its





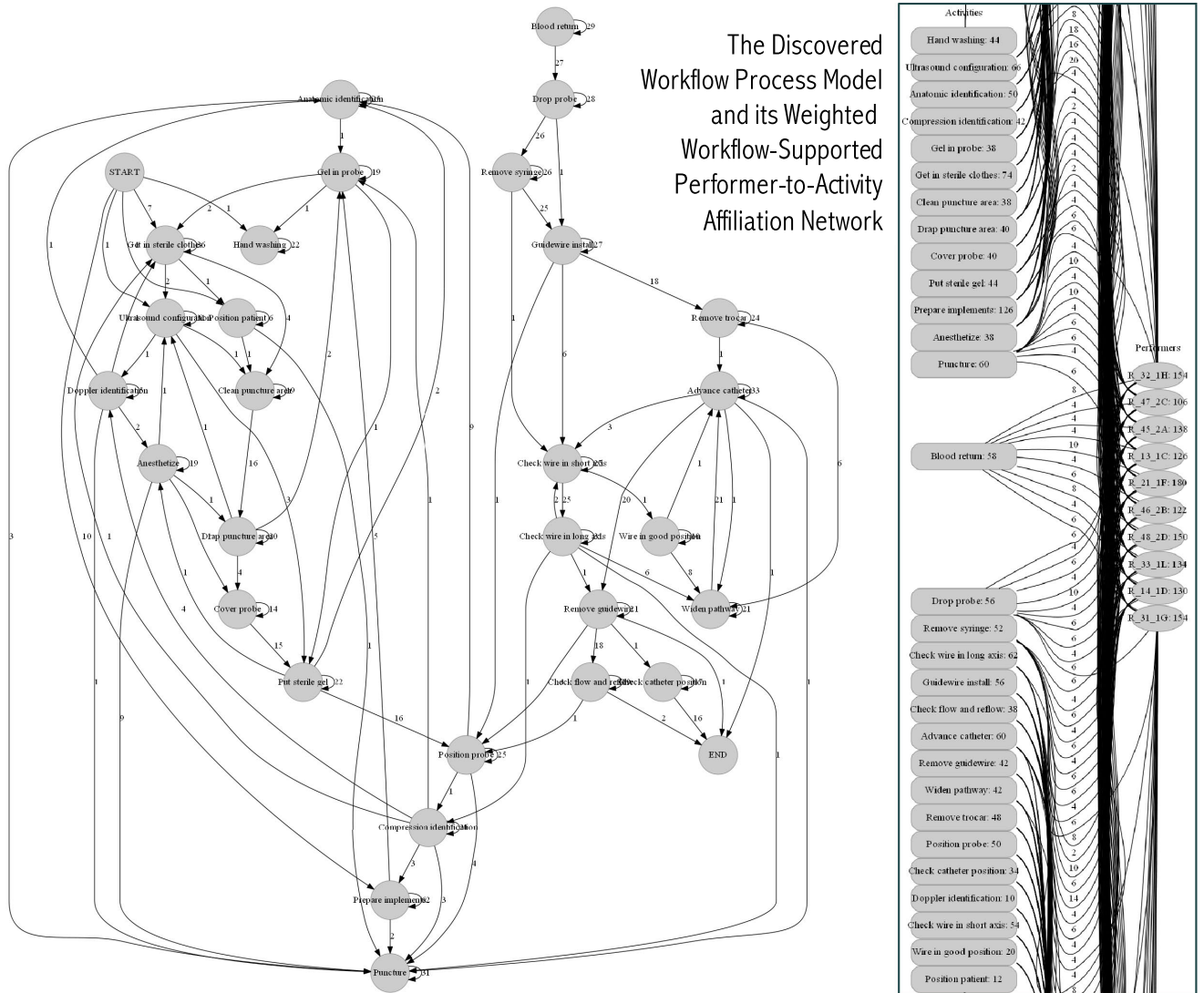
FIGURE 6. The Performer-Centered and the Activity-Centered Organizational Closeness Centrality Measurements: The Screen Snapshots Captured from the Implemented System.

normalized index of the activity  $\alpha_{36}$  are 0.4276 and 0.7500, respectively, and the values are the highest indexes among all the 50 activities associated in the five workflow models of the operational experiment. Through these indexes, it is able to capture the activity-centered organizational knowledge that the activity,  $\alpha_{36}$ , ought to be the most valuable

workflow-activities in fulfilling the corresponding workflow models in the operational experiment.

### 3) SUMMARY OF THE EXPERIMENT

Summarily, by showing the implemented system's operational experiment, it is able to validate the correctness as well as the feasibility of the algorithms not only for the



The Discovered Workflow Process Model and its Weighted Workflow-Supported Performer-to-Activity Affiliation Network

FIGURE 7. The Workflow Process Model and its Weighted Workflow-Supported Performer-to-Activity Affiliation Network Discovered from the Experiment Dataset.

activity-centered geodesic distance measurement but also for the performer-centered geodesic distance measurement. The algorithms and the related equations are applied to the implemented system that aims to discover, analyze and visualize both types of the organizational closeness centrality measures based upon a workflow-supported performer-to-activity affiliation network. The upper part of FIGURE 6 is a screen snapshot captured from the system and shows the measures of the performer-centered organizational closeness centralities for all the 16 performers involved in the five workflow models of the operational experiment. The lower part of FIGURE 6 is also captured from the system and shows the measures of the activity-centered organizational closeness centralities for all the 50 activities arranged in the five workflow models of the operational experiment. Based upon these measures, it is necessary to also measure the extent of group organizational closeness centralization that gives a dispersion measure indicating the hierarchy of organizational closeness centralities

within the workflow-supported performer-to-activity affiliation network discovered from the operational experiment. These measures can be applied to the equations of the indexes of group organizational closeness centralizations on both performers and activities as the followings:

