# Building Implicit Corporate Social Networks: the Case of a Multinational Company

Luka Humski, Darko Striga, Vedran Podobnik, Boris Vrdoljak, Marko Banek, Zoran Skocir and Ignac Lovrek
University of Zagreb
Faculty of Electrical Engineering and Computing, Zagreb, Croatia
Email: luka.humski@fer.hr, darko.striga@fer.hr, vedran.podobnik@fer.hr, boris.vrdoljak@fer.hr, marko.banek@fer.hr, zoran.skocir@fer.hr, ignac.lovrek@fer.hr

***Abstract*—Social networks based on ICT (*Information and Communication Technology*) are nowadays one of the most popular services based on the Internet infrastructure. They are global phenomenon that greatly affects the modern way of life. Contrary to the widespread opinion, which assumes that social networks are interesting only for private users, these networks can produce added value in companies as well. The corporate social network is a system based on the web technologies that enables agile collaboration and information exchange within company. According to the method of making connections, social networks can be divided into two groups: implicit and explicit networks. While in explicit social networks a person herself defines another person to connect with, implicit networking is determined by a person's interests and by the level of communication and collaboration with other persons. This paper addresses the question of developing implicit corporate social networks. Based on the analysed communication performed through several communication channels (i.e. e-mail, file transfer, telephone calls and instant messaging) between the employees of a multinational company, we propose an algorithm for building the social network graph. Our algorithm calculates the level of connection between employees based on the level of communication between them. We verified the proposed algorithm on our prototype test application, the FER CSN Analysis.**

***Keywords*—corporate social network, implicit social networking, communication channels, E2E, data warehouse, data mart**

## I. Introduction

Social networks based on the information and communication technology [1] have appeared approximately simultaneous with development of the Web – in 1990s. Despite being only a drop in the bucket in the world of the web sites of different themes and purposes just a decade ago [2], social web sites are nowadays not only the most popular service based on Internet infrastructure, but also a global phenomenon that greatly affects the modern way of life [3][4], and is therefore an important element of today's social and business environment. After the immense success of general social networks services like Facebook[1] and Twitter[2], social networks started to receive an increasing attention within companies and academia as well. The *Corporate Social Network* (CSN) (also called Enterprise Social Network, ESN) is a system based on the web technologies that enables agile collaboration and information exchange [5][6]. The CSN consists of: i) *external CSN* which deals with process that include interaction between companies and external stakeholders such as customers or business partners; and ii) *internal CSN* which deals with intra-organizational processes that include employee-to-employee (E2E) and business-to-employee (B2E) interactions. This paper focuses on the internal CSN, describing how it improves communication transparency and trust of employees, enhances the exchange of corporation and common news, and introduces new collaboration opportunities.

Explicit social relationships, where each user herself defines other users (friends or followers) she is connected with, are realized by social network services on the Internet. Unlike explicit networking where a person herself defines another person to connect with, implicit networking is determined by a person's interests and by the level of communication and collaboration with other persons [7]. Prior to its creation by analysing the employees' interests, as well as communication and collaboration between employees, the *implicit corporate social network* (iCSN) is "hidden". In this paper we present a solution to creating the iCSN by analysing the usage of corporate communication services.

This paper contributes to the emerging literature on CSNs from two perspectives – i) *methodological* (i.e. *what* information and *which* steps are needed to "discover" the iCSN); and ii) *technical* (i.e. *how* the process of iCSN building can be implemented). While our first contribution should be placed in the academic domain, the second verifies its applicability to the real-world business domain.

The iCSN construction process for a multinational company is based on the analysis of communication between employees (i.e. E2E). The analysed communication channels are: *e-mail*, *file transfer*, *phone calls* and *instant messaging*. Records about communication were stored in different log files, which were studied in detail and data of interest for the analysis was selected. For each communication channel a special data mart was created. Data marts were integrated into a single data warehouse (a data warehouse is a set of technologies and tools that enable flexible analysis of data coming from different sources [8]; a data mart is a subset of the data warehouse that is usually oriented to a specific

---

[1] http://www.facebook.com

[2] http://www.twitter.com

*Humski, Luka; Striga, Darko; Podobnik, Vedran; Vrdoljak, Boris; Banek, Marko; Skocir, Zoran; Lovrek, Ignac.*
*Building Implicit Corporate Social Networks: the Case of a Multinational Company.*
*Proceedings of the 12th International Conference on Telecommunications / Pripuzic, Kresimir; Banek, Marko, editor(s).*
*Zagreb, Croatia: IEEE, 2013. 31-38.*

domain subset e.g. company department). Based on data stored in the data warehouse, the iCSN (i.e. a directed weighted graph of the employees' social connections) was created. The weighted graph consists of the nodes (employees in this case) and edges (links between them). A real number, called the edge weight, is assigned to each edge between two employees. A greater edge weight means a higher degree of connection between two employees. In order to determine the weight of an edge connecting two employees, it is necessary to analyse both *communication intensity* (i.e. how often these employees communicate with each other) and *communication channel* (i.e. how they communicate). This paper introduces a formula that calculates the edge weights for the particular communication channel based on communication frequency connected with that communication channel. The total edge weight for the particular edge that connects two employees is obtained by combining weights for all communication channels. The level by which a particular channel of communication affects the total edge weight can be adjusted manually.

