Simultaneous Analysis of Multiple Big Data Networks: Graph-Mapping in a Data Model

Ahmad Karawash[1, 2], Hamid Mcheick[1], and Mohamed Dbouk[2]
{ahmad.karawash1, hamid_mcheick}@uqac.ca, mdbouk@ul.edu.lb

Abstract. Network analysis is of great interest to companies, largely because of the huge number of social-network users. Analyzing networks is helpful for organizations that profit from how network nodes (e.g. web users) interact and communicate with each other. Currently, network analysis methods and tools support single network analysis. One of the Web 3.0 trends, however, namely personalization, is the merging of several user accounts (social, business, and others) in one place. Therefore, the new web requires simultaneous multiple network analysis. Many attempts have been made to devise an analytical approach that works on multiple big data networks simultaneously. This article proposes a new model to map web multi-network graphs in a data model. The result is a multidimensional database that offers numerous analytical measures of several networks concurrently. The proposed model also supports real-time analysis and online analytical processing (OLAP) operations, including data mining and business intelligence analysis.

Keywords: Multiple big data network analysis, data model, multidimensional database, analysis measures, online network big data analysis.

---

1 Department of Computer Science, University of Quebec at Chicoutimi (UQAC), 555 Boulevard de l'Université Chicoutimi, G7H2B1, Canada.

2 Department of Computer Science, Ecole Doctorale des Sciences et de Technologie (EDST), Université Libanaise, Hadath-Beirut, Lebanon.

1. Introduction

By its very nature, network connection shares big data. The amount of data crossing networks will continue to explode. By 2020, 50 billion devices will be connected to networks and the internet (Cisco IBSG, 2011) and the absolute volume of digital information is predicted to increase to 35 trillion gigabytes, much of it comes from new sources including blog networks, social networks, internet search, and sensor networks. The network can play a valuable role in increasing big data's potential for enterprises. It can assist in collecting data and providing context at high velocity and it can impact the customer's experience.

As the number of online-network communications is increasing sharply, it is difficult to access or analyze relevant information from the web. One possible solution to this problem offered by Web 3.0 is web personalization (Eirinaki & Vazirgiannis, 2003). Personalization aims at alleviating the burden of information overload by tailoring the information presented to individual and immediate user needs (Mobasher et al., 2000). One of the personalization requirements, which can affect a large part of the network data, is the combination of user web accounts to constitute a personal profile for each user.

In response to emerging trends, this article studies how to deal with multiple networks using the data model view. We treat the multiple network idea from the graph model perspective. Indeed, network graphs have been growing rapidly and showing their critical importance in many applications such as the analysis of XML, social networks, the web, biological data, multimedia data and spatial-temporal data.

This article proposes a model which merges multiple network graphs and then maps the obtained graph in the data model, thereby achieving a multidimensional database which enables better network analysis. As a result, an OLAP (online analytical processing) approach (data-mining and business-intelligence analysis) can be applied in numerous networks at the same time. The network's data are collected in such a way that analysis measures are requested by a database query for several networks at the same time. There are many network analysis measures, but this article studies only centrality measures. To explain the proposed idea, this article also consists of a case study simulation of three social network groups. Also, to illustrate the importance of our study, we show two real examples from different domains in which the model is applicable.

The article is organized as follows. Section 2 explains the problem. Section 3 describes the background to the article and the state of the art. Section 4 discusses the proposed model for mapping multi-network graphs in a multidimensional database. Section 5 explains the simulation steps and describes some results. Section 6 highlights the benefits of the proposed model. Section 7 concludes and outlines our future work.

2. Problem Statement

The huge number of random web connections and the unorganized storage of big data in Web 2.0 motivated computer scientists to develop Web 3.0. The new web is based on a wide arrangement of data. One of the problems with Web 2.0 is the random distribution of multi-accounts of users (social, business or other). Web 3.0 proposed the idea of personalization that meant web concepts shifted from working with words to dealing with personal profiles. To achieve a personal profile, all the user's accounts are treated as one block (account aggregation). Although personalization concept can solves many problems, including random accounts and search engine difficulties, it could affect negatively in the analysis phase. Before personalization, analytical methods were easier to apply because the target was one network. In the new web, however, the goal is multi-network analysis (or multidimensional network graph analysis). For example, in the social network case it is easy to apply analysis to one network as a calculation of centrality measures, but how can we analyze several graphs with a different purpose for one person at the same time (e.g. calculating the degree of centrality of a person in both Facebook and Twitter networks at the same time and with one request)?

Currently the available methods and tools deal with one-dimensional graphs. Thus, the challenge to the new web is to analyze the multi-network (multidimensional) graphs simultaneously. What is the degree of online network analysis that can be achieved with Web 3.0?

3. Background and Related Works

“Network” is a heavily overloaded term, and “network analysis” means different things to different people. Specific forms of network analysis are used in the study of diverse structures such as the internet, transportation systems, web graphs, electrical circuits, project plans, and so on (Brandes et al., 2005). Numerous network analysis measures have been developed since the mid-twentieth century: for example, Katz (1953), Hubbell (1965), Hoede's (1978), Taylor’s (1969) and Freeman’s closeness and betweenness (Freeman, 1979), flow betweenness (Freeman et al., 1991), and Bonacich’s eigenvector (1987, 1991), etc.

Although studies on network analysis have been around for decades, and a surfeit of algorithms and systems have been developed for multidimensional analysis in relational databases, none has taken both aspects into account in the multidimensional network scenario.

