Egocentric Database Operations for Social and Economic Network Analysis

Adio Akinwale, Adebayo Adekoya, and Adebukola Onashoga
Department of Computer Science, University of Agriculture, Abeokuta, Nigeria

atakinwale@yahoo.com; **adekoyaaf@unaab.edu.ng**; **bookyy2k@yahoo.com**

Abstract

Recent research development on egocentric social network analysis focuses on the causes and consequences of relationships among sets of people rather than on the features, attribute data of individuals, and their basic database operations. This work describes a record type of network data model that would be suitable to perform database operations in social network analysis (SNA). It describes two schemas that would be needed for web space interface and database operations. A force directed algorithm was used to display membership vertices that show link membership number, name, and pictures. A Breadth first search algorithm was then embedded in the algorithm to avoid endless searching in cycled nodes. Measurement of degree of centrality, betweenness, and closeness was incorporated to determine the relationship among the nodes and their linkages. Implementation of database operations into SNA reveals not only social information but also critical economic reports of the individuals. Observation of the work showed that users can impose any relevant query related to the defined schemas and the results would be available to monitor both social and economic activities of the memberships.

Keywords: Egocentric, database operations, record types, schemas, measurement

Introduction

A social network is basically a set of actors and relations that hold these actors together. Actors can be individuals (egos), organizations, departments, or families. Actors exchange information and knowledge among themselves. These kinds of information exchanges are considered a social network relation where individuals who maintain the relation are said to maintain a tie. The simplest form of a social network consists of actors or events and their connections to each other. The use of a social network depends on the availability of relational schemas rather than the attribute data. It further allows the mapping of relationships among people that can be used to identify knowledge and information flow such as who people seek assistance from during economic hardship or share information with. By using social network analysis, users can be aware of how information flow through social ties and how people acquire information.

Egocentric social network analysis focuses on the network of individuals where each person has his or her own

network of relationships that may traverse to many different groups. These relationships contribute to individual's behaviour and attitudes. The strength of egocentric analysis is its ability to capture the diversity of a social environment. The strength may range from weak to strong which depends on the number and types of social activities and resources. In egocentric studies, participants are asked for a list of their network members (alters). Researchers are then interested in constructing a network that is composed of each individual and his or her links to the alters. Each alter links to other alters. It is difficult to survey each of the alters when there are many links among alters and egos. It is then possible for researchers to rely on participants in reporting their relationships to their alters (Coleman, Kartz, & Menze, 1987).

Researchers are always interested about the existence of a tie and possibly its strength and avoid questions about asymmetric ties because these are unlikely to be well understood by respondents. For example, respondents are not likely to know if two of their co-workers know each other unequally. In addition, consider a situation where there are cooperative societies trying to understand the way each member borrows and pays back loans granted.

Literature Review on Egocentric

Graphs and matrices are normally used to present social network and qualifying important structural properties. Graphs are instruments of reasoning about quantitative information. Often the most effective way to describe and summarise a set of numbers, even a large set, is to look at the pictures of those numbers through graph presentation in a graph. Graphic presentation such as set partition cannot only describe data in different ways but can also facilitate the comparison among sets of data, stimulate scientific innovation, and even encourage theoretical insights.

All work done so far on egocentric social network was based on visualization of every member. Information visualization of egocentric social network consists of an appropriate transformation of input data to output graphics (Chalmers & Chitson, 1992). Accordingly, it can be argued that a visualization method is acceptable only if it clearly identifies the relevant information of ego in relation to alter and defines an appropriate mapping to generate their connections. Effective visual attributes for information coding are position, size, shape, colour, motion, etc. Light (2001) believed visualization should be a means of involving the user in a search for results rather than creating visualization to present results to the users.