The paper is organized as follows. Section II makes an overview of the state-of-the-art in the area of social networks. The system architecture is presented in Section III. The data warehouse that serves as the source of data for creating our iCSN is described in Section IV. In Section V the construction of the iCSN is explained and our prototype test application for

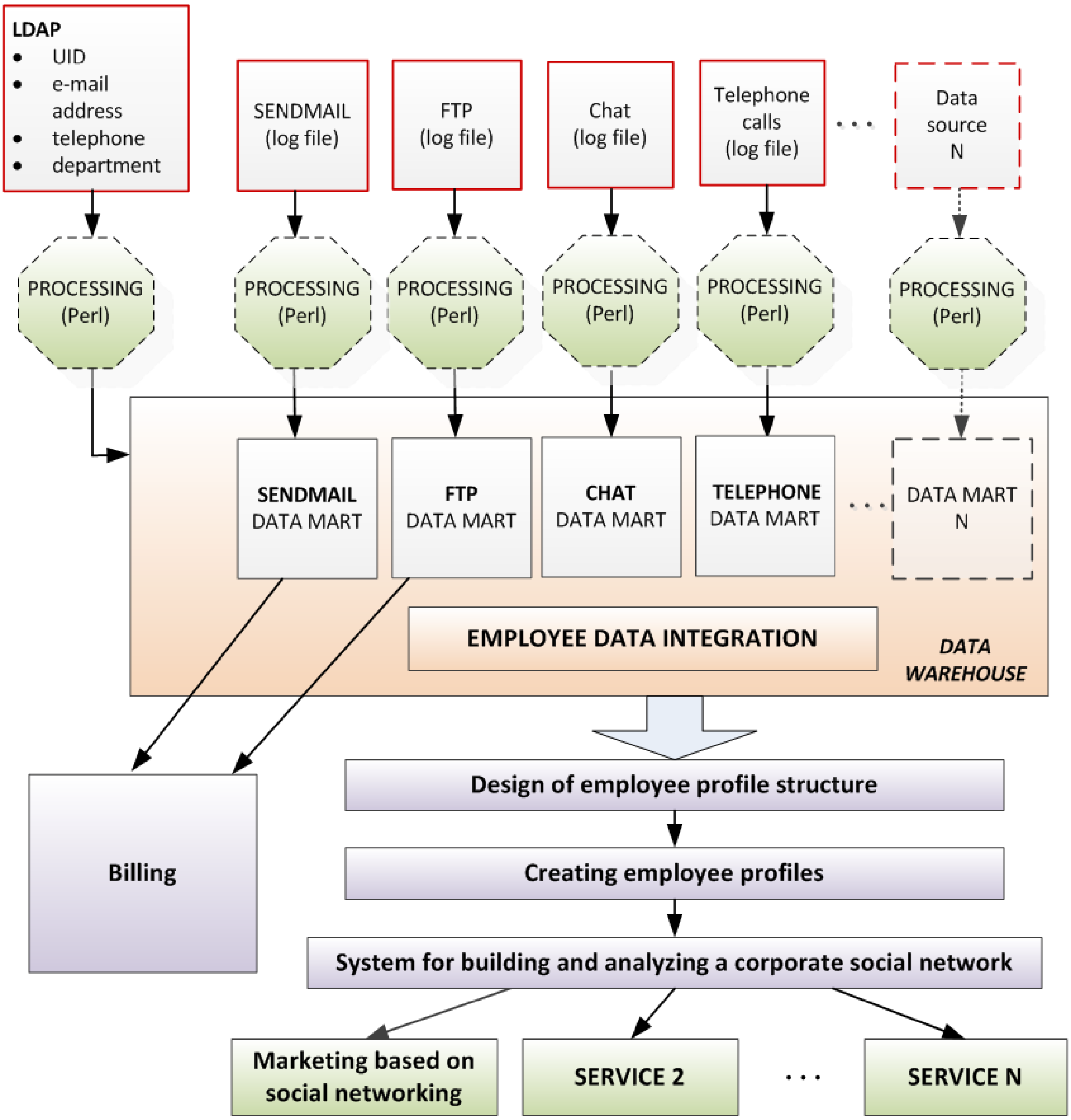


Fig. 1. The architecture of the system for building implicit corporate social network.

network analysis is presented. The drawn conclusions and guidelines for future work are given in Section VI.

