A Closeness Centrality Analysis Algorithm for Workflow-supported Social Networks

Sungjoo Park*, Minjae Park**, Hyuna Kim***, Haksung Kim****, Wonhyun Yoon*, Thomas B. Yoon*, Kwanghoon Pio Kim*

*Dept. of Computer Science, KYONGGI UNIVERSITY, Suwonsi Kyonggido, Korea
**Bistel, LTD., Seoul, Korea
***WoToWiTo, LTD., Suwonsi Kyonggido, Korea
****Dongnam Health University, Suwonsi Kyonggido, Korea

{npk1234, whyoon, toma, kwang}@kgu.ac.kr, mjpark@bistel-inc.com, hakim@wotowito.com, amang@dongnam.ac.kr

***Abstract*—The purpose of this paper is to build a theoretical approach for numerically analyzing closeness centrality measures among workflow-actors of workflow-supported social network models to be formed through BPM(workflow)-driven organizational operations. The essential part of the proposed approach is a closeness centrality analysis equation to calculate each performer's closeness centrality measure on a workflow-supported social network model. In this paper, we try to develop an algorithm that is able to efficiently compute the closeness centrality analysis equation suggested from the conventional social network analysis literature, and eventually the developed algorithm will be applied to analyzing the degree of work-intimacy among those workflow-actors who are allocated to perform the corresponding workflow model.**

Keywords: workflow-supported social networking knowledge, ICN-based workflow model, closeness centrality analysis, organizational knowledge discovery

I. Introduction

Recently, the workflow literature just starts being focused on "People"[1][3]. It starts from the strong belief that social relationships and collaborative behaviors among people who are involved in enacting the specific workflow models affect the overall performance and being crowned with great successes in the real businesses and the working productivity as well. So, research and development issues of applying the concept of social network and its analysis methods to workflow models have been emerging in the literature. There have been existing two main branches of research approaches in adopting social network techniques onto workflow-supported organizations; One is so-called workflow-supported social network *discovery* issues, the other has something to do with so-called workflow-supported social network *rediscovery* issues. The latter is concerned with mining social networking knowledge from workflow event logs, which was firstly issued by [2]; the former is to discover social network knowledge through exploring the human perspective of a group of workflow models, which was issued at first by [7].

The research group of the authors has introduced the basic concept of workflow-supported social networks in [7], and its related formalisms have also been defined through the frameworks[6][7][8] providing organizational knowledge discovery methodologies. However, those frameworks cope only with a bare couple of formalisms, like a degree-centrality analysis technique, in terms of analyzing the workflow-supported social networking knowledge and its representative models. It ought to be definitely necessary for the frameworks to be equipped with more sophisticated and diversified analysis techniques, such as closeness-centrality, betweenness-centrality, eigenvalue-centrality, correspondence analysis, and so on, in order to be practically applied into a real organizational world.

In this paper, we try to conceive an algorithmic formalism of closeness-centrality measurements to quantitatively analyze workflow-supported social networking knowledge and models. The eventual goal of the formalism proposed in this paper is to numerically measure and calculate the degree of work-intimacy among employees involved in a workflow model or a workflow package (a group of inter-relevant workflow models) on a workflow-driven organizational environment.

II. Formal Definition of Workflow-supported Social Networks

In this subsection, we start from introducing the basic concept and definition of workflow-supported social network model that can be used for a knowledge representation theory of workflow-supported social networking knowledge that might be either discovered from workflow models or rediscovered from workflow execution logs. Basically, the origin of the workflow-supported social network model is the actor-based workflow model[1], and its rationale is on where it represents the behaviors of acquisitioning activities among actors in a workflow model, which we would call workflow-supported social relationships that form this special type of social networks.

As given in the formal definition, [Definition 1], of the workflow-supported social network model, the behaviors of the model are revealed through incoming and outgoing directed arcs labeled with activities associated with each of actors. The directed arcs imply two kinds of behaviors—workflow-supported social relationships and activity acquisition of actors—through which we are able to get precedence (candidate-predecessor knowledge/candiate-successor knowledge) knowledge among actors as well as activity acquisition of each actor in a workflow model. In terms of defining actor's predecessors and successors, we would use the prepositional word,"candidate", because a role-actor mapping is an one-to-many relationship knowledge, and the actor selection mechanism will choose one actor out of the assigned actors mapped to the corresponding role during the underlying workflow model's runtime.