- The Index of Group Organizational Closeness Centrality on Performers: The performer having the observed largest standardized index,  $OC_{CP}^S(n^*) = OC_{CP}^S(\phi_{nathaniel}) = 0.5263$ ;
- The Index of Group Organizational Closeness Centrality on Activities: The activity having the observed largest standardized index,  $OC_{CA}^S(m^*) = OC_{CA}^S(\alpha_{36}) = 0.7500$ ;

Note that  $OC_{CP}^S(n^*)$  and  $OC_{CA}^S(m^*)$  denote the observed largest standardized indexes of the organizational closeness centralities of performers and activities, respectively. From the observed standardized index measures of the performers,  $OC_{CP}^S(n^*) = OC_{CP}^S(\phi_{nathaniel})$  and  $OC_{CA}^S(m^*) = OC_{CA}^S(\alpha_{36})$ .

**TABLE 12.** The Activity-to-Performer Affiliations Discovered from the CCC2019 Dataset.

Activity	Enacted Performers
Advance catheter	$P_1, P_2, P_3, P_4, P_5, P_6, P_7, P_8, P_9, P_{10}$
Anatomic identification	$P_1, P_2, P_3, P_4, P_5, P_6, P_7, P_8, P_9, P_{10}$
Anesthetize	$P_1, P_2, P_3, P_4, P_5, P_6, P_7, P_8, P_9, P_{10}$
Blood return	$P_1, P_2, P_3, P_4, P_5, P_6, P_7, P_8, P_9, P_{10}$
Check catheter position	$P_1, P_2, P_3, P_4, P_5, P_6, P_7, P_8, P_9, P_{10}$
Check flow and reflux	$P_1, P_2, P_3, P_4, P_5, P_6, P_7, P_8, P_9, P_{10}$
Check wire in long axis	$P_1, P_2, P_3, P_4, P_5, P_6, P_7, P_8, P_9, P_{10}$
Check wire in short axis	$P_1, P_2, P_3, P_4, P_5, P_6, P_7, P_8, P_{10}$
Clean puncture area	$P_1, P_2, P_3, P_4, P_5, P_6, P_7, P_8, P_9, P_{10}$
Compression identification	$P_1, P_2, P_3, P_4, P_5, P_6, P_7, P_8, P_9, P_{10}$
Cover probe	$P_1, P_2, P_3, P_4, P_5, P_6, P_7, P_8, P_9, P_{10}$
Doppler identification	$P_1, P_2, P_3, P_4, P_9$
Drap puncture area	$P_1, P_2, P_3, P_4, P_5, P_6, P_7, P_8, P_9, P_{10}$
Drop probe	$P_1, P_2, P_3, P_4, P_5, P_6, P_7, P_8, P_9, P_{10}$
Gel in probe	$P_1, P_2, P_3, P_4, P_5, P_6, P_7, P_8, P_9, P_{10}$
Get in sterile clothes	$P_1, P_2, P_3, P_4, P_5, P_6, P_7, P_8, P_9, P_{10}$
Guidewire install	$P_1, P_2, P_3, P_4, P_5, P_6, P_7, P_8, P_9, P_{10}$
Hand washing	$P_1, P_2, P_3, P_4, P_5, P_6, P_7, P_8, P_9, P_{10}$
Position patient	$P_2, P_4, P_6, P_8, P_{10}$
Position probe	$P_1, P_2, P_3, P_4, P_5, P_6, P_7, P_8, P_9, P_{10}$
Prepare implements	$P_1, P_2, P_3, P_4, P_5, P_6, P_7, P_8, P_9, P_{10}$
Puncture	$P_1, P_2, P_3, P_4, P_5, P_6, P_7, P_8, P_9, P_{10}$
Put sterile gel	$P_1, P_2, P_3, P_4, P_5, P_6, P_7, P_8, P_9, P_{10}$
Remove guidewire	$P_1, P_2, P_3, P_4, P_5, P_6, P_7, P_8, P_9, P_{10}$
Remove syringe	$P_1, P_2, P_3, P_4, P_5, P_6, P_7, P_8, P_9, P_{10}$
Remove trocar	$P_1, P_2, P_3, P_4, P_5, P_6, P_7, P_8, P_9, P_{10}$
Ultrasound configuration	$P_1, P_2, P_3, P_4, P_5, P_6, P_7, P_8, P_9, P_{10}$
Widen pathway	$P_1, P_2, P_3, P_4, P_5, P_6, P_7, P_8, P_9, P_{10}$
Wire in good position	$P_1, P_2, P_3, P_4, P_6, P_7, P_8, P_{10}$

$OC_{CP}^S(n^+)$  and  $OC_{CA}^S(m^+)$  denote the theoretically maximum possible standardized indexes of the organizational closeness centralities of performers and activities, respectively, and the maximum standardized indexes of both are the same values of  $OC_{CP}^S(n^+) = 1.0$  and  $OC_{CA}^S(m^+) = 1.0$ .