Ulrik Brandes proposed algorithms to compute centrality indices on large network graphs (2001). Costenbader et al. discussed, in 2003, how to analyze a research network and they used bootstrap sampling procedures research network to determine how sampling affects the stability of several different network centrality measures. In 2008, Chen et al. developed a graph OLAP framework, which presents a multi-dimensional and multi-level view over graphs. In 2010, Tore et al. proposed generalizations that combined centrality measures. Also in 2010, Manuel et al. devised ManyNets to analyze several networks at the same time with visualization. Xi-Nian et al. investigated a broad array of network centrality measures to provide novel insights into connectivity within the whole-brain functional network (2012). Also in 2012, the HMGraph OLAP was developed by Mu et al., that provide more operations on a multi-dimensional heterogeneous information network. In 2013, Daihee et al. devised the NetCube network traffic analysis model using online analytical processing (OLAP) on a multidimensional data cube, which provides a fast and easy way to construct a multi-dimensional analysis of long-term network traffic data. Wararat et al. proposed a framework to materialize this combination of information networks and discussed the main challenges (2013).

4. Multi-Network Graph and Data Model (Proposed Model)

This section highlights the relationship between the graph model and the data model. The new web trend is to use a multi-network model instead of a graph model to deal with the explosive growth of online networks. A graph is a representation of a set of objects wherein some pairs of objects are connected by links. The interconnected objects are represented by mathematical abstractions called vertices, and the links that connect some pairs of vertices are called edges. Typically, a graph is depicted in diagrammatic form as a set of dots for the vertices, joined by lines or curves for the edges (Trudeau & Richard, 1993). The edges may be directed or undirected. A multi-network graph is generally understood to mean a graph in which multiple edges are allowed.

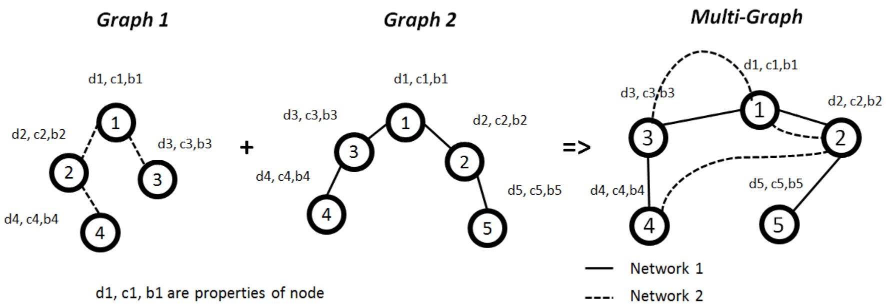

Fig. 1 Merging multiple network graphs in one Multi-Network graph

Figure 1 shows an example of how a multi-graph is obtained from several graphs. Graph1 and Graph2 represent the node connections in two different networks.

A multi-graph is based on *vertices*, *edges*, *belonging network* and *vertex properties*. A multi-graph is an ordered set $M = (V, L, N, P)$ such that:

- V is a set of *vertices*,
- $L = \{\{p, q\}: p, q \, \varepsilon \, V\}$ is a set of links (*edges*) between two *vertices* which are subsets of V,
- $N = \{n1, n2, \dots, nk\}$ is the set of *belonging networks* that node belongs to and
- $P = \{degree\ centrality, closeness, betweenness, \dots, etc\}$ is the set of *properties* of a node.

In order to talk about the relationship between the multi-graph model and the data model, it is necessary first to introduce the entity relationship (***ER***) model. ***ER*** is the most widespread semantic data model. It was first proposed by Chen in 1976 and has become a standard, extensively used in the design phase of commercial applications.

The entity relationship set $\boldsymbol{ER} = (E, R, A)$ is composed of three basic types of sets: *entities*, *relationships*, and *attributes*. An entity set E denotes a set of objects, called *instances*, which have common properties. Element properties are modeled through a set of *attributes* A, whose values belong to one of several predefined domains, such as integer, string, or boolean. Properties that are caused by relations to other entities are modeled through the participation of the *entity* in *relationships*. A *relationship* set R denotes a set of tuples, each of which represents an association among a different combination of instances of the *entities* that participate in the *relationship*.

Let $g: V -> E$ and $h: N -> E$ be two functions mapping the values in set $V \, and \, N$ to set E, in which if $x \, \varepsilon \, V$, then $g\,(x) \, \varepsilon \, E$. Facts $g\,(V)$ and $h\,(N)$**,** derived from the multi-graph M**,** are defined as follows: every *vertex* (node) x in the set of vertices V and every *belonging network* y in the set N is mapped by g and h respectively into *entities* in the set E.

Let $k: L -> R$ be a function such that $k(i) \, \varepsilon \, R$**,** where $i \, \varepsilon \, L$. This means that every *edge* belonging to set L is mapped to *relationship* by k.

Let $w: P -> A$ be a function such that $w\ (c)\ \varepsilon\ A$**,** where $c\ \varepsilon\ P$. This means that every *property* in the multi-graph is mapped in *attribute* in the ***ER*** diagram.