## II. Related Work

The research of social networks, i.e. the research of relationships between people, has its beginnings in the second half of the 20th century. Social network analysis is an area which was traditionally practiced by sociologists and has been primarily focused on human and social aspects. Online social networks (i.e. social web sites), which have become more and more popular in recent years, only offer analysis of the human behaviour on specific social web sites. Therefore, the analysis of connectivity between employees should be extended to other forms of communication – not only on the analysis of the online social networks [9].

The term social network can be interpreted in two ways. On one hand, online social networks are used by employees to connect, interact, exchange data, etc. On the other hand, the term social network is a graph obtained by analysing the interactions between employees. Usual modes of communication which are analysed in different researches are e-mail, call logs, text messaging, document repositories and web 2.0 tools in organizations [9]. Enterprises want to get knowledge of interconnectedness between employees as accurate as possible, so that the management can modify and adjust the organizational structure (if an analysis shows it necessary). In the analysis of the social networks, particular emphasis is put on visualization. Research shows that employees receive 50-75% of the information they use for doing their job from their colleagues [10]. Good visualization provides insight into communications among colleges, how successful the communication between departments is, etc. In large enterprises, unconnected groups of people are common, i.e. social networks exist with structural holes [11]. Bridging the structural holes, i.e. connecting the unconnected groups, has a great potential to achieve better communication and knowledge sharing within the organization [11][12]. This research also shows that employees, who are the bridge between two groups, are more productive and efficient in their work because they have more information than other employees [11].

Different research projects have used different program tools for social network analyses and social networks visualisations are in: SmallBlue [9][10], Pajek [13], Gephi [14], SAS [15], etc.

In addition to standard tools and methods, data mining approaches have also been used for analysing social networks. Heterogeneous social networks, where more than one type of relation exists between two nodes, were analysed in [16] and [17].

## III. The System for Building Implicit Corporate Social Network

The process of building an iCSN can be divided into two parts: i) building a data warehouse; and ii) constructing a corporate social network based on the data from the warehouse.

The first step in developing the system is analysing the channels of communication between employees. We have detected that communication (except for live communication) takes place by e-mail, file transfer, phone calls and instant messaging. The designed system architecture enables the inclusion of data about other channels of communication if useful in future. A data mart was developed separately for each channel of communication. The data marts are loaded with data from the server log files for a particular mode of communication. The data marts share dimensions, thus being integrated into a single data warehouse. A corporate directory (LDAP) is of primary importance for data integration. More precisely, the same employees are identified by different key attributes in different log files (e.g. phone number, internal phone number, user name in instant messaging, e-mail addresses, etc.). All these different identifiers of the same user are interconnected in the LDAP, together with non-key attributes such as surname, first name, sector name, office location, etc., and a surrogate key attribute user ID (UID). The information stored by LDAP is unified within a single dimension *dEmployee*, to which any communication record points unambiguously, regardless of the particular mode of communication. The LDAP directory also stores information about the specific company (which is part of the bigger multinational company) where the employee works, which enables analysing communication both within a single company and between different companies.

The multinational company used in proof-of-concept composed of many separate companies that use common ICT (*Information and Communication Technology*) infrastructure. The fairest model of participation in the cost of the ICT infrastructure maintenance is participation proportional to the degree of system usage. However, the question arising is how to obtain an information about the degree of usage of specific information resource by a particular company. The data in the data warehouse gives answer to this question. The primary purpose of creating the data warehouse is collecting the data for a fair billing process for ICT infrastructure usage. However, creating the iCSN was specified as the secondary purpose and certain features of the warehouse design were aimed specifically at supporting it.

The data stored in the data warehouse can be used to build the iCSN. The first step in building the iCSN is designing the employee profile structure and creating the profiles. A directed weighted graph of social connections (social network graph) between employees is produced based on the data about the frequency (the number of exchanged messages or number of completed calls) and intensity (e-mail message size or duration of a phone call) of communication between them. The social network graph can be analysed with the aim of identifying new and interesting characteristics. For example, if a certain number of Java programmers work in an organization, the level of communication between them, i.e. the level of their social connection, can be analysed. If this level is considered insufficient, the company management may take some actions for improving their connection, with the aim of better exchange of knowledge and experience. Successful exchange of knowledge and experience will finally result in increasing the operating efficiency (the ability to solve the same problem in a shorter period of time). Based on the analysis of CSNs, different services can be offered. An example of one such service is marketing based on social networking. Fig. 1 shows the described system architecture.