**B. EXPERIMENT FOR WEIGHTED AFFILIATIONS BETWEEN PERFORMERS AND ACTIVITIES**

As raised the issue in the previous section, the weighted affiliation concept ought to be meaningfully applicable to the workflow-supported affiliation network models, too. A typical case of the weighted affiliation concept ought to be the workflow process mining activities that discover workflow-supported affiliation networks with affiliated occurrences from datasets of workflow process enactment

event logs. Therefore, this section describes a practical experiment that carries out an affiliation relationship discovery case study and proves that the proposed organizational closeness centrality equations and algorithms are extensively applicable to the weighted affiliation networks.

**1) DISCOVERY OF WEIGHTED PERFORMER-TO-ACTIVITY AFFILIATIONS**

The authors’ research group has successfully developed a workflow process mining system that is also able to discover the weighted affiliations and their occurrences from a dataset of workflow process enactment event log traces. The developed mining system assumes that the traces are formatted in the 1849-2016 IEEE XES (eXtensible Event Stream) [50] of a tag-based language, which aims to provide

$$OC_{GP} = \frac{\sum_{i=1}^{16+50-1} [OC_{CP}^S(n^*) - OC_{CP}^S(n_i)]}{\mathbf{maximum} \sum_{i=1}^{16+50-1} [OC_{CP}^S(n^+) - OC_{CP}^S(n_i)]} = \frac{\left[ \frac{1.5400}{9.1192} \right]}{\left[ \frac{1.5400}{9.1192} \right]} = \left[ 0.1689 \right] \tag{22}$$

$$OC_{GA} = \frac{\sum_{i=1}^{16+50-1} [OC_{CA}^S(m^*) - OC_{CA}^S(m_i)]}{\mathbf{maximum} \sum_{i=1}^{16+50-1} [OC_{CA}^S(m^+) - OC_{CA}^S(m_i)]} = \frac{\left[ \frac{8.7383}{21.2393} \right]}{\left[ \frac{8.7383}{21.2393} \right]} = \left[ 0.4115 \right] \tag{23}$$



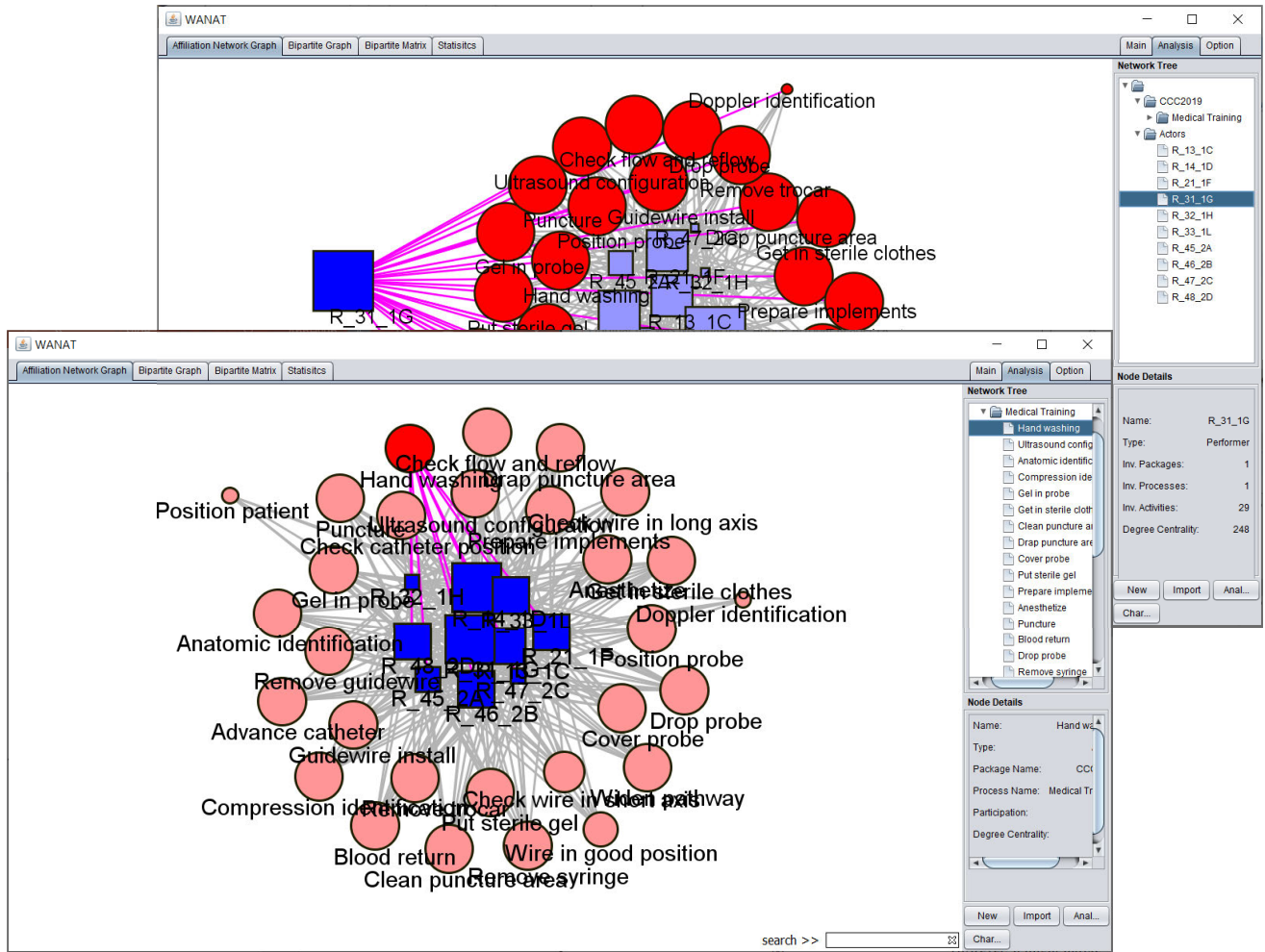


FIGURE 8. The Captured-Screen of Performer-Centered Affiliations and Activity-Centered Affiliations Discovered from the Experimental Dataset.