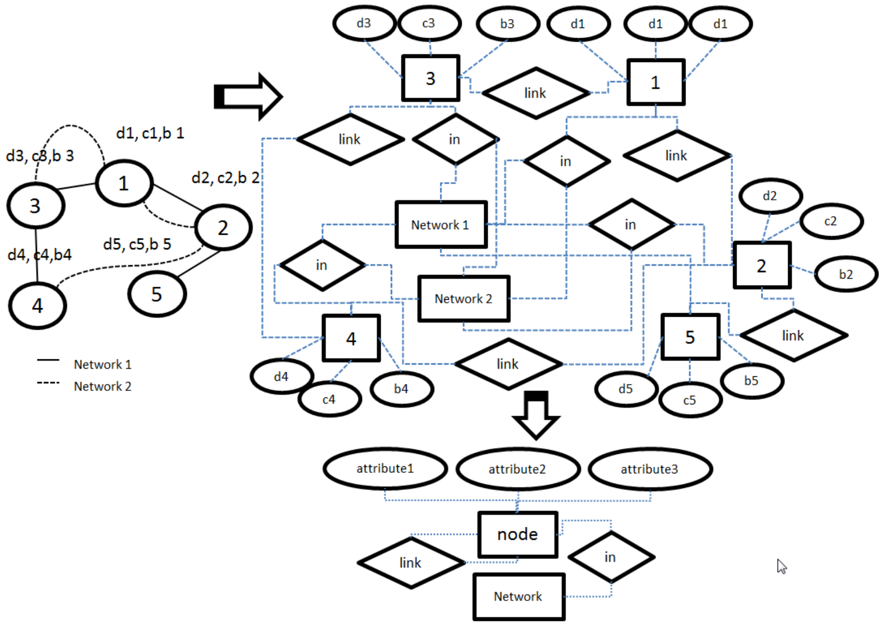

Fig. 2 Mapping a multi-network graph into ER diagram

Figure 2 shows how a multi-network graph is mapped in the ***ER*** diagram. The multi-network graph consists of five nodes each with specific properties. Also, as in the graph in Figure 2, some of the nodes belong to network 1 (lined link) whereas others belong to network 2 (dotted line) , and some may belong to both networks at the same time. As shown in Figure 2, the top ***ER*** diagram forms the result of the translation, in which nodes are translated to entities, properties to attributes and links to relationships. Because the same information is repeated (node name, network type and attributes) the top ***ER*** diagram is optimized into an optimized ***ER*** diagram at the bottom of the figure. The obtained ***ER*** diagram is the same for any multi-network graph (the number of attributes may vary).

5. Multi-network Graph Analysis

This section explains how to benefit from the mapping of the multi-network graph in the ***ER*** diagram in the network analysis.

5.1 Basic Concepts

This section discusses some network analysis concepts. In graph theory and network analysis, there are several types of measures of the centrality of a vertex within a graph that determine the qualified status of a vertex within the graph (e.g. how important a person is within a social network, how important a room is within a building or how well-used a road is within an urban network). Many of the centrality concepts were first developed in social network analysis, such as degree centrality, betweenness, and closeness.

Degree Centrality: The first and conceptually simplest concept, which is defined as the number of links incident upon a node. It is the number of nodes adjacent to a given node (sent = out a degree or received = in degree). The measure is entirely local, saying nothing about how one is positioned in the wider network. Degree centrality is defined by a degree of unit x: $c_D(x) \ = \ degree\ of\ unit\ x$. Relative degree centrality is:

$C_D\ (x) \ = \ c_D\ (x)\ /highest\ degree - 1 \ = \ c_D(x)\ /n - 1$, if n is the number of units in a network, the highest possible degree (network without loops) is n-1.

Closeness Centrality: Measures how many steps away from others one is in the network. Those with high closeness can reach many people in a few steps. Technically it is the sum of network distance to all others. This is not just a local measure, but uses information from the wider network. Sabidussi (1966) suggested a measure of centrality according to the closeness of unit x: $c_c\ (x) = {}^{1}/_{\sum_{y \varepsilon U} d\ (x,y)}$, where $d\ (x;\ y)$ is the length of the shortest path between units x and y, and U is the set of all units. Relative closeness centrality is defined by: $C_c\ (x) \ = \ (n-1)\ *\ c_c(x)$, where n is the number of units in the network.

Betweenness Centrality: Betweenness centrality measures how often a given actor sits “between” others, “between” referring to the shortest geodesic. It detects the actor that has a higher likelihood of being able to control the flow of information in the network. Freeman (1977) defined the centrality measure of unit x according to betweenness in the following way:

$$c_B(\mathrm{x}) = \sum_{\mathrm{y}<z} \frac{\#\text{ of shortest paths between y and z through unit x}}{\#\text{ of shortest paths between y and z}}$$

Suppose that communication in a network always passes through the shortest available paths: the betweenness centrality of unit x is the sum of probabilities across all possible pairs of units that the shortest path between y and z will pass through unit x.

In network analysis, relative betweenness centrality is used; it has two formulas according to the type of network. For undirected graphs of relative betweenness, we have $C_B(x) = c_B(x) \,/\, ((n-1) * (n-2) \,/\, 2)$. For direct graphs of relative betweenness, we have $C_B(x) = c_B(x) \,/\, (n-1) * (n-2)$.

5.2 Analyzing Multi-network Graphs using OLAP

This part maps the obtained ***ER*** diagram in Figure 2 to a multi-dimensional database (cube). In this mapping, we study the three centrality measures (degree centrality, closeness and betweenness) explained in Section 5.1.

Every data analysis is based on a dataset, which is stored in a database. But in our case, we have a multi-dimensional graph. Therefore, we propose to map this type of graph in a multidimensional database. The functions and notations in this part depend on the previous definitions in Section 4 above. Let $L_M(s,x)$ denote a link between s and x where $s, x \,\varepsilon\, V$ *and* $\varphi_s = |\sum_{i\,\varepsilon\, IN} L_M\,(s,x_i) \,/\, n-1\,|$, where n is the number of nodes. Let function $d_M(s,t)$ calculate the shortest path distance between $s,t \,\varepsilon\, V(G_M)$.

Let $S_{st} = |n - 1/d_M(s,t)|$, where n is the number of nodes. let P_{st} denote a set of different shortest paths between s *and* t (such that $s,t \,\varepsilon\, V$) and

$\beta_{st} := |P_{st}|$. For every $v \, \varepsilon \, V$ let $P_{st}(v)$ denote the set of different shortest paths containing v with $s \neq v \neq t$, & $\beta_{st} := |P_{st}(v)|$.

Let $D_{i*j*k}\,(R_{i*k}(D), C_{j*k}(D))$ be a multidimensional database (cube) of order 3, which represents a node in a multi-network graph, as shown in figure 3. $R_{i*k}(D)$ denotes the row i at the k level of the cube and $C_{j*k}(D)$ denotes the column j at the level k of the cube, and $i, j, k \varepsilon IN^{+}$.