## IV. DATA WAREHOUSE AND DATA MARTS

The data warehouse is designed by integrating data marts built for each channel of communication. Data marts are implemented for the following communication modes: *e-mail*, *file transfer*, *phone* and *instant messaging*. Data marts for each communication channel will be described below (except for the file transfer service, which was started concurrently with building the iCSN, and thus is contribution was held irrelevant due to an extremely small amount of data). The warehousing process does not violate the privacy of the corporate users: the e-mail and telephone call logs do not contain the content of the communication but only the metadata. In case of instant messaging, the metadata are extracted from the content at the instant messaging application server, before being submitted to the warehousing process.

### A. E-mail

The proof-of-concept multinational company consists of many companies, some of them being part of the multinational company for more than two decades and therefore connected into an integrated communication backbone, some of them acquired recently and therefore retaining their independent information systems, including those for e-mail exchange. About 90% of the e-mail message exchange is performed via an instance of the open-source *Sendmail server*, while the remaining portion of the traffic is operated by an instance of *Microsoft Exchange Server* ®. Each message arriving at the *Sendmail server* is sequentially processed by several independent threads, which work in parallel with each other, producing a log file with information about one particular message being scattered among information about many other messages. The major challenge in developing a Perl script for processing the log file (which grows for additional 500 MB each week) was the fact that a message is not uniquely identified within the log file, since its internal identifier may change several times during the processing. Likewise, each message received by the *Microsoft Exchange server* results in several lines within its log file (each line contains 26 attributes), but its processing is much easier due to a unique message identifier. The data transformed by the Perl script are inserted into the data mart by a series of SQL (*Structured Query Language*) queries (the combination of a Perl script and SQL queries is used for loading other data marts as well).

Data collected at both servers is stored into a single, integrated data mart consisting of two facts (Fig. 2): *fSend* and *fReceive*, each tuple of *fSend* corresponding to a single message and one or more tuples of *fReceive*, depending on the number of recipients (each tuple of *fReceive* represents one message being delivered to one particular recipient). The facts are mutually linked by the pseudo-dimension *dMessageId* (a single attribute in each of the fact tables). Both facts contain the size of the message, *mMessageSize* as measure (the size of an incoming message can be changed at the server during the processing).

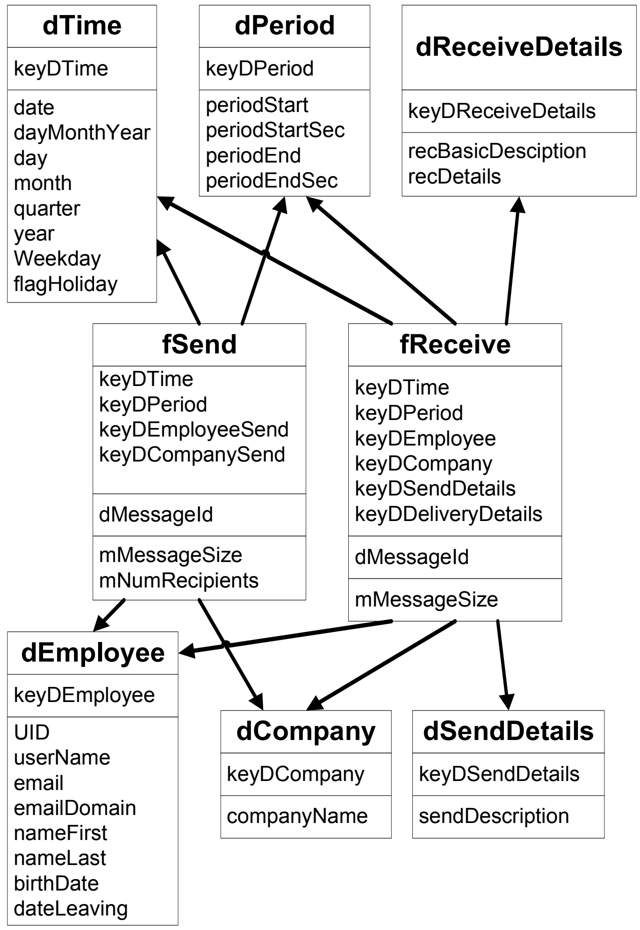


Fig. 2. The star schema of the e-mail data mart.

Each of the facts is described by four other shared dimensions: *dTime* (a day is the smallest grain level), *dPeriodDay*, *dEmployee* and *dCompany*. The *fReceive* fact is additionally associated with two additional measures: *dSendDetails* and *dReceiveDetails*. The warehouse is not designed exclusively for the purpose of billing the ICT services, but also to capture a crucial feature of a CSN: dynamicity. The level of connection between the employees changes continuously. Some employees leave the multinational company (which makes their communication with other employees cease), while others transfer from one particular company of the multinational to another (which causes them to start new connections and reduce or even end the previous ones). This led to specific decisions in data warehouse design: we implemented *dCompany* as a separate dimension instead of making its attributes an upper level of the employee dimension *dEmployee* (which would have become a slowly-changing dimension [18] in that case). During the warehouse loading process each fact record is checked against the LDAP directory (Section III) and associated with the corresponding UID. The attribute *dateLeaving* of the dimension *dEmployee* (which is set to the null-value for current employees) serves to determine whether a total lack of communication is to be assigned to a permanent leave on behalf of an employee.