Nardi et al. (2002) used spatial grouping and colour redundantly code for various community groups within a visualization of a user's e-mail contacts. TouchGraph (http://www.touchgraph.com/navigator) used a force directed layout to present a network visualization of users by allowing personal networks to be expanded or contracted by the users' interactions. Boyd (2004) analyzed users' instant messaging buddy list to present a static network visualization of their contacts. Viégas, Boyd, Nguyen, Potter, and Donath (2004) visualized personal e-mail archives in a zoomable network view and included temporal filtering to visualize contacts at various time periods. They found their work to be particularly well suited as a memory aid for past contexts and contacts. The work was designed to be an on-line dating site with profiles, demographic and interest driven search together with a private messaging system. In this system, people regularly update their profiles, add new photos, change tests dependent on their friends and writing testimonials for their friends in order to encourage reciprocity.

In 2005, Louis Tomsic stated that Stanley Milgram in the 1960s developed small world studies that represented an interesting and less conventional use of egocentric network data. In the studies, respondents in various parts of the United States of America were told the name, occupation, and city of residence of some target people and were asked to mail a packet of papers to a particular person if and only if they knew the target person. They were asked to send the packet to someone who they do know and who they do believe has a chance of knowing the person. Track-

ing the path of the packets provides information on how people know each other and on the average number of links between pairs randomly chosen people in a large society like the United States of America.

Currently, there are many comprehensive tools developed for egocentric analysis. Among them are UCINET (http://www.analytictech.com), Pajek, Agna, NetDraw, NetMiner, MultiNet, and Stocnet. Developers of these tools believe that human eyes are powerful network analytic tools. These tools assist analysts to easily understand the structure of information in egocentric network and statistical models of social networks. In practice, there are many egocentric network developers who have designed and implemented egocentric systems but failed to include database operations.

Problem Statements

Researchers in egocentric social network analysis only do investigation on the existence of a tie between ego and alter and possibly its strength. Apart from adding new members or searching for existing ones in a database, they always avoid the basic database operations such as selection, projection, join or delete a member since it is possible for a respondent to leave the network. Other database operations may include arithmetic operations, union or intersection of member profile for matching pattern and measurement of value of the members in the network. The work done so far on this subject considered a lot of efforts on social network analysis rather than economic network analysis. This work takes a more instructive view on egocentric economic network analysis by examining economic activities of ego in relation to alters and associated interconnections with other alters in another economic environment. It would capture economic information of egos and alters in a single database while social information would be on web space. Through social network analysis, the work would measure the value of economic and social relationship among members. If the value falls within a certain range, the social link would be coloured for effective monitoring. Query operations would be performed on the attribute data of the members.

Measurement of Social Network Analysis

Social Network Analysis (SNA) normally uses data mining techniques to mine a large volume of association data and discover mutual relationship among respondents. Centrality is one of the most studied concepts in SNA and normally used to identify key members who play important roles in a network. According to Ab Raub et al. (2007), centrality measures can be categorized into three: degree, closeness, and betweenness.

Degree Centrality

The out-degree centrality is defined as follows:

$$C_D(n_i) = d(n_{ij}) = \sum_{j=1}^{g} x_{ij}$$

$$where \ x_{ij} = 1 \ if \ there \ is \ link \ between \ nodes \ i \ and \ j,$$

$$and \quad x_{ij} = 0 \ if \ there \ is \ no \ such \ link$$

To measure centralisation in a network, there is a need to observe the difference in the centrality valued of the most central nodes and all the other nodes. Then, to arrive at the centralisation value, we calculate the ratio of the sum of actual differences and the sum of the maximum possible differences. Centralisation is thus defined as:

$$r = \frac{\sum_{i=i}^{g}[\max(D_i) - D_i]}{(g-1)(g-2)}$$

where D_i is the number of people in the network that are directly linked to person i. The number of respondents (actors) is represented by **g** in this equation.