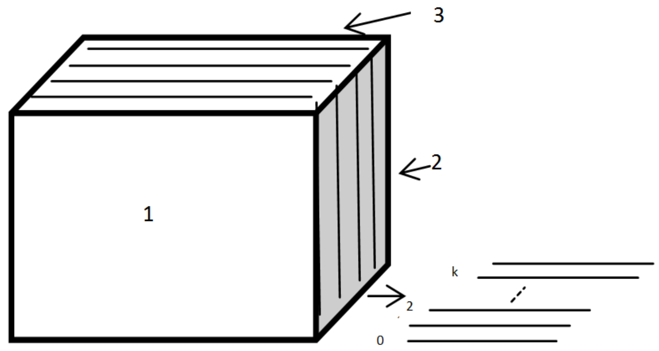

Fig. 3 A structure of a cube with three faces and "k" levels of analysis measures

Let $M_k(\,a_{i,j})$ be a matrix of i, j dimensions, where $a_{i,j}$ is a value of the matrix entity at row i and column j with $i, j, k\varepsilon IN^{*}$. $\mathrm{M_k} := \bigcup_{\substack{0\leq i\leq n \\ k,n\,\in N}} \mathrm{Ri} * \mathrm{k(D)} := \bigcup_{\substack{0\leq j\leq n \\ k,n\,\in N}} \mathrm{Cj} * \mathrm{k(D)}$, which means matrix M_k is formed by the union of cube rows or column at a specific level k. Let R_D denote the set of networks to be studied such that $R_D = \{R_0(D), R_1(D), \ldots, R_N(D)\} = \{Networkname_1, Networkname_2, \ldots, Networkname_N\}$. Let set C_D denote the set of node names such that $C_D = \{C_{0*k}(D), C_{1*k}(D), \ldots, C_{N*k}(D)\} = \{C_0(D), C_1(D), \ldots, C_N(D)\} = \{node\ name_1, node\ name_2, \ldots, node\ name_N\}$ (or $= \{A, B, \ldots, Z\}$ sorted by first letter). Let set C_{D0} denote the set of number of links divided by $n-1$ ($\varphi_{sx} = |\sum_{i\varepsilon IN} L_M\,(s, x_i)\,/n-1\,|$) between a studied node and the other nodes named in C_D, such that $C_{D0} = \{C_{0*0}(D), C_{1*0}(D), \ldots, C_{N*0}(D)\}$ or in other words C_{D0} represents the face of the cube at level zero. Let set C_{D1} denote the set of the distances ($S_{xt^i} = |n-1/d_{GM}\,(x, t^i)\,|$) from a studied node "$x$" to all the other nodes "t^i", such that $C_{D1} = \{\mathrm{C}_{0*1}\,(D), \mathrm{C}_{1*1}\,(D), \ldots, \mathrm{C}_{n*1}(D)\}$. For all the other columns C_{Di}, where $i >= 2$, let set C_{Di} denote the set of different paths between any two nodes passing through a specific node v which is studied by the cube ($\beta_{st} := |P_{st}(v)|$) divided by the sum of different paths between any two nodes ($\beta_{st} := |P_{st}|$), such that $C_{Di} = \{C_{0*i}(D), C_{1*i}(D), \ldots, C_{n*i}(D)\}$.

Table 1 Matrix example represents level 0 of the cube

	Nodename1 (n_1)	Nodename2 (n_2)	Nodename3 (n_3)
Network1 (r_1)	$\boldsymbol{\O s_{n1}^{r_1}}$	$\boldsymbol{\O s_{n2}^{r_1}}$	$\boldsymbol{\O s_{n3}^{r1}}$
Network2 (r_2)	$\boldsymbol{\O s_{n1}^{r_2}}$	$\boldsymbol{\O s_{n2}^{r_2}}$	$\boldsymbol{\O s_{n3}^{r2}}$
Network3 (r_3)	$\boldsymbol{\O s_{n1}^{r3}}$	$\boldsymbol{\O s_{n2}^{r3}}$	$\boldsymbol{\O s_{n3}^{r3}}$

Table 1 explains how the node's cube is structured as a three-dimensional cube of three faces that are divided into "*K*" number of levels (0,1,…, k).

Table 2 Matrix example represents level 1 of the cube

	Nodename1 (n_1)	Nodename2 (n_2)	Nodename3 (n_3)
Network1 (r_1)	$\boldsymbol{S_{sn1}^{r1}}$	$\boldsymbol{S_{sn2}^{r1}}$	$\boldsymbol{S_{sn3}^{r1}}$
Network2 (r_2)	$\boldsymbol{S_{sn1}^{r2}}$	$\boldsymbol{S_{sn2}^{r2}}$	$\boldsymbol{S_{sn3}^{r2}}$
Network3 (r_3)	$\boldsymbol{S_{sn1}^{r3}}$	$\boldsymbol{S_{sn2}^{r3}}$	$\boldsymbol{S_{sn3}^{r3}}$

Table 2 represents the level 0 of the node's cube "s" as a matrix, in which the columns show the other node's name on the graph and the rows show the networks that a node appears in. The values in the matrix entries contain the *degree of centrality* $\boldsymbol{\varphi}$ that node "$\boldsymbol{s}$" has with the other nodes.