a unified and extensible methodology for capturing systems' behaviors from event logs. The basic XML schema specifying the structure of the event log stream contains `<trace>` and `<event>` attributes as the tag names of trace objects and event objects, respectively. For the sake of the operational experiment, a dataset of workflow process enactment event log traces is prepared from the 4TU.Research for Data [51], and the dataset also contains a bunch of the XES-formatted traces. Note that the dataset was for the Conformance Checking Challenge 2019 (CCC2019) that was a co-located event of the 1st international conference on process mining 2019. Moreover, the dataset is about the enactment history of a medical training process model and collected a course delivery from an institution that teaches future doctors how to perform a CVC (Central Venous Catheter) installation and how medical students learn how to install the medical treatment with ultrasound. In the dataset, there are total 20 traces, 1394 events, 29 activities, and 10 performers.

FIGURE 7 shows the workflow process model and its weighted bipartite graph of the workflow-supported performer-to-activity affiliation network discovered from the

CCC2019 dataset. The weighted bipartite graph on right-hand side is discovered from the workflow process model on the left-hand side. As you can see in the right-hand side of the figure, the performer-to-activity affiliation relationships have such numbers as their own occurrences. The total number of affiliation relationships on the weighted bipartite graph is 287 occurrences along with all the edges between 10 performers and 29 activities. TABLE 12 is the list of the affiliation relationships between the activities and the performers<sup>7</sup> discovered from the dataset.

The workflow process mining system finally generates a textual form of the extended XPDL-based workflow process model discovered from the dataset, and then it is able to export to the corresponding file so as for anyone to import the discovered XPDL-based workflow process model to the discovery and analysis system of affiliation networks introduced in the previous experiment. As the final result

<sup>7</sup>The identifiers of the performers are labeled by  $P_1 \sim P_{10}$  in the table. Each of the discovered identifiers is corresponding to R-13-1C, R-14-1D, R-21-1F, R-31-1G, R-32-1H, R-33-1L, R-45-2A, R-46-2B, R-47-2C, R-48-2D, respectively.



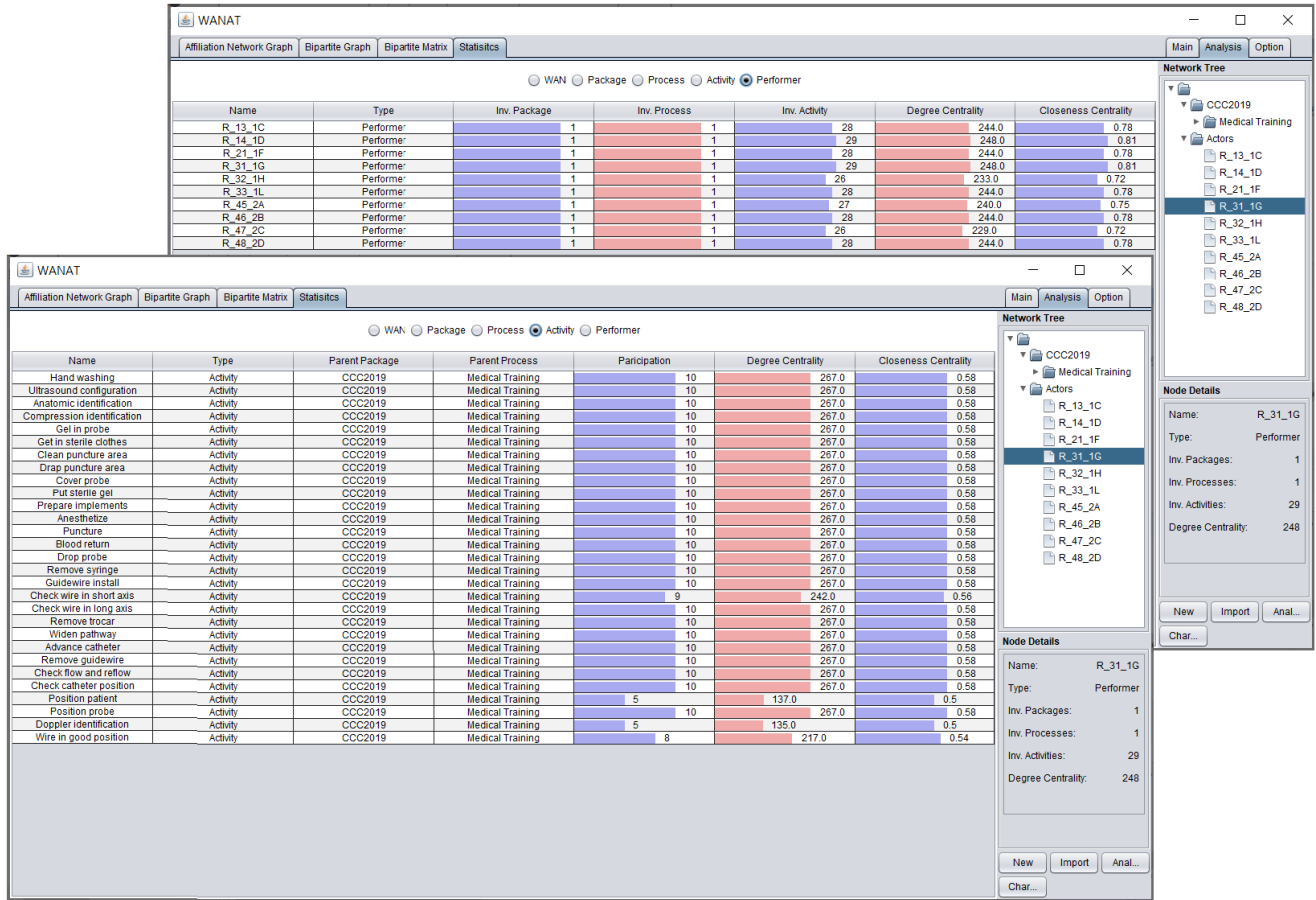


FIGURE 9. The Performer-Centered and the Activity-Centered Organizational Closeness Centrality Measures of the Discovered Affiliation Network.

