Organizational Closeness Centrality Analysis on Workflow-supported Activity-Performer Affiliation Networks

Alyeksandr Battsetseg*, Hyun Ahn*, Yoonjoo Lee*, Minjae Park**, Hyuna Kim***, Wonhyun Yoon*, Kwanghoon Pio Kim*

*Dept. of Computer Science, KYONGGI UNIVERSITY, Suwonsi Kyonggido, Korea
**Bistel, LTD., Seoul, Korea
***WoToWiTo, LTD., Suwonsi Kyonggido, Korea
{alexander, hahn, whyoon, kwang}@kgu.ac.kr, mjpark@bistel-inc.com, hakim@wotowito.com

***Abstract*—This paper analyzes a special type of social networks, which is called 'workflow-supported activity-performer affiliation network.' A workflow model specifies execution sequences of the associated activities and their affiliated relationships with roles, performers, invoked-applications, and relevant data. Especially, these affiliated relationships exhibit a series of valuable organizational knowledge and utilize to explore business intelligence concealed in the corresponding workflow model. In this paper, we particularly focus on analyzing the affiliated relationships between activities and performers in a workflow model by measuring the organizational closeness centralities of performers as well as the organizational closeness centralities of activities. We devise a series of algorithms for analyzing the closeness centralities of activities and performers, and describe the ultimate implications of these analysis results as activity-performer affiliation knowledge in workflow-supported organizations.**

Keywords: workflow-supported affiliation network, ICN-based workflow model, organizational closeness centrality, business process intelligence

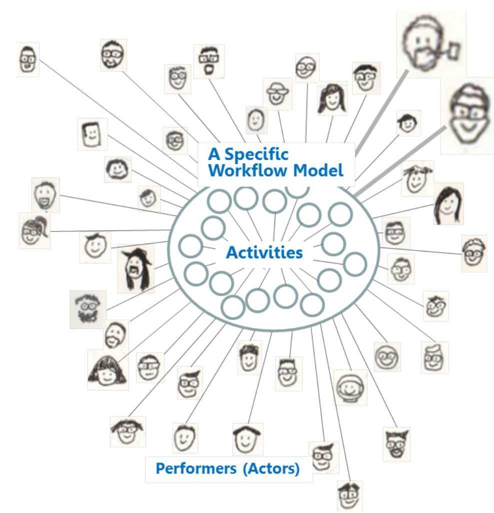

Fig. 1. The Workflow-supported (Activity-Performer) Affiliation Knowledge

I. Introduction

In recent, 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 [12][13], in which the authors formalize a mechanism and its related algorithms to discover workflow-supported affiliation networking knowledge from a conventional workflow model. In a workflow model, the performers (or actors) are linked in activities through joint participation; conversely, the activities are connected to performers through joint involvement; they have dubbed a collection of these links and connections "*workflow-supported performer-activity affiliation network model*."

In this paper, we focus on quantitatively measuring the organizational closeness centralities of workflow-supported affiliation network models, each of which is formed by two key groups of the elements, a set of performers and a collection of activities, as shown in Fig. 1. That is, we are basically concerned about organizational centralities in those affiliation networks, which are coming from the formal properties—two-mode and non-dyadic networks—of affiliation networks. Since those affiliation networks are two-mode networks, a complete measurements should be done by giving centrality indices for both performers and activities. Traditionally, there are four organizational centrality analysis techniques, such as degree, closeness, betweenness, and eigenvector centralities, and we particularly measure the closeness centralities of performers and the closeness centralities of activities in this paper.

By analyzing the organizational closeness centralities from activity-performer affiliated relationships, it is ultimately possible to visualize and numerically express how much the performers and the activities are interrelated and collaboratively closed in enacting a specific workflow model. Ultimately, the main purpose of this paper is to theoretically develop an algorithm for calculating the organizational closeness centralities for activities and performers in a 'workflwo activity-performer affiliation network.

