Minimal Implications Base for Social Network Analyzes

Article *in* International Journal of Web Information Systems · February 2018

DOI: 10.1108/IJWIS-04-2017-0028

4 authors:

Paula Silva
Institute for Systems and Computer Engineering, Technology and Science (INESC …

Sérgio M. Dias
Federal Service of Data Processing (SERPRO)

Mark A. J. Song
Pontifícia Universidade Católica de Minas Gerais

Luis E. Zárate
Pontifícia Universidade Católica de Minas Gerais

International Journal of Web Information Systems

Minimal implications base for social network analysis

Paula Raissa, Sérgio Dias, Mark Song, Luis Zárate,

Article information:

To cite this document:
Paula Raissa, Sérgio Dias, Mark Song, Luis Zárate, (2018) "Minimal implications base for social network analysis", International Journal of Web Information Systems, Vol. 14 Issue: 1, pp.62-77, https://doi.org/10.1108/IJWIS-04-2017-0028
Permanent link to this document:
https://doi.org/10.1108/IJWIS-04-2017-0028

Received 28 April 2017
Revised 9 July 2017
Accepted 15 July 2017

Minimal implications base for social network analysis

Paula Raissa
Campus São Gabriel, Pontificia Universidade Catolica de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil

Sérgio Dias
SERPRO Av. José Cândido da Silveira, Cidade Nova, Belo Horizonte, Brazil, and

Mark Song and Luis Zárate
Campus São Gabriel, Pontificia Universidade Catolica de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil

Abstract

Purpose – Currently, social network (SN) analysis is focused on the discovery of activity and social relationship patterns. Usually, these relationships are not easily and completely observed. Therefore, it is relevant to discover substructures and potential behavior patterns in SN. Recently, formal concept analysis (FCA) has been applied for this purpose. FCA is a concept analysis theory that identifies concept structures within a data set. The representation of SN patterns through implication rules based on FCA enables the identification of relevant substructures that cannot be easily identified. The authors' approach considers a minimum and irreducible set of implication rules (stem base) to represent the complete set of data (activity in the network). Applying this to an SN is of interest because it can represent all the relationships using a reduced form. So, the purpose of this paper is to represent social networks through the steam base.

Design/methodology/approach – The authors' approach permits to analyze two-mode networks by transforming access activities of SN into a formal context. From this context, it can be extracted to a minimal set of implications applying the NextClosure algorithm, which is based on the closed sets theory that provides to extract a complete, minimal and non-redundant set of implications. Based on the minimal set, the authors analyzed the relationships between premises and their respective conclusions to find basic user behaviors.

Findings – The experiments pointed out that implications, represented as a complex network, enable the identification and visualization of minimal substructures, which could not be found in two-mode network representation. The results also indicated that relations among premises and conclusions represent navigation behavior of SN functionalities. This approach enables to analyze the following behaviors: conservative, transitive, main functionalities and access time. The results also demonstrated that the relations between premises and conclusions represented the navigation behavior based on the functionalities of SN. The authors applied their approach for an SN for a relationship to explore the minimal access patterns of navigation.

Originality/value – The authors present an FCA-based approach to obtain the minimal set of implications capable of representing the minimum structure of the users' behavior in an SN. The paper defines and analyzes three types of rules that form the sets of implications. These types of rules define substructures of the network, the capacity of generation users' behaviors, transitive behavior and conservative capacity when the temporal aspect is considered.

Keywords Social networks, Formal concept analysis, Stem base

Paper type Research paper

The authors would acknowledge the financial support received from the Foundation for Research Support of Minas Gerais state, FAPEMIG; the National Council for Scientific and Technological Development, CNPq; Coordination for the Improvement of Higher Education Personnel, CAPES. The authors would like to thank the partial support of the Federal Service of Data Processing – www.serpro.gov.br

International Journal of Web Information Systems
Vol. 14 No. 1, 2018
pp. 62-77

1744-0084
DOI 10.1108/IJWIS-04-2017-0028

1. Introduction

A typical *social network* (SN) is formed by authors and "themes", which are combined with some relationship class. An SN can represent relationships between people, organizations, research fields, websites, a combination between them, etc. Currently, the *SN analysis* (SNA) is focused on the discovery of social relationship patterns that represents actors and their possible interactions, such as author-author or author-theme (Yamamoto *et al.*, 2016; Bianchini *et al.*, 2016; Sander *et al.*, 2017; Smatana and Butka, 2017; Jonnalagadda and Kuppusamy, 2016; Hao *et al.*, 2016). However, as shown by Freeman (2005) and Oliver *et al.* (2015) a correct identification of the network substructures requires appropriate methods to facilitate the visual analysis, extraction and representation of the information. Moreover, according to Getoor and Diehl (2005), relationships in SN are not easily and completely observed. Therefore, it is relevant to discover substructures and potential behavior patterns in SN.