after importing the discovered XPD L-based workflow process model, the discovery and analysis system discovers the workflow-supported performer-to-activity affiliation network as shown in FIGURE 8. The upper captured-screen in the figure shows a group of red-colored circles (representing activities) affiliated with the specific performer named as “R\_31\_1G,” whereas the lower captured-screen shows a group of blue-colored boxes (representing performers) affiliated with the specific activity titled as “Hand washing.”

2) MEASUREMENTS OF THE ORGANIZATIONAL CLOSNESS CENTRALITIES

For the weighted bipartite graph model of the discovered workflow-supported performer-to-activity affiliation network discovered in the previous subsection, another operational experiment is carried out for measuring the organizational closeness centralities on all the performers (e.g. 10 performers) as well as all the activities (e.g. 29 activities). The operational experiment applies the implemented discovery and analysis system to the weighted bipartite graph in order to measure the performer-centered organizational closeness centralities as well as the activity-centered organizational closeness centralities as followings:

- 1) *Performer-Centered Organizational Closeness Centrality Measurement:* First of all, by using the implemented system, the experiment measured the performer-centered organizational closeness centralities with the performer-centered geodesic distance matrix and the indexes and their standardized indexes of organizational closeness centralities of the performers. The upper captured-screen of FIGURE 9 shows the organizational closeness centrality measures on all the performers in the screen snapshot captured from and produced by the discovery and analysis system, and the left part of TABLE 13 is to list all the standardized indexes of the organizational closeness centralities on all the performers involved in the discovered workflow process model.
  - The measures in the table show that the standardized index of organizational closeness centrality of the performers,  $P_2 : R-14-1D$  and  $P_4 : R-31-1G$ , are 0.81 equally, and the measures are the highest index among all the performers involved in the discovered workflow process model. Through these indexes, it is able to capture the human-centered organizational knowledge that the performers,  $P_2 : R-14-1D$  and  $P_4 : R-31-1G$ , are not only interacting with others

**TABLE 13. The Performer-Centered and the Activity-Centered Organizational Closeness Centrality Measures.**

Performer	$OC_{CP}^S(n_i)$	Activity	$OC_{CA}^S(m_i)$
$P_1 : R - 13 - 1C$	0.78	Advance catheter	0.58
$P_2 : R - 14 - 1D$	0.81	Anatomic identification	0.58
$P_3 : R - 21 - 1F$	0.78	Anesthetize	0.58
$P_4 : R - 31 - 1G$	0.81	Blood return	0.58
$P_5 : R - 32 - 1H$	0.72	Check catheter position	0.58
$P_6 : R - 33 - 1L$	0.78	Check flow and reflow	0.58
$P_7 : R - 45 - 2A$	0.75	Check wire in long axis	0.58
$P_8 : R - 46 - 2B$	0.78	Check wire in short axis	0.56
$P_9 : R - 47 - 2C$	0.72	Clean puncture area	0.58
$P_{10} : R - 48 - 2D$	0.78	Compression identification	0.58
		Cover probe	0.58
		Doppler identification	0.50
		Drap puncture area	0.58
		Drop probe	0.58
		Gel in probe	0.58
		Get in sterile clothes	0.58
		Guidewire install	0.58
		Hand washing	0.58
		Position patient	0.50
		Position probe	0.58
		Prepare implements	0.58
		Puncture	0.58
		Put sterile gel	0.58
		Remove guidewire	0.58
		Remove syringe	0.58
		Remove trocar	0.58
		Ultrasound configuration	0.58
		Widen pathway	0.58
		Wire in good position	0.54

by communicating most directly or through most few intermediaries but also participating most directly to the workflow-activities in performing the instances of the discovered workflow process model. Conclusively speaking, it is analytically predicted that the performers,  $P_2 : R - 14 - 1D$  and  $P_4 : R - 31 - 1G$ , ought to be the most interactive employees with others as well as the most knowledgeable employees on business activities in the enactment histories of all the instances spawned from the discovered workflow model.

2) *Activity-Centered Organizational Closeness Centrality Measurement*: Next, the operation experiment applied the discovery and analysis system again to the weighted bipartite graph model discovered from the CCC2019 dataset, and measured the organizational closeness centralities with the activity-centered geodesic distance matrix and the normalized indexes of organizational closeness centralities of all the activities. The lower captured-screen of FIGURE 9 shows the screen snapshot captured from the discovery and analysis system, on which you can see the measured organizational closeness centralities of all the activities. the right part of TABLE 13 is to list the all the measured standardized indexes of the organizational closeness centralities of all the activities.