II. Activity-Performer Affiliation Network Model

In order to represent the workflow-supported activity-performer affiliation knowledge, [13] has recently defined a graphical (Bipartite Graph) and formal representation model, which is dubbed activity-performer affiliation network model. An activity-performer affiliation network model, which is abbreviated as APANM, consists of two types of nodes—a set of performers and a set of activities—and a set of relationships between those nodal types. Thus, an activity-performer affiliation network model is a two-mode network model with aiming to accomplish 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-performer affiliation networks—binary activity-performer affiliation network and valued activity-performer affiliation network. In the binary activity-performer affiliation network, its value (0 or 1) implies a binary relationship of involvement (or participation), while values in the valued activity-performer affiliation network may represent various implications according to their application domains; typical examples of values might be stochastic (or probabilistic) values, strengths, and frequencies. The formal knowledge representation of activity-performer affiliation network model is defined in the following [Definition 1].

[Definition 1] Activity-performer Affiliation Network Model. An activity-performer affiliation network model is formally defined as $\Lambda = (\sigma, \psi, \mathbf{S})$, over a set $\mathbf{C}$ of performers (actors), a set $\mathbf{A}$ of activities, a set $\mathbf{V}$ of weight-values, a set $\mathbf{E}_p \subseteq (\mathbf{C} \times \mathbf{A})$ of edges (pairs of performers and activities), and a set $\mathbf{E}_a \subseteq (\mathbf{A} \times \mathbf{C})$ of edges (pairs of activities and performers), where, $\wp(\mathbf{A})$ represents a power set of the activity set, $\mathbf{A}$:

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

Furthermore, the affiliation knowledge representation can be graphically depicted by an affiliation graph. So, an activity-performer affiliation network's graphical model 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 non-directed edges between two nodal types, which means that a workflow affiliation network is a non-directed graph. That is, in an activity-performer affiliation graph, non-directed lines connect performers aligned on one side of the diagram to the workflow activities aligned on the other side. Importantly, an activity-performer affiliation graph does not permit lines among the performers nor among the workflow activities. Therefore, an activity-performer affiliation graph with g performers and h workflow activities can be transformed into a matrix with 2-dimension of $g \times h$.

III. Organizational Closeness Centrality Analysis

In general, an affiliation networking graph 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. Based upon the activity-performer affiliation networking graph and its affiliation matrix, it is possible to analyze a variety of knowledge analytics issues[11], such as mean rates analysis[6], density measurements[6], centrality measurements[11], and so on, raised from the social networking literature. In this paper, our focus concentrates upon the centrality measurements of the workflow-supported affiliation network model. More precisely speaking, we try to propose an algorithmic formalism for analyzing organizational centrality measurements, particularly closeness centrality measurements, of a workflow-supported performer-activity affiliation network.

A. Definition of Affiliation Matrix

Eventually, it is necessary for activity-performer affiliation network model to be analyzed in a mathematical representation. An activity-performer affiliation network model is graphically represented by a bipartite graph, and at the same time it is mathematically represented by an affiliation matrix. The affiliation matrix can be realized by either an involvement matrix or a participation matrix. That is, an activity-performer affiliation network model is mathematically transformed into an activity-performer affiliation 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 ϕ_i attends a workflow activity α_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 activity-performer affiliation matrix as $\mathbf{Z}$, its $x_{i,j}$ values meet these conditions:

$$x_{i,j} = \begin{cases} \mathbf{1} & \textit{if performer}, \phi_i, \textit{is affiliated with activity}, \alpha_j \\ \mathbf{0} & \textit{otherwise} \end{cases} \tag{1}$$

- The row total, also called row marginals, ($\overline{D}_r$), of activity-performer affiliation matrix $\mathbf{Z}$ 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.

$$\overline{D}_r = \left[\sum_{j=1}^{h} x_{i,j} \right]_{i=1}^{g} \tag{2}$$

- The column marginals, ($\overline{D}_c$), 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.