[Definition 1] Workflow-supported Social Network Model. A Workflow-supported Social Network Model is formally defined as $\Lambda = (\sigma, \psi, \mathbf{S}, \mathbf{E})$, over a set $\mathbf{C}$ of performers, and a set $\mathbf{A}$ of activities, where

- $\mathbf{S}$ is a finite set of coordinators or coordinator-groups connected from some external workflow-supported social network models;
- $\mathbf{E}$ is a finite set of coordinators or coordinator-groups connected to some external workflow-supported social network models;
- $\sigma = \sigma_i \cup \sigma_o$ /* Social Relationships: successors and predecessors */
  where, $\sigma_o : \mathbf{C} \longrightarrow \wp(\mathbf{C})$ is a multi-valued function mapping a performer to its sets of (immediate) candidate-successors, and $\sigma_i : \mathbf{C} \longrightarrow \wp(\mathbf{C})$ is a multi-valued function mapping a performer to its sets of (immediate) candidate-predecessors;
- $\psi = \psi_i \cup \psi_o$ /* Acquisition of Activities */
  where, $\psi_i : \mathbf{C} \longrightarrow \wp(\mathbf{C})$ is a multi-valued function returning a bag[1] of previously worked activities, $(\mathbf{K} \subseteq \mathbf{A})$, on directed arcs, $(\sigma_i(o), o)$, $o \in \mathbf{C}$, from $\sigma_i(o)$ to o; and $\psi_o : \mathbf{C} \longrightarrow \wp(\mathbf{C})$ is a multi-valued function returning a set of acquisition-activities, $(\mathbf{K} \subseteq \mathbf{A})$, on directed arcs, $(o, \sigma_o(o))$, $o \in \mathbf{C}$ from o to $\sigma_o(o)$;

In principle, a workflow-supported social network model is graphically represented by a directed graph characterized by multiple-incoming arcs, multiple-outgoing arcs, cyclic, self-transitive, and multiple-activity associations on arcs. Additionally, it can be also transformed to an undirected graph for analyzing closeness centralities among the associated performers.

III. Algorithmic Formalism for Closeness Centrality Measurements

There exist several social network analysis techniques and algorithms for the centrality measures. Particularly among them, the most widely used centrality measures are degree, closeness, and betweenness, and these measures not only vary in their applicability to nondirected and directed relations, but also differ at the individual actor and the group or complete network levels. As stated in the previous section, we are interested in quantitatively measuring the degree of closeness centrality of a workflow-supported social network by borrowing the well-known techniques[4] in the social network analysis literature. The analyzed measurements of closeness centrality reflect how near a performer is to the other performers in a workflow-supported social network. Ultimately, the implication of the algorithmic formalism to be deployed in this section aims to answer for the following question:

- How quickly can a performer interact with others in enacting the associated workflow procedure by communicating directly or through very few intermediaries?

A. SocioMatrix

In principle, the social network analysis literature suggests two classes—*binary directed/nondirected SocioMatrix* and *valued directed/nondirected SocioMatrix*—of SocioMatrices and graphs to analyze cognitive social structures, and it uses them to construct a sociogram[4] that is a two-dimensional diagram for displaying the relations among nodes in a bounded social system. The term, *directed*, indicates directed relations or ties from the node at the tail to the node at the arrowhead (e.g. giving advice); while the term, *nondirected* (no arrowheads), implies mutual relations. Likewise, when a directed/nondirected graph is transformed to a SocioMatrix, the term, *binary*, implies the most basic measurement, the presence or absence of a tie, which is a dichotomy indicated by binary values of 1 and 0, respectively; also SocioMatrices may include *nonbinary* or *valued* cells, reflecting the intensity of relations or ties, such as frequency of contacts, tie strength, or magnitude of associations, and therefore the cell entries in SocioMatrix can vary from 0 to the maximum level of dyadic interactions.