– The measures in the table show that almost all activities have the same normalized indexes in

their organizational closeness centralities except four activities, such as Check wire in short axis, Doppler identification, Position patient, and Wire in good position activities, and their index values are 0.56, 0.50, 0.50, and 0.54, respectively. The highest index value among all the 29 activities associated in the discovered workflow process model is 0.58. Through these indexes, it is able to capture the activity-centered organizational closeness centrality knowledge that almost all of the activities ought to be the most valuable workflow-activities in the enactment histories of all the instances spawned from the discovered workflow model.

## VI. CONCLUSION

This paper has investigated the algorithmic formalisms of the organizational closeness centralities for measuring “the degrees of farnesses” on a workflow-supported performer-to-activity affiliation network, after all which will be used for quantifying the work-intimacy (and work-farness) degrees and the work-connectivity degrees between performers as well as between activities, respectively, in a workflow-supported organization. This paper has also introduced a series of numerical formulae and their algorithms of organizational closeness centralities on a workflow-supported performer-to-activity affiliation network and its implications as meaningful mechanisms for human-centered organizational knowledge discovery and

analysis. Particularly, it has restated the mathematical equations for the two types (performer-centered and activity-centered) of organizational closeness centrality measurements, and developed their functional algorithms that can be used for implementing those organizational closeness centrality measurements. In addition, the proposed concept of organizational closeness centrality has been mathematically as well as systematically verified, and its functional equations and algorithms are also validated by carrying out two cases of operational experiments. In the operational experiments, we carried out not only applying the proposed formalisms and algorithms to an information control net of the synthetic five workflow models selected for the operational experiment, but also implementing them as an analytics and visualization system.

As future works, we have plans not only to implement those concepts and algorithms for the organizational closeness centrality measurements as the fundamental analysis functions of a human-centered organizational knowledge and intelligent management system, but also to extend them to be applicable to the advanced properties of the organizational closeness centrality on the workflow-supported performer-to-activity affiliation networks. Furthermore, we will develop a series of equations and algorithms dealing with the remaining centrality types like betweenness and eigenvalue centralities, too.

## ACKNOWLEDGMENT

The authors thank their colleagues, and their institutions for sponsoring this research. Particularly, appreciation is extended to the Kyonggi University Contents Convergence Software Research Institute research professor Dr. Dinh-Lam Pham for his devotional contribution to the implementation of the workflow-supported affiliation network discovery and analysis system based upon the concept and its algorithms proposed in this paper.