Normally, the structure of an SN is modeled as a graph, whose nodes can be actors and the edges are links that represent a relationship between these actors. When the nodes are similar to each other, the network is defined as a one-mode network (Tamassia, 2013). However, with two different sets of nodes, such as users and websites, the network can be defined as two-mode networks (Breiger, 1974). Figure 1 shows an example of a network that represents users' connections to websites. The resulting graph represents a two-mode network, to which we applied the force atlas 3D algorithm (Jacomy *et al.*, 2014).

With the purpose of exploring the structures in SN in a way easy and complete, a new approach has been applied recently, the *formal concept analysis* (FCA) (Wille, 1982). The FCA is a concept analysis theory, which identifies concept structures within a data set. It applies lattice theory to hierarchically organize data from a formal context built from objects, attributes and their incidences (or relationships). The FCA has been widely studied and applied in many diverse scientific fields (Poelmans *et al.*, 2013; Li *et al.*, 2015; Mouakher and Yahia, 2016; Elloumi *et al.*, 2016; Zárate *et al.*, 2008; Zárate and Dias, 2009; Singh *et al.*, 2016).

In the past years, several authors have applied the FCA for SNA (White, 1993; Freeman, 1996; Rome and Haralick, 2005; Aufaure and Le Grand, 2013; Krajči, 2014; Castellanos *et al.*, 2017). One of the first works that used the FCA to represent network data is White (1993). Freeman (1996) and Freeman (2003) transform a one-mode network into a formal context and present an extensive meta-analysis of techniques to analyze two-mode networks to determine social groups. Freeman (2005) and Rome and Haralick (2005) explore web communities. Cuvelier and Aufaure (2011) also carry out a study unifying the FCA and

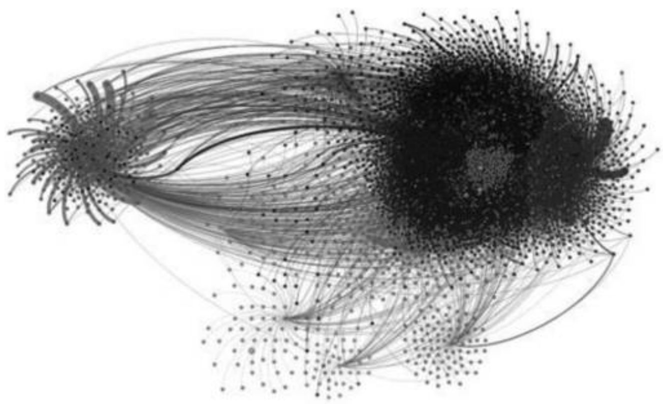

Figure 1.
Two-mode network generated from accesses of a social network for relationships – *Orkut*

complex network theories. A socially-oriented work is presented by Poelmans *et al.* (2011), where a semi-automatic process to expose a network of criminal organizations and their members is proposed. Aufaure and Le Grand (2013) also explore the FCA applicability in SNA. The authors describe the lattice expressiveness, especially when associated with ontologies. Cordero *et al.* (2015) applied the FCA to identify user's influence in an SN and using implications to build a structure containing the complete set of influences among users. Finally, Castellanos *et al.* (2017) using the FCA for topic detection in SN. Note that, these studies were motivated by the interest in understanding and interpreting SNs through mathematical formalism.

Applying the FCA in different problem domains as SNs depends on the formal context construction representing a specific problem. From these formal contexts, it is possible to apply algorithms to obtain formal concepts, concept lattice, and implications, to represent and extract knowledge from a network activity. The knowledge embedded in a formal context, concept lattice or implication set are equivalent (Ganter and Wille, 2012). In this paper, we used *implication rules* as an alternative expression of the knowledge portrayed by a formal context or concept lattice.