Closeness Centrality

A node that lies at a short distance between many other nodes is considered as closeness. Closeness centrality is calculated as follows:

$$C_c(n_i) = \frac{J_i/(g-1)}{[\sum_{j=1}^{g} d(n_i n_j)]/J_i}$$

where n_i stands for node i

           $d(n_i, n_j)$ defines the length of the shortest path from node i to node j

           g   is the total member of nodes in a graph

           J_i is the total number of actors who are reachable from n_i

Betweenness Centrality

Betweenness centrality measures the extent to which a particular node lies in between the other nodes of the network. It is the number of shortest paths that pass through a given node. The betweeness centrality of node i, denoted as b(i) is obtained as follows:

$$b(i) = \sum_{j,k} \frac{g_{jik}}{g_{jik}}$$

where g_{jik} is the number of shortest paths from node j to node k (j, k $\neq$ i)

Another measurement of SNA is network density. It represents the actual number of ties in a network as a ratio of the total maximum ties that are possible with all the nodes of the network. A fully dense network has a network density value of 1 which indicates that all nodes are connected to each other. A network with a density value near 0 indicates that it is a sparsely-knit network. For an undirected graph as the case of SNA, network density is calculated as follows:

$$D = \frac{2N_i}{N(N-1)}$$

where N_i are the number of the ties and N is the number of the nodes

Society Network Data Model

Society network data model can be 1:1, 1:N, or M:N relationships. Figure 1 illustrates the link relationship between the record types ego and the alter. In general, each ego may relate to many

alters and each alter may be related with many egos. The fact that alters are related in both ego1 and ego2 as shown in Figure 1 makes this an N:M relationship. An N:M relationship can be represented without using N:M links between record types. Instead an intermediate record type and two 1:N links can be used. In this way a functional mapping is maintained. The introduction of the intermediate record type is not always artificial. For instance, it can be used to contain data common to both egos and alters. For database operations, the network database is a set of record types and links defined as follows:

- there is a set of named links connecting the record types in data structure of both ego and alter

- every link is functional in at least one direction (partial functionality is allowed)

  $ego_1 \rightarrow alter_1$ that is ego_1 is functional determined $alter_1$ or $alter_1$ is functional dependent on ego_1.

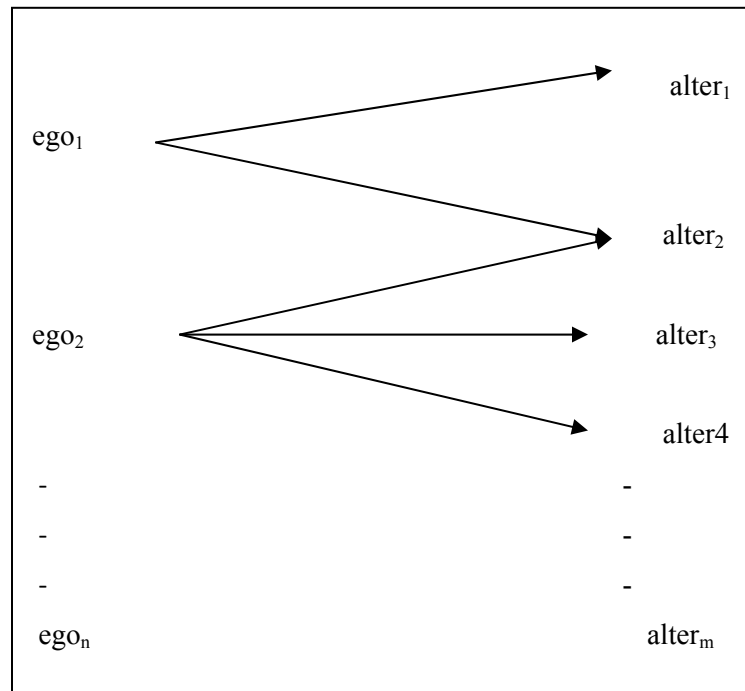- there is no link in form of ego_{ii} that is from a record type to itself (reflexive)