## REFERENCES

- [1] W. M. P. van der Aalst, H. A. Reijers, and M. Song, "Discovering social networks from event logs," *Comput. Supported Cooperat. Work (CSCW)*, vol. 14, no. 6, pp. 549–593, Dec. 2005.
- [2] A. Battsetseg, H. Ahn, M. Park, H. Kim, W. Yoon, and K. P. Kim, "Organizational closeness centrality analysis on workflow-supported activity-performer affiliation networks," in *Proc. IEEE Int. Conf. Adv. Commun. Technol.*, IEEE/Global IT Research Institute, Feb. 2013, pp. 154–157.
- [3] K. P. Kim, "Discovering activity-performer affiliation knowledge on ICN-based workflow model," *J. Inf. Sci. Eng.*, vol. 29, pp. 79–97, Jan. 2013.
- [4] H. Oinas-Kukkonen, U. of Oulu, K. Lyytinen, and Y. Yoo, "Social networks and information systems: Ongoing and future research streams," *J. Assoc. Inf. Syst.*, vol. 11, no. 2, pp. 61–68, Feb. 2010.
- [5] A. J. Rembert, "Automatic discovery of workflow models," Ph.D. dissertation, Dept. Comput. Sci., Univ. Colorado Boulder, CO, USA, 2008.
- [6] M. Škerlavaj, V. Dimovski, and K. C. Desouza, "Patterns and structures of intra-organizational learning networks within a knowledge-intensive organization," *J. Inf. Technol.*, vol. 25, no. 2, pp. 189–204, Jun. 2010.
- [7] J. Song, M. Kim, H. Kim, and K. Kim, "A framework: Workflow-based social network discovery and analysis," in *Proc. 4th Int. Workshop Workflow Manage. Service Cloud Comput.*, 2010, pp. 421–426.
- [8] J. Won, "A framework: Organizational network discovery on workflows," Ph.D. dissertation, Dept. Comput. Sci., Kyonggi Univ., Republic Korea, Suwon, South Korea, 2008.
- [9] H. Jeong, H. Kim, and K. P. Kim, "Betweenness centralization analysis formalisms on workflow-supported org-social networks," in *Proc. 16th Int. Conf. Adv. Commun. Technol.*, Feb. 2014, pp. 1168–1172.
- [10] S. Park, H. Jeong, H. Ahn, and K. P. Kim, "A closeness centrality analysis algorithm for workflow-supported social networks," in *Proc. IEEE Int. Conf. Adv. Commun. Technol.*, IEEE/Global IT Research Institute, Feb. 2013, pp. 158–161.
- [11] A. Saxena and S. Iyengar, "Centrality measures in complex networks: A survey," 2020, *arXiv:2011.07190*. [Online]. Available: <http://arxiv.org/abs/2011.07190>
- [12] K. Faust, "Centrality in affiliation networks," *Social Netw.*, vol. 19, no. 2, pp. 157–191, Apr. 1997.
- [13] D. Knoke and S. Yang, *Social Network Analysis*, 2nd ed. Newbury Park, CA, USA: Sage, 2008.
- [14] Z. Shao, N. Guo, and Y. Gu, "Efficient closeness centrality computation for dynamic graphs," in *Database Systems for Advanced Applications (Lecture Notes in Computer Science)*, vol. 12113, Y. Nah, B. Cui, S. W. Lee, J. X. Yu, Y. S. Moon, and S. E. Whang, Eds. Cham, Switzerland: Springer, Sep. 2020, pp. 534–550.
- [15] M. Park, H. Ahn, and K. P. Kim, "Workflow-supported social networks: Discovery, analyses, and system," *J. Netw. Comput. Appl.*, vol. 75, pp. 355–373, Nov. 2016.
- [16] E. Femeley and R. Helms, "Editorial: The special issue on social networking," *J. Inf. Technol.*, vol. 25, no. 2, pp. 107–108, 2010.
- [17] M. Griffiths, "Oxygen: Social intranets, collective intelligence and government practices," *Electron. J. e-Government*, vol. 5, no. 2, pp. 177–190, 2007.
- [18] R. Lui and A. Kumar, "Impact of socio-technical network on process performance," in *Proc. 10th IEEE Int. Conf. Collaborative Comput., Netw., Appl., Worksharing*, 2014, pp. 243–252.
- [19] P. Busch, "Business process management, social network analysis and knowledge management: A triangulation of sorts?" in *Proc. 21st Australas. Conf. Inf. Syst.*, Brisbane, QLD, Australia, 2010, pp. 1–10.
- [20] Y. Jia, S. Ge, and N. Wang, "Analyzing enterprise information system's feature use: A data-driven perspective," *Inf. Technol. People*, vol. 34, no. 1, pp. 375–398, Apr. 2020.
- [21] K. Kim, "Actor-oriented workflow model," in *Proc. 2nd Int. Symp. Cooperat. Database Syst. Adv. Appl.*, Mar. 1999, pp. 27–28.
- [22] S. Poelmans, H. A. Reijers, and J. Recker, "Investigating the success of operational business process management systems," *Inf. Technol. Manage.*, vol. 14, no. 4, pp. 295–314, Dec. 2013.
- [23] M. P. Bach, V. B. Vukšić, D. S. Vugec, and A.-M. Stjepic, "BPM and BI in SMEs: The role of BPM/BI alignment in organizational performance," *Int. J. Eng. Bus. Manage.*, vol. 11, no. 5, pp. 1–16, Sep. 2019.
- [24] H. Ahn, D.-L. Pham, and K. P. Kim, "An experimental analytics on discovering work transference networks from workflow enactment event logs," *Appl. Sci.*, vol. 9, no. 11, pp. 2368–2390, Jun. 2019.
- [25] H. Ahn and K. P. Kim, "Formal approach for discovering work transference networks from workflow logs," *Inf. Sci.*, vol. 515, pp. 1–25, Apr. 2020.
- [26] D.-L. Pham, H. Ahn, and K. P. Kim, "Discovering redo-activities and performers' involvements from xes-formatted workflow process enactment event logs," *KSII Trans. Internet And Inf. Syst.*, vol. 13, no. 8, pp. 4108–4122, 2019.
- [27] A. Senderovich, J. J. Schippers, and H. A. Reijers, "Socially-aware business process redesign," in *Business Process Management (Lecture Notes in Computer Science)*, vol. 12168, D. Fahland, C. Ghidini, J. Becker, and M. Dumas, Eds. Cham, Switzerland: Springer, Sep. 2020, pp. 75–92.
- [28] H. Kim, H. Ahn, and K. P. Kim, "A workflow affiliation network discovery algorithm," *ICIC Exp. Lett.*, vol. 6, no. 3, pp. 765–770, 2012.
- [29] H. Jeong, H. Ahn, and K. P. Kim, "A workflow-supported social networking knowledge visualization system," in *Proc. Int. Conf. Internet*, Korean Society for Internet and Information, Dec. 2013, pp. 187–193.
- [30] I. Ra and K. P. Kim, "ccWSSN-GraphML: An extended graph markup language for visualizing closeness-centrality measurements of workflow-supported org-social networks," in *Proc. KSII 8th Asia Pacific Int. Conf. Inf. Sci. Technol. (APICIST)*, Korean Society for Internet and Information, Jun. 2013, pp. 97–100.
- [31] M.-J. Kim, H. Ahn, and M.-J. Park, "A GraphML-based visualization framework for workflow-performers' closeness centrality measurements," *KSII Trans. Internet Inf. Syst.*, vol. 9, no. 8, pp. 3216–3230, 2015.
- [32] M.-J. Kim, H. Ahn, and M.-J. Park, "A theoretical framework for closeness centrality measurements in a workflow-supported organization," *KSII Trans. Internet Inf. Syst.*, vol. 9, no. 9, pp. 3611–3634, 2015.