The authors' research group had devised a series of algorithms[7][8] that is able to transform a workflow-supported social network model to any possible types of SocioMatrices. In this paper we simply introduce one of the algorithms as followings, which produces a binary nondirected SocioMatrix, $\mathbf{Z}[N, N]$, where N is the number of performers, from a formally defined workflow-supported social network model.

Binary NonDirected SocioMatrix Generation Algorithm
Input A Workflow-supported Social Network, $\Lambda = (\sigma, \psi, \mathbf{S}, \mathbf{E})$;
Output A Binary Nondirected SocioMatrix, $\mathbf{Z}[N, N]$, where N is the number elements in the set of $\mathbf{C}$ performers.
Begin Procedure
  Initialize all entries of $\mathbf{Z}[N, N]$ **To** $Zeroes$;
  For ($\forall o \in \mathbf{C}$) **Do**
    Begin
      /* Set the Incoming Relations to $\mathbf{Z}[N, N]$ */
      Set One **To** *entries of* $\mathbf{Z}[o$, *each member of* $\sigma_i(o)]$;
      /* Set the Outgoing Relations to $\mathbf{Z}[N, N]$ */
      Set One **To** *entries of* $\mathbf{Z}[o$, *each member of* $\sigma_o(o)]$;
    End
End Procedure

B. Equations

Based upon the SocioMatrix, $\mathbf{Z}[N, N]$, we are able to calculate the closeness centrality measures by applying the equations given in [4]. That is, through the closeness centrality concept and its measurements we can obtain a reasonable level of analysis results, which is enough to answer to the issued question stated in the beginning of the section. The closeness centrality measures can be applied to the individual performer (*individual closeness centrality*) as well as the group of performers (*group closeness centrality*).

[1]The bag theory is same to the set theory except allowing duplicated members.

1) Individual Closeness Centrality Equation: An individual performer's closeness centrality is a function of its geodesic distance to all other performers. The geodesic distance implies the length of the shortest path connecting a dyad in a workflow-supported social network. The conceptual implication of the individual closeness centrality refers to how quickly a performer can interact with others by communicating directly or through very few intermediaries. Conclusively, for a binary nondirected workflow-supported social network with g performers, the index of individual closeness centrality[4] is computed as the inverse of the sum of the geodesic distances between performer i and the $(g-1)$ other performers, where $d(N_i, N_j)$ represents the geodesic distance between two performers, i and j.

- The Index of Individual Closeness Centrality

$$C_C(N_i) = \frac{1}{\left[\displaystyle\sum_{j=1}^{g} d(N_i, N_j)\right]} (i \neq j) \tag{1}$$

- The Standardized Index of Individual Closeness Centrality

$$C_C^S(N_i) = (g-1) \cdot [C_C(N_i)] \tag{2}$$

As you see, the measures computed from the equation (1)[4] can never be 0.0 because division by zero is mathematically undefined. Thus, closeness centrality scores cannot be computed for isolated performer, which is the case of that only a single performer is assigned to enacting the corresponding workflow procedure. Also, we can predict that the lowest closeness centrality values, which is the case of the highest sum of the geodesic distances between a focal performer and others, result either from a performer in a relatively large network or a performer in a small network with relatively long geodesic distances from others.

The equation (2)[4] is for standardizing the index of individual closeness centrality by multiplying by $(g-1)$. Suppose that an individual performer is close to all others, which means that the performer has a direct tie to everyone in the network. Thus the computed index values will be vary according to their network sizes. In order to control the size of the network, it is necessary for the individual index to be standardized so as to allow meaningful comparisons of performers across different networks.