Figure 1: Relationship between record types of ego and alter

In addition, all logical units of data are declared in a schema and subschema. This type of set membership of records is optional and automatic; an optional member can join and leave the network society at any time while an automatic member is automatically assigned to other existing members (Lightstone et. al., 2007). In the system, a relational identification number or text is used as database key. In addition, a record can have an independent existence apart from the links among record types. That is, a record does not have to be connected to another record even though a link is defined between the two record types. Deletion of a record in an egocentric network model does not necessarily trigger the deletion of all records connected to it. Just as a record is explicitly connected to another record an effort is made to disconnect a record from another record. Therefore, it is possible to remove all of the connections and then delete the records. It is also possible to have the removal of certain connections implied by the deletion operations or to trigger the deletion of certain records connected to the deleted record. In this system, it is possible

to combine deletion and removal of connection between records so that certain connections are automatically removed while other connections cause the associated records to be deleted. In this type of network database, there can be several paths from one record to another record. This makes it necessary to have several current records in the database in form of pointers. These pointers keep track of where it is in the database. It is also possible to have cycles or endless loop. This work sets the times of a cycle to one.

Force Directed Algorithm of the Network

The work used force directed algorithm of Koren and Harel (2002) that could display a large number of vertices. Figure 2 illustrates the force directed algorithm of the network. The algorithm was fine tuned to reflect the semantic relationship among the alters, egos and their central degree of measurements. The vertices represent picture of the members, names, and link membership numbers. Breadth First Search algorithm was integrated into search to avoid repetition in a cycle.

```
//The aim is to find Layout, L  of graph G
Begin
  Set_Time();
  Layout(G(V,E));
     k = minSize = 10,  radius = 7;
      While k ≤ |V|  do begin
         Centres := k_Centres(G(V,E), k));
          Search link_identification number using BFS(k);
            If(photo_member_identification_existing()==false and k_existing()== true )
               Add photo identification number into database
               L := add(if(photo_identification_number != existing()))
                If L_existing() then
                   L := (linkid, name, member_Picture)
                   localLayout(L(centres), radius, 4)
             for every v ε V do begin
               L(v) ε L(centre(v))  + rand
                 K:= kRatio
               return L
             End
             Compute Xᵢⱼ for 1 ≤ i≠ j ≤ n; Compute Gᵢⱼₖ for 1 ≤ i ≠ j ≠ k ≤ n;
             Compute  Dij for  1 ≤ i≠ j ≤ n;
   end
end.
```

$$\text{Compute } X_{ij} \text{ for } 1 \le i \ne j \le n; \text{ Compute } G_{ijk} \text{ for } 1 \le i \ne j \ne k \le n;$$
$$\text{Compute } D_{ij} \text{ for } 1 \le i \ne j \le n;$$

Figure 2: Force Directed Algorithm of the Network

```
K_Centres(G(V,E), k)
S:= {v} for some arbitrary v ε V
  For i:= 2 to k do begin
    Find the vertex u farthest away from S
     S := S ⋃ {u}
     Return S
 end
```

```
LocalLayout(L, k, 4)
  For i:=1 to 4 * |v| do begin
   Choose the vertex v with the maximal Δᵥᵏ
   Compute BFS( ∂ᵥᵏ )
    L(v)  := L(v)  + ( ∂ᵥᵏ (x),  ∂ᵥᵏ (y))
  end
```

Figure 2a and 2b: Called Functions of Force Directed algorithm of the Network

Data Collection

The process was applied to cooperative society network analysis in Abeokuta, Nigeria, as a case study of social and economic network analysis. There are sixty five cooperative societies in Abeokuta, Nigeria, with a minimum number of 251 memberships and a maximum number of 500. There is an assumption that needs to be taken in calculating the measurement of degree in cooperative societies in Abeokuta, Nigeria. First, the proportion of alters in cooperative society are members of the population. It is assumed that every member has an equal chance to know each member from a given cooperative society. This means there is probability that every member has a perfect knowledge of other members. More so, every member has the right to join more than one cooperative society. Bernard, Johnson, & Robinson (1989) expressed the equation as follows:

$$\frac{m}{c} = \frac{e}{t}$$

where **m** is the average of alters that respondents know in the subpopulation,

        c   is the average size of the respondent's network

        e   is the total size of the subpopulation and

        t   is the total population

Relational data are collected from cooperative societies along with attribute data. The relational and attribute data are information that is stored in schema and subschema. The schema contains social network analysis data of each member as follows:

SNA_Data$_1$, SNA_Data$_2$, SNA_Data$_3$,…,SNA_Data$_n$ (member identification number, link identification number, member name, age, gender, e-mail, phone number, passport photography)

The information on this schema serves as egocentric online social network input interface and they would appear on web space. The subschema contains cooperative network analysis (CNA) data of each member as follows:

CNA_Data$_1$, CNA_Data$_2$, CNA_Data$_3$,…,CNA_Data$_n$(member identification number,
        link identification number, passport photograph, member name,
        occupation, registration number cooperative society, name
        cooperative society, loans amount, past loan amount, outstanding
        loan amount, date loan given, first sureties name, second sureties
        name, borrower comment)

This information on both schema and subschema are stored in Microsoft SQL Server. For security purpose and to able to track members' information, member identification number, link identification number, and passport photograph are used as candidate keys while passport photograph is used as primary key. This would assist tracking a situation whereby a member can join different cooperative societies with different particulars such as name, occupation, age, photograph, etc. The attribute of borrower comment is created as text and two questions are needed to be provided by the member. They are as follows:

- to whom do member turn for assistance during financial needs

- who are the friends among his or her cooperative members who always stand for him/her as guarantor?

These statements serve as security for checking the economic network analysis of the alters.

Implementation and Results

The force directed algorithm was implemented in Java Programming Language and connected into Microsoft SQL Server, which serves as back end. On execution, the program loads the network environment, which consists of web space and database section. Web space displays the attributes of the membership while the database section keeps values of both schema and subschema. Figure 3 shows the user input interface of social network analysis concept space that contains the details of add, search, edit, and delete. As shown in Figure 3, nodes are labelled with link membership number, name, and member passport photograph. Moving the cursor over a node will pop up the tool tip showing the attributes of the memberships. The user can select as many operations as the user wants from the input interface and conduct database operations such as projection, selection, join, union, sort, add, delete, and search. It is also possible to view the number of alters in the network that are directly linked to an ego or the number of egos reachable from any node.

For example, to add a new member into the network, the add detail panel is required to be filled. Deleting a member from the web space or database requires both search and deletes detail panels. Figure 3 shows a sample of seven members being added using add panel details while Figure 4 illustrates the delete operation by reducing the member to six using search and delete panel details. On clicking the photograph, a box appears which requires the password to be entered by any executive member of the cooperative society. After supplying the password, another query operation box will also appear for entering any database query operation. For example it is possible to check if a member has joined more than two cooperative societies and his/her financial outstanding amount as follows:

Select passport photograph, sum(outstanding loan amount)