### B. Phone

The telephone communication within the multinational company is performed using *Cisco*[3] *technology for IP telephony*. Information about the answered calls is recorded in log files. Seven attributes in the log files serve as the source for the data warehouse: *calling party number*, *called party number*, *calling party login user ID*, *called party login user ID*, *connect time*, *disconnect time* and *duration*.

The data mart for the phone communication is composed of fact table *fCall*, which stores call duration as a measure, and the four previously mentioned shared dimensional tables actually serving as eight dimensions: *dTime* (a double reference, separately recording the start time and the end time of a call on a day basis), *dPeriodDay* (again a double reference, recording the exact start and end time of a call), *dEmployee* (a double reference, recording the data about the calling and the called party) and *dCompany* (a double reference, recording data about companies where the calling and the called employee work).

### C. Instant messaging

Within the multinational company instant messaging is implemented by using *XMPP* (*Extensible Messaging and Presence Protocol*, an open technology for real time communication using XML as a basic format for the information exchange). Information about the exchanged messages is recorded in log files. Six attributes are used for the analysis of communication between employees through instant messaging system: *message ID*, *message type* (*chat* or *groupchat*), *message sender*, *message receiver*, *sending time* and *length* (number of signs that message contains). Instant messaging by type can be bilateral communication (only two persons communicate simultaneously) – *chat* or conferencing (more than two persons communicate simultaneously) – *groupchat*.

The data mart for the instant messaging communication contains a fact table *fMessage*, which stores length of the sent message (number of characters), and the four shared dimensional tables serving as six dimensions: *dTime*, *dPeriodDay*, *dEmployee* (a double reference, denoting the sender and receiver of the message) and *dCompany* (a double reference, denoting the companies where sender and receiver work). A pseudo-dimensional flag attribute of the fact table records whether the message is exchanged through *chat* or *groupchat*. Messages in the *groupchat* are stored as bilateral messages between all *groupchat* participants. For example, if *users A*, *B* and *C* communicate through the *groupchat*, a message sent by *user A* will be stored as two messages: a message that *user A* sent to *user B* and another message that *user A* sent to *user C*. For both messages the flag is set to indicate that messages are exchanged through the *groupchat*.

## V. CORPORATE SOCIAL NETWORK

Nowadays, social networks based on information and communication technology are one of the most popular services based on the Internet infrastructure, which makes them a significant component of the contemporary business and social environment. Social networks represent free web services which emphasize the role of common interest [19]. This common interest interconnects the users of a social network. The process of building the corporate social network prototype and an example of its analysis is shown below.

### A. A Model of an Implicit Corporate Social Network "The Multinational Company"

The process of building the "The Multinational Company" iCSN (Fig. 3) has two main inputs. The first input is the source data obtained from LDAP files (Section III) and the three data marts: e-mail, instant messaging and phone calls (Section IV). The second input are the initial weights assigned to each of the three communication means: $W_{email}$, $W_{chat}$ and $W_{phone}$.

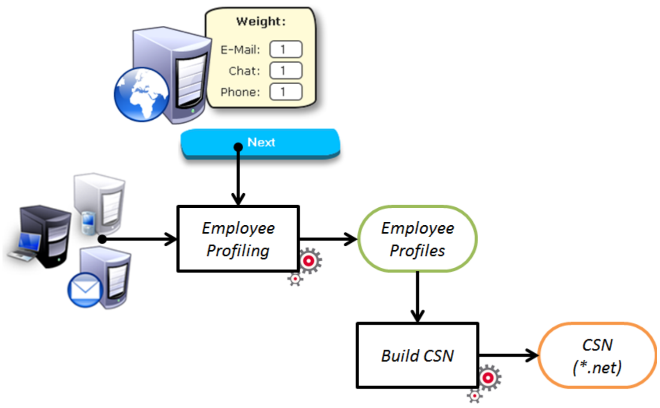


Fig. 3. The process of building the "The Multinational Company" implicit corporate social network.

Using the *JavaServer Pages* technology [20], we implemented a web-based application, which analyses the two inputs, generates employee profiles and finally builds the iCSN.

The employee profiles are created by retrieving identifiers from an LDAP file: unique employee identifiers (UIDs), e-mail addresses, chat user names, phone numbers. In the particular proof-of-concept test case we generated 125 user profiles.

Building the iCSN starts with an analysis of the e-mail communication, instant messaging (i.e. chat; both private messages between users and conference messages are analysed) and telephone calls.