In the current scenario, SNs are undergoing an exponential growth, which makes it harder to observe patterns in their relationships (Figure 1). In general, the approaches based on implication rules permit the identification of access and behavior patterns typically found in the SN structure. The representation of these behaviors through implication rules based on the FCA could be very promising because it can encourage user interest and attention in SNs. In our work, we show the potential of building computational models based on implications to represent and analyze the SN.

Our approach is based on the FCA, specifically applying minimal implications base called *stem base* (or *Duquenne-Guigues base*) (Ganter and Wille, 2012) to analyze relationships in SNs. The stem base which is non-redundant provides a complete implication set so that any valid implication on formal context or concept lattice can be extracted. The stem base applied in an SN is of interest because it can represent all the relationships between attributes of a formal context using a reduced form.

Furthermore, through the set of minimal implications, it is possible to identify canonic substructures and relevant patterns that cannot be easily identified and that are capable and necessary to generate all the SN structure. We identify important patterns, in the stem base, such as SN patterns for which there are no other activities (behaviors) that lead to the same conclusion; patterns that result in the same activity, the most influential activities in the network; and patterns that exhibit transitive behavior. Note that, as we work with a minimal and non-redundant base, these implications are exclusive in the stem base. We can also obtain conservative behaviors based on rules that are repeated over time.

As a case study, we considered the access data of Orkut SN[1] which is focused on friend relationships and shared information. For characterization of the SN, we used real traffic data from a residential internet provider. The main idea is to analyze the two-mode network by transforming the activity of SN into a formal context. From this context, extract a minimal set of implications that represents premises and conclusions as nodes, and incidences as edges. In other words, the purpose is to increase the knowledge regarding SN finding a minimal set of user access patterns capable of representing the network structure. The importance of implications base is related to its ability to represent the complete set of data (activity in the network) with a minimum and irreducible set of implications rules. Thus, the NextClosure algorithm (Ganter *et al.*, 2005) was used. This algorithm is based on the closed sets theory and permits to extract a complete, minimal and non-redundant set of implications. This algorithm was chosen because it is the most widely used algorithm to get a minimum set of implications into the FCA applications.

Goncalves *et al.* (2010) and Benevenuto *et al.* (2009) also used Orkut SN as a case study for their researches. These studies were focused on characterization of the Orkut network, analyzing aspects such as access frequency and time, available activities, and interactions among users. We also explored Orkut network data as a case study, but in users' access perspective. It is important to highlight that the proposed methodology can be applied in any SN or websites, where the goal is to extract information about the access patterns. In our case, the main reason is to identify the minimal set of implications rules which represent the SN. In this paper, we focus on extract a minimal set of implications in SNs. It corresponds to an extension of preliminary results presented by Jota Resende *et al.* (2015). According to our reference review process, it was not possible to find any related work from other authors that applies a minimal set of implications to analyze access patterns on the SN.

The paper is organized as follows: In Section 2, we present the theoretical foundations concerning FCA to explain our approach. In Section 3, we present our methodology to obtain a minimal implication base and its application on the case study considered. Finally, the conclusions and future works are laid out.

2. Formal concept analysis

Firstly, we introduce some core concepts of the FCA mainly to explain the terminology. A better understanding of such concepts can be found in Ganter and Wille (2012).

In FCA, formal context is an elementary structure used to represent data and information through a cross table. This table represents a structure that defines objects (rows), attributes (columns) and their respective incidence by binary relations.

From formal context, it is possible to extract formal concepts, building concept lattices and obtaining rules which describe the relationships among objects and attributes. In the following, some preliminary definitions are presented:

Definition 2.1. Formally, a *formal context* is a triple (G,M,I), where G is a finite set of objects (rows), M is a finite set of attributes (columns) and I are incidences, which is defined as $I \subseteq G \times M$.

If an object $g \in G$ and an attribute $m \in M$ have a relationship i, their representation is $(g,m) \in I$ or gIm, which can be read as "the object g has the attribute m".

Table I presents an example of formal context, where rows are objects representing Orkut users, and the columns are Orkut pages accessed during sessions. In a formal context, when an object has an attribute, this incidence is identified and represented by "×".