From $CNA\_data_1, CNA\_data_2, \ldots, CNA\_data_n$

Where passport photograph $==$ "image"
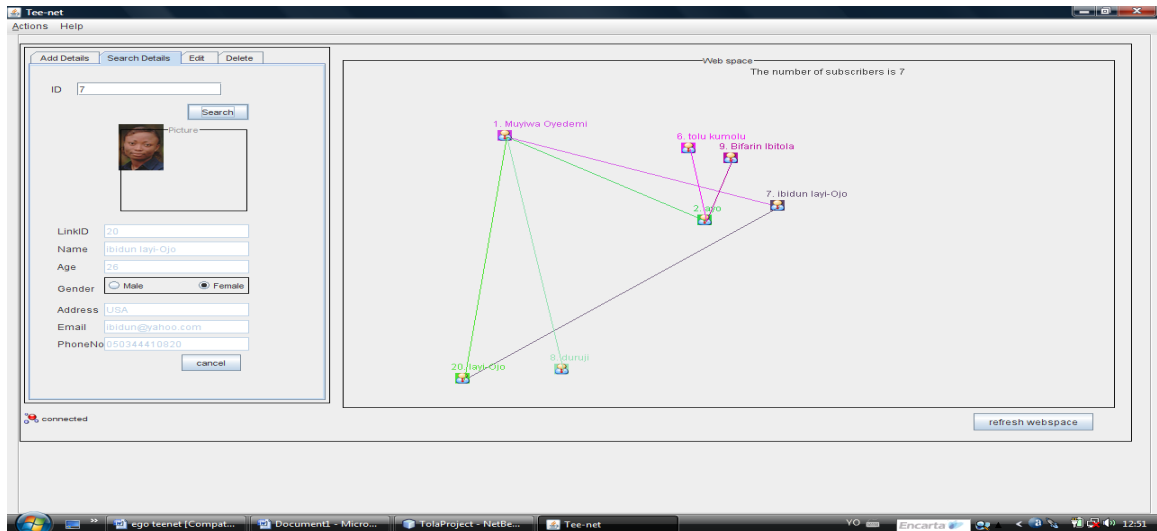
Group by passport photograph

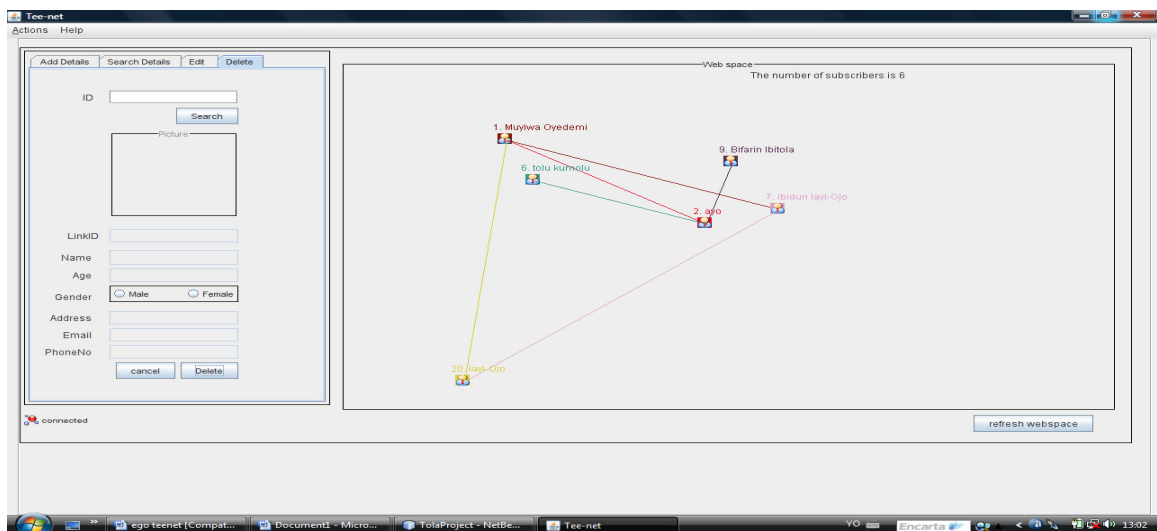**Figure 3: Using add panel detail to add member into web space**

Figure 4: Using search and delete panel details to remove a member

Assuming the result of the above query indicates that a member joined more than one cooperative society and the outstanding loan balance was more than 0.5million naira, a trigger function would be called to show a blinking red edge link between the member ego and alter. This permits the executive member of the cooperative societies to pay a great attention to such member. After execution the query, another browser is invoked to present database operation results for report analysis. The operation panel permits the users to manipulate different database attributes of members.