$$\overline{D}_c = \left[\sum_{i=1}^{g} x_{i,j} \right]_{j=1}^{h} \tag{3}$$

Also, assuming an affiliation networking graph has g performers and h activities, then its bipartite affiliation matrix has dimensions $(g + h) \times (g + h)$. Consequently, using the involvement affiliation matrix ($\mathbf{Z}_p$) and the participation affiliation matrix ($\mathbf{Z}_a$) forms an affiliation bipartite matrix, $\mathbf{X}^{P,A}$, which can be schematically represented as the following equations, (4) and (5).

$$\mathbf{X}^{P,A} = \begin{bmatrix} \mathbf{0} & \mathbf{Z}_p \\ \mathbf{Z}_a & \mathbf{0} \end{bmatrix} \tag{4}$$

$$\mathbf{X^P} = \mathbf{Z_p} \cdot \mathbf{Z_a} \qquad \mathbf{X^A} = \mathbf{Z_a} \cdot \mathbf{Z_p}; \tag{5}$$

B. Organizational Closeness Centrality

[11] gives a series of well-described equations that can be applied to calculating the organizational closeness centralities based upon the bipartite matrix of a workflow-supported activity-performer affiliation network model. Before we develop an algorithm of the organizational closeness centrality measurements in the next subsection, we need to restate those closeness centrality equations, and 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.

Basically, the meaning of closeness centrality index in a social network implies the average geodesic distance that a node is from all other nodes in the graph. In other words, it is to calculate the 'farness' of a node from other nodes in the graph. As described in the previous section, the performer-activity affiliation network is a special type of social networks, and it is represented by a bipartite graph with relationships (or connections) between performers and activities. Thus, calculating the geodesic distances in a bipartite graph begins with a function of the distances from the activities to the performers which each of them belongs. The distance from a node i representing a performer to any node j (either performer or activity) is $d(i, j) = 1 + \min\{d(k, j)\}_k$, for every activity node k adjacent to i. Given this properties, the organizational closeness centrality of a performer in the bipartite graph can be expressed with a function of the distances from the performer's activities, k:

$$\sum_{j=1}^{g+h} d(i,j) = \sum_{j=1}^{g+h} \left[1 + \min\{d(k,j)\}_k\right], i \neq j \tag{6}$$

1) Organizational Closeness Centrality of Performers:

Based on the distance function of (6), the following expressions are the index and the standardized index of the organizational closeness centrality of a performer with a function of the minimum geodesic distances from its activities to other actors and to other activities, respectively. Note that every activity n_a is adjacent to performer n_i.

- The Index of Organizational Closeness Centrality of Performers

$$OC_C(n_i) = \left[\sum_{j=1}^{g+h} d(i,j) \right]^{-1} \quad (i \neq j) \tag{7}$$

$$OC_C(n_i) = \left[1 + \sum_{j=1}^{g+h} \min\{d(n_a, n_j)\}_a \right]^{-1} \quad (i \neq j) \tag{8}$$

- The Normalized Index of Organizational Closeness Centrality of Performers

$$OC_C^S(n_i) = (g + h - 1) \cdot [OC_C(n_i)] \tag{9}$$

$$OC_C^S(n_i) = \left[1 + \frac{\sum_{j=1}^{g+h} \min\{d(n_a, n_j)\}_a}{g + h - 1} \right]^{-1} \quad (i \neq j) \tag{10}$$

2) Organizational Closeness Centrality of Activities:

By revising the distance function of (6), it is also necessary to make the expressions for the index and the standardized index of the organizational closeness centrality of an activity with a function of the minimum geodesic distances from its performers to other activities and to other performers. Note that every performer m_p is adjacent to activity m_j.