Table 3 Matrix example represents level 2 of the cube

	Nodename1 (n_1)	Nodename2 (n_2)	Nodename3 (n_3)
Net_1 (r_1)	$\sum_{\forall xi,yi\, \varepsilon\, V} \frac{\boldsymbol{\beta xiyi}\,[\mathbf{r1}]}{\boldsymbol{\beta xiyi}(\boldsymbol{n1})[\mathbf{r1}]}$	$\sum_{\forall xi,yi\, \varepsilon\, V} \frac{\boldsymbol{\beta xiyi}\,[\mathbf{r1}]}{\boldsymbol{\beta xiyi}\,(\boldsymbol{n2})[\mathbf{r1}]}$	$\sum_{\forall xi,yi\, \varepsilon\, V} \frac{\boldsymbol{\beta xiyi}\,[\mathbf{r1}]}{\boldsymbol{\beta xiyi}\,(\boldsymbol{n3})[\mathbf{r1}]}$
Net_2 (r_2)	$\sum_{\forall \boldsymbol{xi,yi}\, \varepsilon\, \boldsymbol{V}} \frac{\boldsymbol{\beta xiyi}\,[\mathbf{r2}]}{\boldsymbol{\beta xiyi}\,(\boldsymbol{n1})[\mathbf{r2}]}$	$\sum_{\forall \boldsymbol{xi,yi}\, \varepsilon\, \boldsymbol{V}} \frac{\boldsymbol{\beta xiyi}\,[\mathbf{r2}]}{\boldsymbol{\beta xiyi}\,(\boldsymbol{n2})[\mathbf{r2}]}$	$\sum_{\forall \boldsymbol{xi,yi}\, \varepsilon\, \boldsymbol{V}} \frac{\boldsymbol{\beta xiyi}\,[\mathbf{r2}]}{\boldsymbol{\beta xiyi}\,(\boldsymbol{n3})[\mathbf{r2}]}$
Net_3 (r_3)	$\sum_{\forall \boldsymbol{xi,yi}\, \varepsilon\, \boldsymbol{V}} \frac{\boldsymbol{\beta xiyi}\,[\mathbf{r3}]}{\boldsymbol{\beta xiyi}\,(\boldsymbol{n1})[\mathbf{r3}]}$	$\sum_{\forall xi,yi\, \varepsilon\, V} \frac{\beta xiyi\,[\text{r3}]}{\beta xiyi\,(n2)[\text{r3}]}$	$\sum_{\forall \boldsymbol{xi,yi}\, \varepsilon\, \boldsymbol{V}} \frac{\boldsymbol{\beta xiyi}\,[\boldsymbol{r3}]}{\boldsymbol{\beta xiyi}\,(\boldsymbol{n3})[\boldsymbol{r3}]}$

Table 3 represents level 2 of the node's cube "s" as a matrix. The values in the matrix entries, however, contain the result of calculating the number of different paths between any two nodes passing through a node "**s**"($\beta_{st}(s) := |P_{st}(s)|$) divided by the sum of different paths between any two nodes ($\beta_{st} := |P_{st}|$).

A database cube is obtained that represents a multi-network graph at the same time. As a result, it is easy to calculate centrality measures for each node depending on its cube ($D_{\boldsymbol{i*j*k}}$) and by directly applying queries on cube values. In order to calculate the degree centrality and the closeness centrality, the contents of cube levels $k = 0$ and $k = 1$ are invoked, respectively. For betweenness centrality, the cube level $k = 2$ is invoked. If the studied graph is undirected, then we divide the result by $((n-1) * (n-2)/2)$; otherwise the result is divided by $(n-1) * (n-2)$.

6. Simulation and Results

In this section, a simulation of a real multi-network graph example is applied to show the analytical benefits of the proposed mapping of models. In the network analysis domain, social networks retain the first level of importance. There's absolutely no doubt that social networks continue to play an increasingly important part in many people's lives. By 2017, the worldwide social network users will total 2.55 billion (*eMarketer*, 2013). Figure 4 shows the distribution of internet users through the social network services. As shown, more than 50% of internet users are Facebook users.

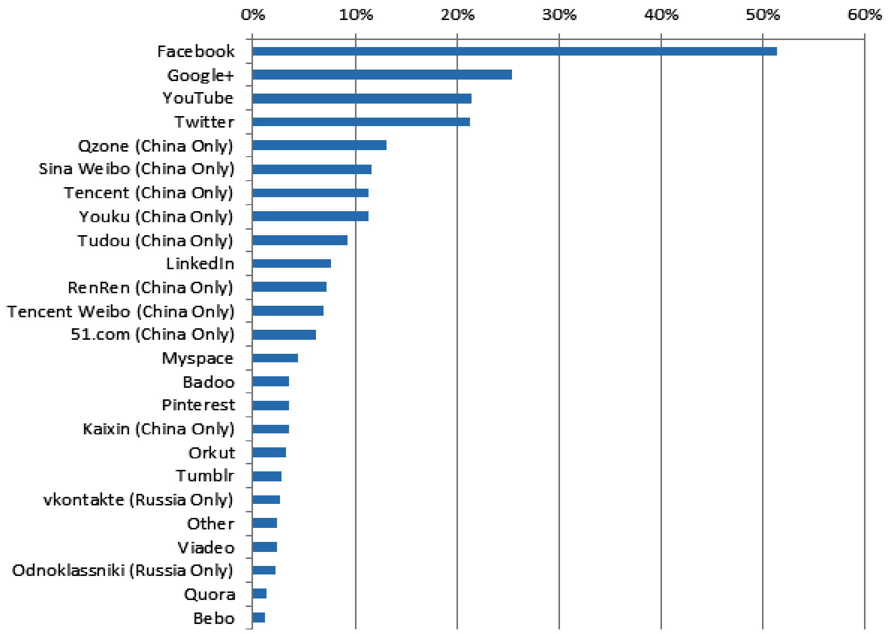

Fig. 4 Percentage (%) of global internet users (source: GlobalWebIndex.com, 2013)

With regard to the importance of social networks in global network data analysis, we have collected and studied three small sets of social data (Facebook, Twitter and Google+). The first is an undirected graph of a Facebook group of 104 members, the second a directed Twitter graph of 76 members and the third an undirected graph of a Google+ group network of 61 members. We tried to gather the same people together in a large part of the data set that shared accounts in different networks. We initially applied one of the traditional network analysis tools, Gephi 0.8.2, to analyze each of the networks alone to get the centrality results.