- [33] H. Kim and K. P. Kim, "Formulating closeness centralities on workflow-supported performer-activity affiliation networks," *ICACT Trans. Adv. Commun. Technol.*, vol. 5, no. 6, pp. 960–965, 2016.
- [34] H. Kim, H. Ahn, and K. P. Kim, "Modeling, discovering, and visualizing workflow performer-role affiliation networking knowledge," *KSII Trans. Internet Inf. Syst.*, vol. 8, no. 2, pp. 691–708, 2014.
- [35] H. Kim, H. Ahn, and K. P. Kim, "Discovering workflow performer-application affiliation knowledge and its implications," in *Proc. Int. Conf. Innov. Comput., Inf., Control*, 2015, pp. 1049–1056.
- [36] K. P. Kim, "A workflow performer-activity affiliation networking knowledge discovery system," *J. Amer. Acad. Bus.*, vol. 19, no. 2, pp. 172–178, 2014.
- [37] H. A. Reijers, M. Song, and B. Jeong, "Analysis of a collaborative workflow process with distributed actors," *Inf. Syst. Frontiers*, vol. 11, no. 3, pp. 307–322, 2009.
- [38] A. Kumar, R. Dijkman, and M. Song, "Optimal resource assignment in workflows for maximizing cooperation," in *Business Process Management (Lecture Notes in Computer Science)*, vol. 8094, F. Daniel, J. Wang, and B. Weber, Eds. Berlin, Germany: Springer, 2013, pp. 235–250.
- [39] E. Kim and P. Busch, "Workflow interpretation via social networks," in *Knowledge Management and Acquisition for Intelligent Systems (Lecture Notes in Computer Science)*, vol. 9806, H. Ohwada and K. Yoshida, Eds. Cham, Switzerland: Springer, Aug. 2016, pp. 241–250.
- [40] M. Jalili, "Graph theoretical analysis of Alzheimer's disease: Discrimination of ad patients from healthy subjects," *Inf. Sci.*, vol. 384, pp. 145–156, Apr. 2017.
- [41] M. Lozano, C. Harcia-Matrinez, F. J. Rodriguez, and H. M. Trujillo, "Optimizing network attacks by artificial bee colony," *Inf. Sci.*, vol. 377, pp. 30–50, Jan. 2017.
- [42] E. Xamena, N. B. Brignole, and A. G. Maguitman, "A structural analysis of topic ontologies," *Inf. Sci.*, vol. 421, pp. 15–29, Dec. 2017.
- [43] P. Chamoso, A. Rivas, S. Rodríguez, and J. Bajo, "Relationship recommender system in a business and employment-oriented social network," *Inf. Sci.*, vols. 433–434, pp. 204–220, Apr. 2018.
- [44] Z. Ghalmane, C. Cherifi, H. Cherifi, and M. E. Hassouni, "Centrality in complex networks with overlapping community structure," *Sci. Rep.*, vol. 9, no. 1, pp. 1–29, Jul. 2019.
- [45] B. Aronson, K.-C. Yang, M. Odabas, Y.-Y. Ahn, and B. L. Perry, "Comparing measures of centrality in bipartite social networks: A study of drug seeking for opioid analgesics," *SocArXiv*, vol. 82, pp. 1–47, Jun. 2020.
- [46] H. Ahn, C. Park, and K. Kim, "A BPM activity-performer correspondence analysis method," *J. Korean Soc. Internet Inf.*, vol. 14, no. 4, pp. 63–72, Aug. 2013.
- [47] C. A. Ellis, K. Kim, A. Rembert, and J. Wainer, "Investigations on stochastic information control nets," *Inf. Sci.*, vol. 194, pp. 120–137, Jul. 2012.
- [48] C. A. Ellis and G. J. Nutt, "Office information systems and computer science," *ACM Comput. Surv.*, vol. 12, no. 1, pp. 27–60, Mar. 1980.
- [49] C. A. Ellis, "Formal and informal models of office activity," in *Proc. World Comput. Congr.*, 1983, pp. 11–22.
- [50] *IEEE Standard for Extensible Event Stream (XES) for Achieving Interoperability in Event Logs and Event Streams*, Standard IEEE 1849-2016, 2106.
- [51] *Bpi Challenges and Conformance Checking Challenge*, TU Center for Research Data, BPM, Boston, MA, USA, 2019.



**HYUN AHN** received the B.S. and M.S. degrees and the Ph.D. degree in computer science from Kyonggi University, South Korea, in 2011, 2013, and 2017, respectively. He is currently an Assistant Professor with the Department of Software Convergence, Hanshin University, South Korea. He is also a Research Professor with the Division of Computer Science and Engineering as well as of the Contents Convergence Software Research Institute, Kyonggi University. His research interests include process-aware enterprise information systems, business process intelligence, process mining, predictive process monitoring, and deep learning and knowledge engineering using active contents big data.



**KWANGHOON PIO KIM** received the B.S. degree in computer science from Kyonggi University, South Korea, in 1984, the M.S. degree in computer science from Chungang University, in 1986, and the M.S. and Ph.D. degrees from the Computer Science Department, University of Colorado Boulder, in 1994 and 1998, respectively. He is currently a Full Professor with the Division of Computer Science and Engineering as well as the Founder and a Supervisor of the Data and Process Engineering Research Laboratory, Kyonggi University. At Kyonggi, he has been in charge of the Dean of the Computerization and Informatics Institute and the Director of the Contents Convergence Software Research Institute that is newly designated as a National Science and Engineering Research Institute, where he is leading and fulfilling a multi-million dollars research project during 2020–2029 fully funded by the Korea National Research Foundation of the Education Ministry of South Korea. He had worked as a Researcher and a Developer at Aztek Engineering, American Educational Products Inc., and IBM, USA, as well as the Electronics and Telecommunications Research Institute (ETRI), South Korea. His research interests include groupware, workflow and business process management systems, BPM, CSCW, collaboration theory, Grid/P2P distributed systems, process warehousing and mining, predictive process monitoring, workflow-supported social networks discovery and analysis, process-aware information systems, data intensive workflows, process-aware Internet of Things, predictive process modeling, process deep-learning, and active contents big data engineering and applications supporting crime prevention and prediction. He is a Vice-Chair of the BPM Korea Forum. He has been in charge of a country-chair (South Korea) and ERC vice-chair of the Workflow Management Coalition. He has also been on the editorial board of the journal of *KSII*, and the committee member of the several conferences and workshops.

• • •