2) Group Closeness Centrality Equation: Group closeness centrality is a dispersion measure indicating the hierarchy of closeness centralities within a network. That is, the group closeness centrality implies the extent to which performers in a given network differ in their closeness centralities. The index of group closeness centrality is computed as followings:

- The Index of Group Closeness Centrality

$$C_C = \frac{\displaystyle\sum_{i=1}^{g} \left[C_C^S(N^*) - C_C^S(N_i)\right]}{\frac{[(g-2)(g-1)]}{(2g-3)}} \tag{3}$$

In the equation (3)[4], $C_C^S(N^*)$ denotes the largest individual closeness centrality observed in a network, and $C_C^S(N_i)$ are the individual closeness centrality of the $(g-1)$ other performers. The maximum value of the indexed group closeness centrality ought to be 1.0 when the corresponding network forms a completely uneven distribution of individual closeness centralities, in the case of that a single performer has the maximum closeness centrality and all others have the minimum closeness centralities. In contrast, the index of group closeness centrality equals to 0.0 when every performer has the same individual closeness centrality score.

Conclusively, the index of group closeness centrality measurement in a workflow-supported social network may take values between 0.0 and 1.0. The closer that group closeness centrality measure is to 1.0, the more uneven or hierarchical is the closeness centrality of performers in a workflow-supported social network; while on the other hand, the closer the measure is to 0.0, then the more the closeness centrality of the workflow-supported social network is evenly dispersed.

C. Algorithm

Based upon the equation (1) of the index of individual closeness centrality, we have devised an algorithm with a recursive function, which is able to compute the closeness centrality measures, $C_C(N_1), \ldots, C_C(N_n)$, for all of the individual performers. The details of the algorithm (named iCcMeasurement and iDistance) are followings:

```
The Individual Closeness Centrality Analysis Algorithm:

Global A Binary Nondirected SocioMatrix, Z[N, N];
Global A Set of Individuals, C;
Global A Set of Traversed Individuals, T;
Global A Set of Distance Values, depth[N];

Procedure Name: iCcMeasurement
Input A Binary Nondirected SocioMatrix, Z[N, N];
Output Individual Closeness Centralities, C_C(o_1), ..., C_C(o_n);
Local A Distance Matrix, Depth[N, N];
Begin Procedure
    For (∀o_i ∈ C)
        For (∀o_j ∈ C, o_i ≠ o_j )
            Depth(o_i, o_j) ← 1;
            Switch (Z[o_i, o_j])
                case 1: /* direct tie between o_i and o_j. */
                        C_c(o_i) ← C_c(o_i) + Depth(o_i, o_j);
                        break;
                case 0: /* no direct tie between o_i and o_j. */
                        Initialize
                            (depth(o_1) ... depth(o_n)) ← N;
                        T ← o_i;
                        Depth(o_i, o_j) ← iDistance(o_i, o_j);
                        C_c(o_i) ← C_c(o_i) + Depth(o_i, o_j);
                        break;
        Rof
    Rof
    Return 1/C_c(o_1), ..., 1/C_c(o_n);
End Procedure

Procedure Name: iDistance
Input The source individual, o_s, and the destination individual, o_d;
Output The shortest distance between o_s and o_d;
Local A Set of Direct-tied Individuals, D;
Begin Procedure
    D ← ∅;
    T ← T ∪ {o_s};
    For ( ∀o_i ∈ C )
        If ( Z[o_s, o_i] = 1 )    D ← D ∪ {o_i};
    Rof;
    For ( ∀o_i ∈ D )
        If (Z[o_i, o_d] = 1 )
            depth(o_i) ← 1;
            depth(o_i) ← depth(o_i) + 1;
            Return depth(o_i);
        Fi;
    Rof;
```

```
For (∀o_i ∈ D ∧ o_i ∉ T )
    depth(o_i) ← iDistance(o_i, o_d) + 1;
    D ← D − {o_i};
Rof;
Return Minimum{depth(o_1), . . . , depth(o_m)};
        /* m is the no. of members in D. */
End Procedure
```

As you see, the time complexity of the algorithm is $\bigcirc(n^2)$. The main algorithm named iCcMeasurement() procedure forms a typical double-loop construct with a recursive function, iDistance(), that can be computed in a constant time, $\bigcirc(1)$, because the number of members of the set, **D**, ought to be much smaller than the number of individual performers. Consequently, by using the algorithm conceived in this paper we are able to measure not only the standardized index of individual closeness centrality but also the group closeness centrality for a workflow-supported social network. Due to the page limitation, we won't describe the details of the remainder algorithms and their application examples to verify those algorithms. However, we strongly believe that the proposed algorithms work correctly without any logical mistakes.