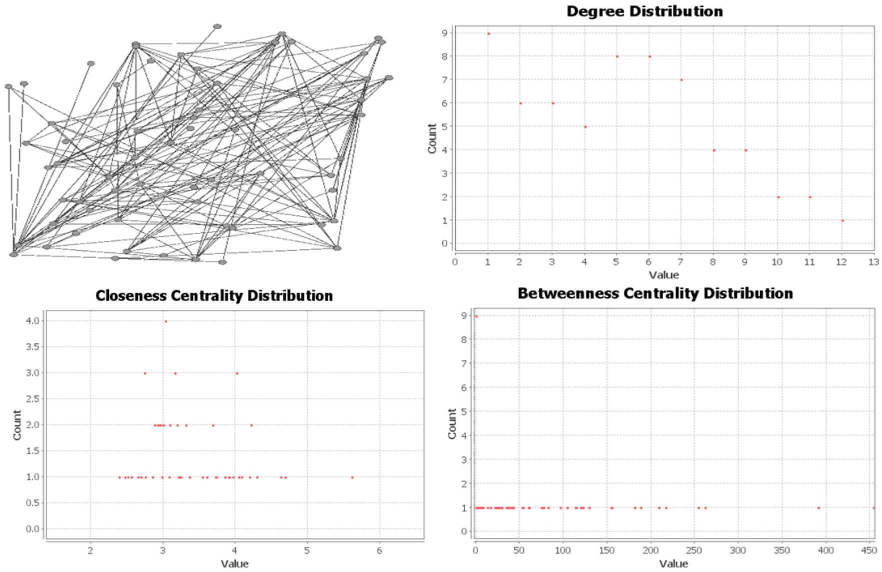

Fig. 5 Centrality measure and graph of the Google+ group

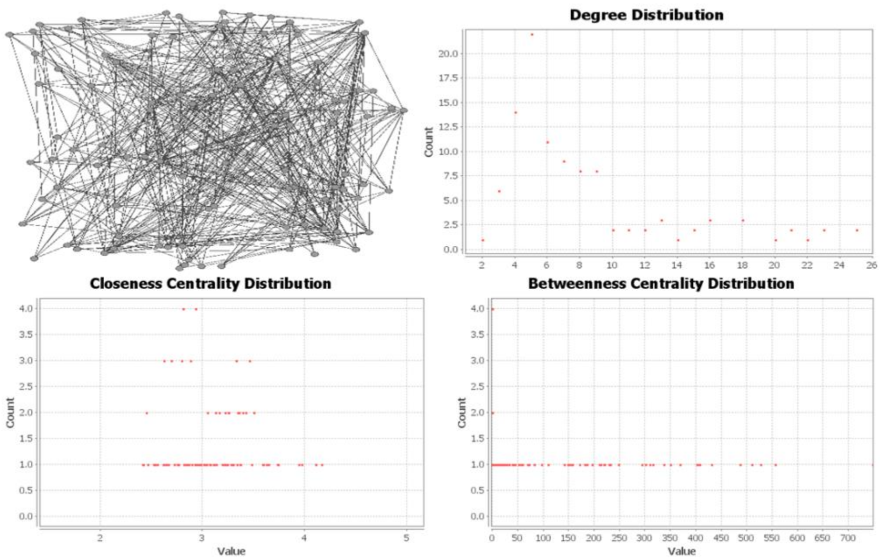

Fig. 6 Centrality measure and graph of Facebook group

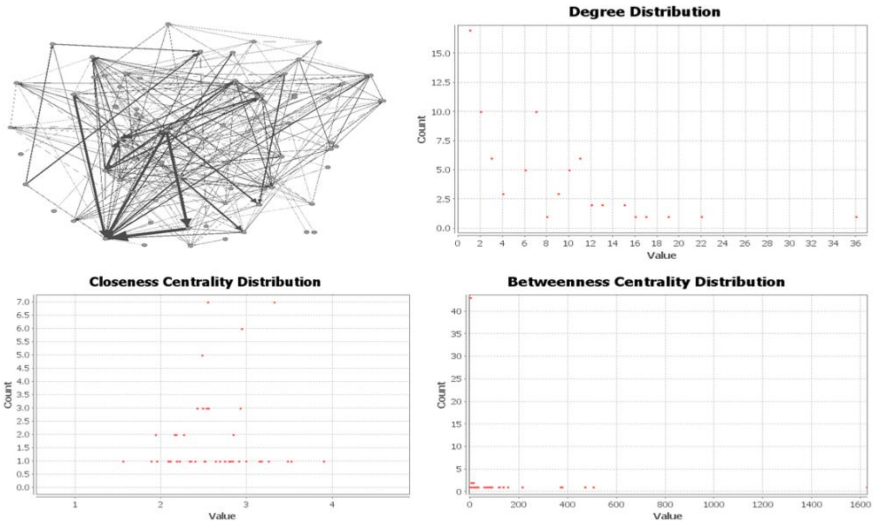

Fig. 7 Centrality measure and graph of a Twitter group

Figures 5, 6, and 7 present the results of analyzing the studied social network groups (Google+, Facebook and Twitter groups respectively). The top left diagram shows the connections between nodes through networks. The top right scatter diagram shows the degree distribution over the group. The bottom left scatter diagram shows the closeness centrality distribution over the group. The bottom right scatter diagram shows the betweenness centrality distribution over the group.

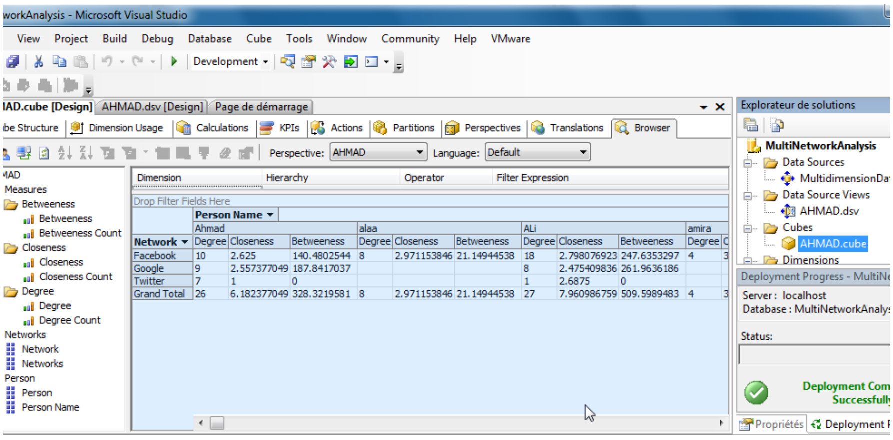

Fig. 8 Building data warehouse using business intelligence and SQL server 2008

Figure 8 shows how to apply the traditional multidimensional database in multi-networks. First, we extract the centrality measures from the given graphs using specific Java codes; this step is similar to the extraction stage in the *OLTP* (online transaction processing). The second step is summarized by building the multi-database schema for the networks using the *SQL server 2008* operations. Then, we customize this database as a data source to build the required cube measures and dimensions using the *SQL business intelligence studio*. Now we can apply simultaneously in multiple networks all the *OLAP* operations on the obtained cube and several data mining algorithms such as: *decision tree*, *clustering*, *association rules* and *neural network*.