The initial weight parameters $W_{email}$, $W_{chat}$ and $W_{phone}$ express the analyst's preference of a particular means of communication in the process of calculating the total weight of the edges between the iCSN nodes. The total weight of an edge, $connWeight$, is calculated using the formula expression (1):

$$connWeight = W_{email} * T_{email} + W_{chat} * T_{chat} + W_{phone} * T_{phone} \quad (1)$$

where, $T_{email}$, $T_{chat}$ and $T_{phone}$ are the total number of the exchanged emails, chat messages and telephone calls, between the two user profiles, respectively. A newly created CSN is stored in a **.net* file aimed at further analysis by our prototype test application. Our application comprises the *Pajek* [21] tool and therefore takes a **.net* file formatted according to the rules defined by *Pajek* at the input. *Pajek* [22] (Slovene word for Spider) is a program for analysis and visualization of large networks, which is used later for a further analysis of the constructed iCSN.

An excerpt of the mentioned *.net* file that stores the CSN is shown in Fig. *4*.

| | |
|---|---|
| 1 | *Vertices 125 |
| 2 | 1 userX |
| 3 | 2 userY |
| ... | ... |
| 126 | 125 userZ |
| 127 | *Arcs |
| 128 | 1 34 2 |
| 129 | 1 55 2 |
| 130 | 1 37 3 |
| ... | ... |
| 2022 | 124 60 1 |

Fig. 4. An example of **.net* file that stores an iCSN.

It contains 2022 lines:

- *line* 1 – ("*Vertices 125") indicates that the iCSN consists of 125 user profiles;
- *line* 2 – ("1 userX") indicates that the first profile identifier of the CSN is "userX";
- *line* 127 – ("*Arcs") indicates that in the next part of the file (directed weighted) connections between the profiles of the iCSN users are described;
- *line* 128 – ("1 34 2") indicates that the profile numbered 1 (i.e. *userX*) is connected with the profile numbered as 34, and that they are connected with an edge weight equal to 2. In practice, this means that the *profile* 1 sent an e-mail or an instant message to and/or had a phone conversation with the *profile* 34 during the observed period of time. The weight, which in this case equals 2, is obtained by using the formula (1).

*B. An analysis of the corporate social network*

In the remaining part of this section, the implementation of our test application for analysing iCSN is explained. The application comprises the *Pajek* tool. The analysis is performed indirectly, through a web service.

*1) Architecture of CSN System Analysis*

Fig. 5 shows the architecture of the system for the iCSN analysis. The system consists of three parts:

- the client;
- the web service server (which is a web server);
- the application server (which is another web server).

The user initiates the analysis by sending the network stored as a **.net* file to the application (Fig. 5, interaction 1) and selects the type of analysis she wants to perform on that network. Upon the analysis being completed, a custom analysis result is received from the web server (Fig. 5, interaction 6).

The main task of the application server is to adapt the user input data so that it could be sent to the web service for analysis (Fig. 5, interaction 2). This includes transforming the network into a format that can be used as an input for *Pajek* (*Input.net*). Also, the application domain is responsible for the adaptation and presentation of the results to the client that requested the analysis (Fig. 5, interaction 6). Beside these two fundamental functions, we implemented many other features, such as the presentation of the history of analyses performed, comparing the results of previous analyses, etc.

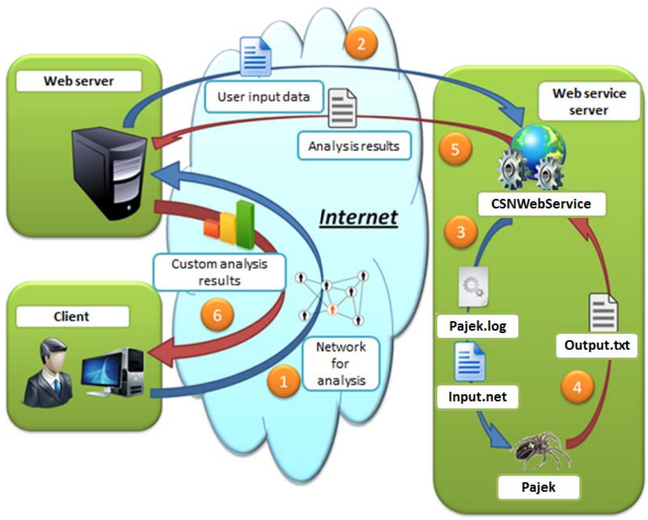


Fig. 5. The architecture of the system for network analysis.

The web service domain is in charge of carrying out the analysis of the received input data. Two entities are located in this domain - the web service (*CSNWebService*) and the tool *Pajek*. The web service *CSNWebService* prepares all the necessary files for analysis and runs the *Pajek* module (Fig. 5, interaction 3). Upon finishing the analysis, *Pajek* generates a file with the analysis result (Fig. 5, interaction 4), and then the web service forwards the result to the web server (Fig. 5, interaction 5).