IV. Related Work

Recently, the workflow literature just starts being focused on social and collaborative work analysis on process-oriented organizations. Particularly, our work, workflow-supported social networking knowledge analysis, is directly related with a converged issue of social networks analysis issue and its visualization issue, which we need to dig into more specifically and profoundly as the future works of this research results. A typical one of a few research results on the social network analysis issue might be [6]. In this Ph.D. research, the thesis tried to build a fundamental theory of discovering organizational work-sharing networks, which would be a special type of social networks, from a specific workflow procedure. The organizational work-sharing networks discovered from the workflow procedure have been analyzed by a new statistical analysis approach. And [7] has also proposed a framework for discovering and analyzing workflow-supported social networks. However, the framework's analysis functionality supports only the degree centrality analysis capabilities, which is the simplest analysis techniques out of the four typical centrality analysis techniques, such as degree, closeness, betweenness, and eigen-value centrality measurements. Based upon the results of this paper, we have a plan to extend the framework by developing the closeness centrality analysis functionality in the near future.

V. Conclusion

In this paper, we suggested a possible way of viewing the knowledge and collaborative behaviors among workflow-supported people by converging the social network analysis techniques[4] and the workflow discovering techniques[2][3][6][7]. At this moment, it is important to emphasize that the workflow-supported social network model won't be modeled or designed but automatically discovered from a workflow model. So, we need to devise an automatic analysis functionality and methodology for the discovered workflow-supported social network model. The algorithm proposed in this paper ought to be an impeccable solution for developing the automatic analysis and visualization functionalities for workflow-supported social networking knowledge management systems. Likewise, as a future work, we need to develop the remainder centrality analysis techniques, like betweenness and eigen-value centralities, to be applied to workflow-supported social network models.

Acknowledgement

This research was supported by the Gyeonggi Regional Research Center, Contents Convergence Software Research Center, at Kyonggi University, funded by the Gyeonggi Province, Republic of Korea.

References

[1] Kwanghoon Kim,"Actor-oriented Workflow Model," Proceedings of the 2nd international symposium on Cooperative Database Systems for Advanced Applications, WOLLONGONG, AUSTRALIA, March 27-28, pp. 163-177, 1999

[2] 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

[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] David Knoke, Song Yang, SOCIAL NETWORK ANALYSIS - 2^{nd} Edition, *Series: Quantitative Applications in the Social Sciences*, SAGE Publications, 2008

[5] Kwanghoon Kim, Clarence A. Ellis, "Section II / Chapter VII. An ICN-based Workflow Model and Its Advances," *Handbook of Research on BP Modeling*, pp. 142-172, IGI Global, ISR, pp. 142-172, 2009

[6] Jaekang Won, "A Framework: Organizational Network Discovery on Workflows," *Ph.D. Dissertation*, Department of Computer Science, KYONGGI UNIVERSITY, 2008

[7] Jihye Song, et al., "A Framework: Workflow-based Social Network Discovery and Analysis," Proceedings of the 4rd International Workshop on Workflow Management in Service and Cloud Computing, Hongkong, China, December, pp. 421-426, 2010

[8] Kwanghoon Kim, "A Workflow-based Social Network Discovery and Analysis System," Proceedings of the International Symposium on Data-driven Process Discovery and Analysis, Campione d'Italia, ITALY, June 29-July 1, pp. 163-176, 2011

[9] Haksung Kim, Hyun Ahn, Kwanghoon Kim, "A Workflow Affiliation Network Discovery Algorithm," ICIC Express Letters, Vol.6, No.3, pp. 765-770, 2011

[10] E. Ferneley, R. Helms, "Editorial of the Special Issue on Social Networking," Journal of Information Technology, Vol.25, No.2, pp. 107-108, 2010

[11] 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, 2010

[12] Katherine Faust, "Centrality in Affiliation Networks," Journal of Social Networks, Vol. 19, pp. 157-191, 1997