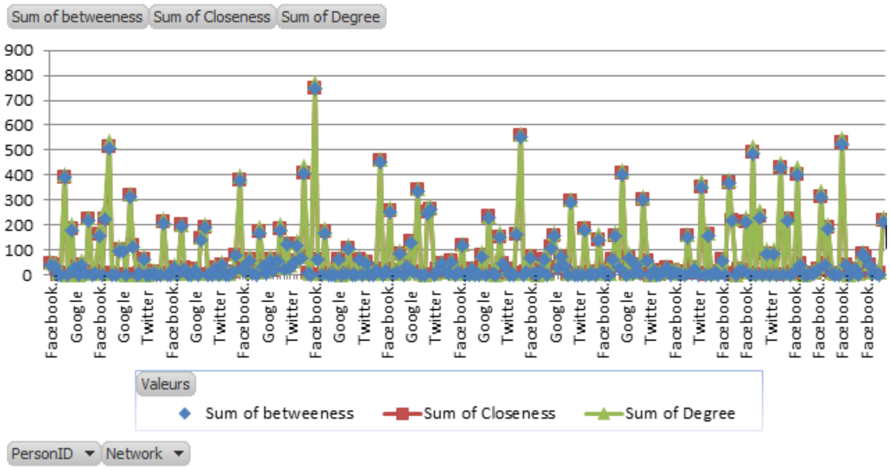

Fig. 9 The distribution of centrality measures over three social network groups

In order to obtain a visual analytical report about the obtained multidimensional database, we have imported the obtained cube in the *SQL business intelligence studio* to *Microsoft Excel 2012*. Figure 9 shows a line chart of the distribution of the centrality measures as a function of the network name.

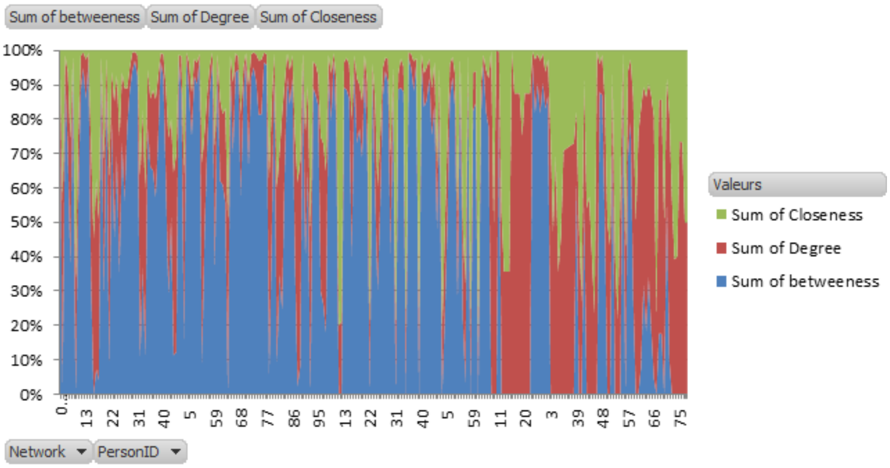

Fig. 10 Area chart reflecting the distribution of analysis measures for the three social network groups concurrently

Figure 10 shows the distribution of centrality measures as a function of both networks and node id. The above results give some idea how analysis can benefit from simultaneous multi-network analysis. This is not everything, however; in fact the analyzer can now apply several queries (see below), which cannot be achieved when each network is analyzed separately:

- Give the name of the person who has the highest degree of centrality in all his/her social network services.
- Give the name of the most important person (according to the centrality measure) in all the social network services.
- Give the name of the person who has a centrality degree greater than 5 in all his/her social network services.
- Give the number of friends that the most important person has through all his/her social network services.

More advanced queries can be applied by means of the proposed model to understand node behaviors in several networks.

7. Benefits and Facilities

The proposed model has attained a multidimensional database for three measures of network analysis and it can be extended to all network analysis measures. Given the large amount of network data, it is hard to analyze the database directly, however. Therefore, a data warehouse has been built. The data warehouse (and its steps extract, transform and load) facilitates reporting and analysis and provides access to structured and unstructured information and operational and transactional data in real time. The obtained data warehouse allows the analyzer to study relations among different networks. It also makes it easier to access simultaneously the node behaviors in multiple networks and answers such plain-language questions as "What happened?" and "Why?" and then predicts what may happen on the basis of strong analytical results. To show the importance of the proposed model in the big network analysis domain, we offer two real examples where the multi-network graph data warehouse is useful. The first example is the analysis of the US election in 2012. The main challenge was to circumscribe web content (web sites, RSS feeds, tweets) coming from the States under scrutiny (Papadopoullos A., 2012). The analysis team collected information from the social networks (multiple networks) over three months before the election. In fact, building a complete multi-network graph data warehouse, that studies all analysis measures, could result a better analyze of election than applying a temporal study. The second example is from Bioinformatics, and concerns Alzheimer’s disease. Alzheimer's is a widespread disease that affects the patient's memory and needs a permanent watch to be kept on the patient's activities. Some technologies have been devised to remind patients about their required activities. These technologies depend on a group of smart sensors that record daily activities. Every sensor recognizes the patient as an object and provides an *activity network* that reflects how s/he deals with the other objects (people, machines,… etc.). Each environment, however, has its own *activity network* and to offer a unique solution suitable for all environments, scientists analyze multiple networks, but not simultaneously. Thus, the solution is to map all the *activity networks* (as a multi-network graph) from sensors to a data warehouse.