Conclusion and Future Works

The paper included database operations into social network analysis for performing relevant queries related to the defined schemas and subschema. Effort was made to incorporate force directed algorithm to avoid edge crossings and uniform distribution of nodes and edges. The embedded structural query language trigger constraint served as an objective of discovery of defaulters. The process permits viewing all egos in relation to the alters and monitoring their social and economic activities. For future works, in case the number of membership reaches 200 or higher then, the

input interface may appear crowded and complex. To solve this problem, an alternative two layer hierarchical interface is proposed as future works that first displays society cluster and later enlarge the cluster to see each member detail.

References

Ab Raub, R., Nor Hamzah, A. H., Mohamad, A., Mohamed Bandi, A. F., Hasan, S., Abdullah, S., & Jaafar, M. D. (2007). Network analysis in analyzing potential fraudster. *Proceedings of the International Conference on Electrical Engineering and Informatics*, June 17-19, Indonesia. Available http://repository.gunadarma.ac.id:8000/679/2/eRepository-eprint-679.txt

Bernard, H. R., Johnson, E. C., & Robinson, J. (1989). *Estimating the size of an average personal network and of an event subpopulation.* In M. Kochen (Ed.), *The small world* (pp. 159-175). Norwood, NJ: Ablex.

Boyd, D. (2004). Publicly articulated social networks. *Proceedings of the Conference on Human Factors and. Computing Systems (CHI 2004)* Vienna, Austria, 1279-1282.

Chalmers, M. & Chitson, P. (1992). Beed: Exploration on information visualization. *Proceeding of the 15th Annual International ACM/SIGIR Conference on Research and Development in Information Retrieval*, pp 330-337,

Coleman, J. S., Kartz, E., & Menzel, H. (1987). The diffusion of an innovation among physicians. *Sociometry, 20*(4), 253-270.

Koren, Y., & Harel, D. (2002). A fast multi-scale method for drawing large graph. *Journal of Graph Algorithms and Applications, 6*(3).

Light, J. (2001). Portable document indexes. In U. M. Fayyad, G. G. Grinstein, & A. Wierse (Eds.), *Information visualization in data mining and knowledge discovery* (pp. 99-102). San Francisco, CA: Morgan Kaufmann.

Nardi, B., Whittaker, S., Isaacs, E., Creech, M., Johnson, J., & Hainsworth, J. (2002). Integrating communication and information through ContactMap. *Communications of the ACM, 45*(4), 89-95. http://portal.acm.org/citation.cfm?id=505251&amp%3Bdl=ACM&amp%3Bcoll=portal

Tomsic, L. (2005). *Establishing a community of practice for collaborative learning and social network analysis*.

Viégas, F., Boyd, D., Nguyen, D., Potter, J., & Donath, J. (2004). Digital artefacts for remembering and storytelling: Posthistory and social network fragments. *Proceedings of the 37th Annual Hawaii International Conference on System Sciences, HICSS 04*.

Biographies

Akinwale Adio Taofiki received his Magister in Informatics from Warsaw University, Warsaw, Poland, M.Sc./PhD in Economic Cybernetics and Computer Science from Oskar Lange University, Wroclaw, Poland. His area of research interest is knowledge database system and query database optimization.

Adekoya, Adebayo is a Lecturer in Department of Computer Science, University of Agriculture, Abeokuta, Nigeria. He has B.Sc (Hons) and M.Sc in Computer Science from Ogun State University (Now Olabisi Onabanjo University), Ago-Iwoye, Nigeria in 1993 and University of Agriculture, Abeokuta, Nigeria in 2002 respectively. He is currently at an advanced stage of his Ph.D research work in Computer Science. His area of research interests include Computer Networks & Data Communications, Ontology and Semantic Web, and Artificial Intelligence.

Onashoga Adebukola is a lecturer in the Department of Computer Science, University of Agriculture, Abeokuta, Nigeria. She just completed her research work for the award of doctorate degree in Computer Science from the Department of Computer Science, University of Agriculture, Abeokuta Nigeria. She has published in both international and local journals. Her current research interests include Computer Security, Data mining and Artificial Intelligence.