*Pajek* [22] is a standalone desktop application which starts running by opening its *.exe* file. The *Pajek.log* file is the key for the implementation of *CSNWebService*. That file contains macro-commands that trigger certain actions in the *Pajek* tool [23].

Web service *CSN WebService* utilizes the possibilities of macro-commands in *Pajek* and thus allows its use in other applications. The web service *CSNWebService* is implemented as an *ASP.NET* web service [24]. The service uses standardized web protocols and formats such as SOAP (*Simple Object Access Protocol*), WSDL (*Web Services*

*Description Language*), HTTP (*Hypertext Transfer Protocol*) and XML (*Extensible Markup Language*). Due to the use of standardized technology, the use of a web service is independent of the client technology. After completing the analysis, the web service reads the results and the analysis of the results are returned to the client.

*2) Web application FER CSN Analysis*

In order to demonstrate how the web service *CSNWebService* works, we implemented a web application named *FER CSN Analysis* (Fig. 6). The web application is realized as the *Web Forms ASP.NET* web Application [25]. The *.NET Framework version* 4.0 was used. Web forms allow the separation of presentation (*HyperText Markup Language*, HTML) from the application programming logic. Program logic can be implemented in any *.NET* programming language. In the *FER CSN Analysis* application it is implemented by using *C#*. The application is hosted within the application server *IIS* (*Internet Information Services*).

The implemented functionalities of this application enable authentication and user authorization. The nodes *Hub* and *Authority* can be displayed for the analysed iCSN. Furthermore, the user is enabled to perform one of the analyses that are not built into the application, but then she must independently create a separate file with macro-commands for the analysis she wants to execute. The application also enables each user to store iCSNs and macro-files individually. Finally, the user is allowed to browse the analysis execution history, i.e. when that analysis was performed and which iCSN was analysed.


Fig. 6. The FER CSN Analysis web application home page.

*3) Hubs and Authorities*

In directed networks two important types of vertices can usually be identified: hubs and authorities. A node is a good hub if it points to many good authorities, and it is a good authority, if it is pointed to by many good hubs. Authority nodes are nodes that represent a significant source of information, while Hub nodes represent information collectors [26]. The authority score indicates the value of the node itself and hubs estimates the value of the links outgoing from the node.

In next example (Fig. 7) each node has two numbers: an authority weight $a_i$, and a hub weight $h_i$. Nodes with a higher $a_i$ number are better authorities, and nodes with a higher $h_i$ number are better hubs. Let $A$ be the adjacency matrix of the graph. Let $v$ be the authority weight vector and let $u$ be the hub weight vector:

$$v = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix}, \quad u = \begin{bmatrix} h_1 \\ h_2 \\ h_3 \\ h_4 \end{bmatrix}.$$

The adjacency matrix of the graph is:

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \end{bmatrix},$$

with its transpose:

$$A^t = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Assuming that the initial hub weight vector is:

$$u = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix},$$

we compute the authority weight vector as:

$$v = A^t * u = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} * \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 1 \\ 0 \end{bmatrix}.$$

The updated hub weight is:

$$u = A * v = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \end{bmatrix} * \begin{bmatrix} 1 \\ 3 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \\ 3 \\ 5 \end{bmatrix}.$$

Node 2 is the most authoritative, since it is the only one with three incoming arcs. Node 4 is the most important hub, while nodes 1 and 3 are equally important hubs.

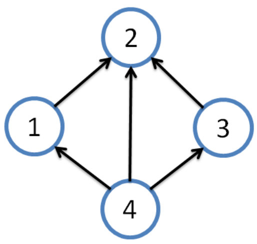


Fig. 7. A simple case-study directed network.

## VI. CONCLUSION AND FUTURE WORK

In this paper we presented a proof of concept for a general approach to developing implicit corporate social networks (iCSN) based on analysing the communication between employees through different communication channels. We analysed communication via e-mail, phone, instant messaging and file transfer within a multinational company consisting of many subsidiary companies. Separate data marts with a common backbone of shared dimensions were constructed for each communication mode. Log files for each communication mode were analysed and the data were extracted, transformed and loaded into the data marts. Based on the data stored in the corresponding data mart, the level of connection between employees was calculated for each communication mode. The total level of connection between employees was calculated by summing the levels of connection for each communication mode weighted by manually entered weights (level of importance) for each communication mode.

In our future work we will check the possibility of finding real levels of importance for each communication channel, i.e. of defining empirically the levels of importance. The research will answer the question which communication channel most firmly suggests a strong connection between employees.


## ACKNOWLEDGEMENTS

The authors acknowledge the support of research projects "Content Delivery and Mobility of Users and Services in New Generation Networks" (036-0362017-1639) and "Networked Economy" (036-0362027-1638), funded by the Ministry of Science, Education and Sports of the Republic of Croatia.