8. Conclusion

Web researchers make strenuous efforts to convert the information retrieval web (Web 2.0) into a semantic web (Web 3.0). One of the new concepts of Web 3.0 is personalization that requires aggregation of web user accounts. Indeed, every web user has several web accounts (social, business, study … etc.). If all the networks of these accounts are treated as one, without effect on the individual characteristics of each network, then a multi-network graph of big data web accounts can be achieved for every user. Also, from the network analysis perspective, it is harder to analyze multi-networks concurrently. To solve this problem we have proposed a novel model that maps multi-networks graphs in a multidimensional database. To validate our idea, we applied our model to some network analysis measures of centrality: degree centrality, closeness and betweenness. By means of a simulation, we have discussed the simultaneous analysis of three social networks. In future work, we hope to expand this model to cover ontology and apply it in real complex networks such as Bioinformatics networks.

Acknowledgments. This work has been supported by the University of Quebec at Chicoutimi and the Lebanese University (AZM Association) .

References

Bonacich, P., 1987. *Power and centrality: a family of measures*. American Journal of Sociology 92, 1170–1182.

Bonacich, P., 1991. *Simultaneous group and individual centralities*. Social Networks 13, 155–168.

Brandes, U., 2001. A *faster algorithm for betweenness centrality*. The Journal of Mathematical Sociology, Volume 25, Issue 2, DOI: 10.1080/0022250X.2001.9990249, pages 163-177.

Brandes, U., Erlebach, T., 2005. *Network Analysis*, Lecture Notes in Computer Science, Springer, Theoretical computer science, Vol.3418.

Cisco IBSG. *The Internet of Things: How the Next Evolution of the Internet Is Changing Everything*, 2011.

Chen, P., 1976. The *Entity-Relationship Model-Toward a Unified View of Data*, ACM Transactions on Database Systems, Vol. 1, No. 1. March 1976, Pages 9-36.

Chen, C., Yan, X., Feida, Z., Jiawei H., 2008. *Graph OLAP: Towards Online Analytical Processing on Graphs*, Data Mining, 2008. ICDM '08. Eighth IEEE International Conference, Pisa, DOI:10.1109/ICDM.2008.30, Pages 103 – 112

Daihee, P., Jaehak, Y., and Jun-Sang, P., 2013. *NetCube: a comprehensive network traffic analysis model based on multidimensional OLAP data cube*, International Journal of Network Management, Volume 23, Issue 2, pages 101–118

Eirinaki, M., and Vazirgiannis, M., 2003. *Web Mining for Web Personalization*, ACM Transactions on Internet Technology, 3(1) (2003) 1-27

Costenbader, E., et Valente, T. W., 2004. *The stability of centrality measures when networks are sampled*, Elsevier, Social Networks, Volume 25, Issue 4, October 2003, Pages 283–307

eMarketer report, 2013. *Worldwide Social Network Users: 2013 Forecast and Comparative Estimates.Freeman*, L.C., Borgatti, S.P., White, D.R., 1991. Centrality in valued graphs: a measure of Betweenness based on network flow. Social Networks 13, 141–154.

Freeman, L.C., 1979. *Centrality in networks: Conceptual clarification*. Social Networks 1, 215–239.

Hoede, C., 1978. A *new status score for actors in a social network*. Department of Mathematics, Twente University, unpublished manuscript.

Hubbell, C.H., 1965. An *input–output approach to clique identification*. Sociometry 28, 377–399.

Katz, L., 1953. *A new index derived from sociometric data analysis*. Psychometrika 18, 39–43.

Manuel, F., Catherine, P., Ben, S., Jen, G., 2010. *ManyNets: An Interface for Multiple Network Analysis and Visualization*, ACM CHI.

Mobasher, B., Cooley, R., Srivastava, J., 2000. *Automatic personalization based on web usage mining. Commun*. ACM 43 (8) (2000) 142–151.

Mu, Y., Bin, W., and Zengfeng, Z., 2012. *HMGraph OLAP: a novel framework for multi-dimensional heterogeneous network analysis*, DOLAP'12 Proceedings of the Fifteenth International Workshop on Data Warehousing and OLAP, Pages 137-144.

Taylor, M., 1969. *Influence structures*. Sociometry 32, 490–502.

Tore, O., Filip, A., John, S., 2010. *Node centrality in weighted networks: Generalizing degree and shortest paths*, Elsevier, Social Networks, Volume 32, Issue 3, July 2010, Pages 245–251

Wararat, J., Cécile, F., and Sabine, L., 2013. *OLAP on Information Networks: a new Framework for Dealing with Bibliographic Data*, 1st International Workshop on Social Business Intelligence (SoBI 2013), Genoa, Italy.

Xi-Nian, Z., Ross, E., Maarten, M., Davide, I., F.Xavier, C., Olaf, S., Michael, P., M., 2012. *Network Centrality in the Human Functional Connectome*, Oxford Journals, Life Sciences & Medicine, Cerebral Cortex, Volume 22, Issue 8, Pp. 1862-1875.

Freeman, L.C., 1979. *Centrality in networks: I. Conceptual clarification. Social Networks* 1, 215–239.

Freeman, L.C., Borgatti, S.P., White, D.R., 1991. *Centrality in valued graphs: a measure of betweenness based on network flow*. Social Networks 13, 141–154.

Sabidussi, G., 1966. "*The centrality index of a graph*". Psychomatrika 31 :S81-603.

Papadopoullos, A., 2013. *CASE STUDY Social Network Analysis of the 2012 US Elections*, Chief Technology Officer, Semeon.

Trudeau, Richard J., 1993. *Introduction to Graph Theory*. New York: Dover Pub.. pp. 19. ISBN 978-0-486-67870-2. http://store.doverpublications.com/0486678709.html