- The Index of Organizational Closeness Centrality of Activities

$$OC_C(m_i) = \left[1 + \sum_{j=1}^{g+h} \min\{d(m_p, m_j)\}_p \right]^{-1} \quad (i \neq j) \tag{11}$$

- The Normalized Index of Organizational Closeness Centrality of Activities

$$OC_C^S(m_i) = \left[1 + \frac{\sum_{j=1}^{g+h} \min\{d(m_p, m_j)\}_p}{g + h - 1} \right]^{-1} \quad (i \neq j) \tag{12}$$

The equations (10) and (12) are for normalizing the index of organizational closeness centrality by multiplying by $(g+h-1)$. Suppose that a performer is close to all others, which means that its adjacent activity has a direct tie to every performer in the bipartite graph. Thus the computed index values will be vary according to their graph sizes. In order to control the size of the graph, it is necessary for the individual index to be normalized so as to allow meaningful comparisons of performers across different graphs. This explanation can be identically applied to the normalized index for activities.

C. Algorithm for the geodesic distances

Based upon the equations (10) and (12), we can calculate the normalized indices of organizational closeness centralities of performers and activities, respectively, for a workflow-supported performer-activity affiliation network. As you know, the essential part of those equations must be the function of the geodesic distances, $d(i,j) = 1 + \min\{d(k,j)\}_k$. In this subsection, we devise an algorithm with a recursive function, which is able to compute the geodesic distances for all of the performers. The algorithm operates on the affiliation bipartite matrix, $\mathbf{X}^{P,A}$, and its detailed function (named '*gDistance*') is the followings:

The Geodesic Distances Algorithm for a Performer:

Global A Binary Affiliation Bipartite Matrix, $X^{P,A}[g+h, g+h]$;
Global A Set of Performers, **P**;
Global A Set of Activities, **A**;

Procedure Name: AllgDistances
Input A Binary Affiliation Bipartite Matrix, $X^{P,A}[g+h, g+h]$;
A Performer, n_i;
Output A Set of gDistances, $G_d[g+h-1]$, from n_i to all others;
Begin Procedure
Local A Distance Vector, $G_d[g+h-1]$;

For ($\forall n_g \in$ **P**, $n_i \neq n_g$)
$G_d[g] \leftarrow$ **gDistance**(n_i, n_g);
Rof
For ($\forall m_h \in$ **A**)
$G_d[h] \leftarrow$ **hDistance**(n_i, m_h);
Rof
Return $G_d[s], s = 1..(g+h-1), \sum_{s=1}^{g+h-1} G_d[s]$;
End Procedure

Procedure Name: gDistance
Input The input performer, n_i, and the destination performer, n_j;
Output The minimum geodesic distance between n_i and n_j;
Begin Procedure
Local An Activity Distance Vector, $G_k[h]$;
Local A Performer Distance Vector, $H_k[g]$;
For ($\forall m_k \in$ **A**, m_k adjacent to n_i)
Switch $(X^{P,A}[m_k, n_j])$
case 1: /* direct tie between m_k and n_j. */
$G_k[m_k] \leftarrow 2$;
break;
case 0: /* no direct tie between m_k and n_j. */
For ($\forall n_s \in$ **P**, n_s adjacent to m_k, $n_s \neq n_i$)
$H_k[n_s] \leftarrow 2 +$ **gDistance**(n_s, n_j);
Rof
$G_k[m_k] \leftarrow$ Minimum$\{\forall H_k[n_s]\}$;
break;
Rof
If ($\forall G_k[1..h] = zero$)
Return $zero$;
Else
Return Minimum$\{G_s \subseteq G_k[h], \forall\alpha(\in G_s) \neq zero\}$;
End Procedure

Procedure Name: hDistance
Input The input performer, n_i, and the destination activity, m_j;
Output The minimum geodesic distance between n_i and m_j;
Begin Procedure
Local A Distance Vector, $G_k[h]$;
Local A Performer Distance Vector, $H_k[g]$;
For ($\forall n_k \in$ **P**, n_k adjacent to m_j)
Switch $(X^{P,A}[n_k, m_j])$
case 1: /* direct tie between n_k and m_j. */
$H_k[n_k] \leftarrow 1$;
break;
case 0: /* no direct tie between n_k and m_j. */
For ($\forall m_s \in$ **A**, m_s adjacent to n_k)
$G_k[m_s] \leftarrow 1 +$ **hDistance**(m_s, m_j);
Rof
break;
Rof
If ($\forall H_k[1..g] = zero$)
Return $zero$;
Else
Return Minimum$\{G_s \subseteq H_k[g], \forall\alpha(\in G_s) \neq zero\}$;
End Procedure