## REFERENCES


[1] D. Boyd, N. Ellison, *Social Network Sites: Definition, History, and Scholarship*, Journal of Computer-Mediated Communication 13(1), pp. 210-230, 2007.

[2] J. C. Westland, *Critical mass and willingness to pay for social networks*, Electronic Commerce Research and Applications 9(1), pp. 6-19, 2010.

[3] V. Podobnik, *Multi-Agent System for Telecommunication Service Provisioning based on User Profiles*, doctoral thesis, Zagreb, 2010.

[4] I. Bojic, T. Lipic, V. Podobnik, *Bio-inspired Clustering and Data Diffusion in Machine Social Networks*, Computational Social Networks: Mining and Visualization / editor: A. Abraham, London, Springer Verlag, pp. 51-79, 2012.

[5] N. Medman, *Doing your own thing on the net*. Ericsson Business Review 2006(1), pp. 48-53, 2006.

[6] M. Butler, D. Butler, J. Chester, *Enterprise Social Networking and Colaboration.* Martin Butler Research Ltd., East Yorkshire, UK, 2010.

[7] V. Podobnik; I. Lovrek, An Agent-based Platform for Ad-hoc Social Networking, Lecture Notes in Computer Science 6682, pp. 74-83, 2011.

[8] B. Vrdoljak, M. Banek, Z. Skocir, *Integrating XML Sources into a Data Warehouse*, Lecture Notes in Computer Science 4055, pp. 133-142, 2006.

[9] C. Y. Lin, L. Wu, Z. Wen, H. Tong, V. Griffiths-Fisher, L. Shi, D. Lubensky, *Social Network Analysis in Enterprise*, Proceedings of the IEEE, 100 (9), pp.2759-2776, Sept. 2012.

[10] C. Y. Lin, K. Ehrlich, V. Griffiths-Fisher, C. Desforges, *SmallBlue: People Mining for Expertise Search*, MultiMedia, IEEE , 15(1), pp.78,84, Jan.-March 2008.

[11] R. S. Burt, *Structural Holes and Good Ideas*, American Journal of Sociology , 110(2), pp. 349-399, September 2004.

[12] A. McAfee, *Enterprise 2.0: New collaborative tools for your organization's toughest challenges*, Harvard Business School Press, November 2009.

[13] V. Batagelj, A. Mrvar, *Pajek— Analysis and Visualization of Large Networks*, Graph Drawing, Lecture Notes in Computer Science 2265, pp 477-478, 2002.

[14] M. Bastian, S. Heymann, M. Jacomy, *Gephi: An Open Source Software for Exploring and Manipulating Networks*, Proceedings of the Third International ICWSM Conference, pp 361- 362, 2009.

[15] S. Hornibrook, *Social Network Analysis Using The SAS System*, Charlotte, NC, 2006.

[16] V. Stroele, J. Oliveira, G. Zimbrao, J. M. Souza, *Mining and Analyzing Multirelational Social Networks*, Computational Science and Engineering, CSE '09. International Conference on Computational Science and Engineering, pp.711-716, 29-31 Aug. 2009.

[17] B. T. Dai, F. C. T. Chua, E. P. Lim, *Structural Analysis in Multi-Relational Social Networks*, Structural Analysis in Multi-Relational Social Networks, pp. 451-462, 2012.

[18] R. Kimbal, T*he Data Warehouse Toolkit, 2nd edition, John Wiley & Sons,* 2002.

[19] D. M. Boyd, N. B. Ellison, *Social Network Sites: Definition, History and Scholarship*. Journal of Computer - Mediated Communication, vol. 13,no. 1, pp. 210-230, 2007.

[20] M. Hall, L. Brown, *Core Servlets and JavaServer Pages*: Core Technologies, Vol. 1 (2nd Edition), Sun Microsystems Press 2004.

[21] *Pajek NET Format* - https://gephi.org/users/supported-graph-formats/pajek-net-format/, March 2013.

[22] V. Batagelj, A. Mrvar, *Pajek - Program for Analysis and Visualization of Large Networks.* Ljubljana, September 24, 2011.

[23] V. Batagelj, A. Mrvar, *Using Macros in Pajek*, University of Ljubljana, http://vlado.fmf.uni-lj.si/pub/networks/pajek/doc/macro.htm, March 2013.

[24] R. Howard, *Web Services with ASP.NET*. Microsoft Corporation, http://msdn.microsoft.com/en-us/library/ms972326.aspx, February 22, 2001.

[25] P. D. Sheriff, *Introduction to ASP.NET and Web Forms*, PDSA, Inc., http://msdn.microsoft.com/en-us/library/ms973868.aspx, November 2001.

[26] J. M. Kleinberg, *Hubs, Authorities and Communities*, ACM Computing Survey, vol.31, 1999.