IV. Conclusion

In this paper, we propose an algorithmic description for analyzing organizational closeness centralities of a workflow-supported performer-activity affiliation network model representing involvement and participation behaviors between workflow-based people and workflow-based activities. We have introduced the basic concept of workflow-supported affiliation network and its implications as a meaningful mechanism of organizational knowledge and intelligence. Particularly, we restate the mathematical equations for the organizational closeness centrality measurements, and develop an functional algorithm for implementing those organizational closeness centrality equations. As a future work, we have a plan to implement those concept and algorithms for organizational closeness centralities as a fundamental function of the organizational knowledge and intelligent management system.

Acknowledgement

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology (Grant No. 2012006971).

References

[1] Clarence A. Ellis, Gary J. Nutt, "Office Information Systems and Computer Science," ACM Computing Surveys, Vol. 12, No. 1, 1980
[2] Kwanghoon Kim, "Actor-oriented Workflow Model," Proceedings of the 2nd international symposium on Cooperative Database Systems for Advanced Applications, WOLLONGONG, AUSTRALIA, March 27-28, 1999
[3] Harri Oinas-Kukkonen, et al., "Social Networks and Information Systems: Ongoing and Future Research Streams," JOURNAL OF THE ASSOCIATION OF INFORMATION SYSTEMS, Vol. 11, Issue 2, pp. 61-68, 2010
[4] Mary Griffiths, "Oxygen: Social Intranets, collective intelligence and Government Practices," ELECTRONICS JOURNAL OF E-GOVERNMENT, Vol. 5, Issue 2, pp. 177-190, 2007
[5] Wil M. P. van der Aalst, Hajo A. Reijers, Minseok Song, "Discovering Social Networks from Event Logs," COMPUTER SUPPORTED COOPERATIVE WORK, Vol. 14, No. 6, pp. 549-593, 2005
[6] David Knoke, Song Yang, SOCIAL NETWORK ANALYSIS - 2^{nd} Edition, *Series: Quantitative Applications in the Social Sciences*, SAGE Publications, 2008
[7] Jaekang Won, "A Framework: Organizational Network Discovery on Workflows," *Ph.D. Dissertation*, Department of Computer Science, KYONGGI UNIVERSITY, 2008
[8] Jihye Song, Minjoon Kim, Haksung Kim, Kwanghoon Kim, "A Framework: Workflow-based Social Network Discovery and Analysis," Proceedings of the International Workshop on Hongkong, China, December, 2010
[9] E. Ferneley, R. Helms, "Editorial of the Special Issue on Social Networking," Journal of Information Technology, Vol.25, No.2, pp. 107-108, 2010
[10] Miha Skerlavaj, Vlado Dimovski, Kevin C Desouza, "Patterns and structures of intra-organizational learning networks within a knowledge-intensive organization," Journal of Information Technology, Vol. 25, No. 2, pp. 189-204, June 2010
[11] Katherine Faust, "Centrality in Affiliation Networks," Journal of Social Networks, Vol. 19, pp. 157-191, 1997
[12] Haksung Kim, Hyun Ahn, Kwanghoon Pio Kim, "A Workflow Affiliation Network Discovery Algorithm," ICIC Express Letters, Vol. 6, No. 3, pp. 765-770, 2012
[13] Kwanghoon Pio Kim, "Discovering Activity-Performer Affiliation Knowledge on ICN-based Workflow Model," Journal of Information Science and Engineering, Vol. 29, pp. 79-97, 2013