Definition 2.2. Given a subset of objects $A \subseteq G$ of the formal context (G,M,I), there is an attribute subset of M common to all objects of A, even if it is empty. Likewise, given a set

User	Profile	Photo	Friends	Scraps	Album	Community	Testimonial
user1			×				
user2					×	×	
user3	×	×	×	×	×	×	×
user4	×	×	×	×	×	×	×
user5						×	
user6							×
user7					×	×	
user8				×	×		×

Table I. Example of formal context

$B \subseteq M$, there is an object subset that shares the attributes of B, even if it is empty. These relationships are defined by the *derivation operations*:

$$A' = \{m \in M | gIm \forall g \in A\} \tag{1}$$

$$B' = \{g \in G | gIm \forall m \in B\} \tag{2}$$

Definition 2.3. A formal context (G, M, I) is called *clarified context*, iff for all object $g, h \in G$ and for all attribute $m, n \in M$ it always respects the following conditions: $g' = h' \Rightarrow g = h$ and $m' = n' \Rightarrow m = n$.

The clarification process consists in maintaining one element (objects and attributes) from a set of equal elements eliminating others. In this process, the number of objects and attributes is reduced retaining the lattice structure (Ganter *et al.*, 2005). To obtain a minimal implication base, our approach considers clarified formal contexts.

Definition 2.4. The *formal concept* is defined as a pair (A,B), where $A \subseteq G$ is called extension and $B \subseteq M$ is called intention, where $A = B'$ and $B = A'$.

Considering the formal context given in Table I, we can extract, for example, the formal concept $(A,B) = (\{usu_3, usu_4\},\{Photo, Profile\})$, where elements of B are $\{Photo, Profile\}$, that, by derivation [equation (2)], $A = \{usu_3, usu_4\}$.

Applying the operator given by equation (1), we again obtain B, a condition that defines the pair (A,B) as a formal concept. This concept represents the subset of users who accessed *photo* and *profile* pages.

The set of formal concepts is ordered by the partial order such that for any two formal concepts (A_1,B_1) and (A_2,B_2), $(A_1,B_1) \preceq (A_2,B_2)$ iff $A_1 \subseteq A_2$ (equivalently, $B_2 \subseteq B_1$). The set of concepts ordered by constitutes a complete lattice, so it's called *concept lattice* (Ganter and Wille, 2012). The concept lattice obtained from a formal context (G,M,I) is denoted $B(G,M,I)$.

The first part of the *basic theorem* on concept lattices says that a concept lattice $B(G,M,I)$ is a complete lattice in which for any arbitrary set $C \subseteq B(G,M,I)$ the *infimum* and *supremum* are given by ${}^{V}C = ({}^{T}X,({}^{S}Y)')$ and $\vee C = ((\cup X)',\cap Y)$, where $X = \{A \mid (A,B) \in C\}$ and $Y = \{B \mid (A,B) \in C\}$.

2.1 Implication rules

Given a formal context (G,M,I) or a concept lattice $B(G,M,I)$, we can extract exact implication rules from them (from now on named as *implications*). The definition of an implication is given as follows:

Definition 2.5. Being a formal context whose attributes set is M. An implication is an expression $P \rightarrow Q$, which $P,Q \subseteq M$.

An implication $P \rightarrow Q$, extracted from a formal context, or respective concept lattice, has to be such that $P' \subseteq Q'$. In other words: all object which has the attributes of P, it also has the attributes of Q.

Note that, if X is a set of attributes, then X *respects* an implication $P \rightarrow Q$ iff

$P\, 6\!\subseteq X$ or $Q \subseteq X$. An implication $P \rightarrow Q$ *holds* in a set $\{X_1,\ldots,X_n\} \subseteq M$ iff each X_i respects $P \rightarrow Q$; and $P \rightarrow Q$ *is an implication of the context* (G,M,I) iff it holds in its set of object intents (an object intent is the set of its attributes). An implication $P \rightarrow Q$ *follows from* a set of implications I, iff for all set of attributes X if X respects I, then it respects $P \rightarrow Q$. A set of implications I is said to be *complete* in (G,M,I) iff every implication of (G,M,I) follows from I. A set of implications I is said to be *redundant* iff it contains an implication $P \rightarrow Q$

that follows from $I \setminus \{P \rightarrow Q\}$. Finally, an implication $P \rightarrow Q$ is considered *superfluous* iff $P \cap Q \neq \emptyset$.

For example, the case study considered in this paper, the implication {*Profile, Album, Photo*} → {*Home, >2 h*} is an implication where the premise is formed by a set of attributes {*Profile, Album, Photo*}, and the conclusion is represented by the set of attributes {*Home, >2 h*}. The example states that if a user visits the pages *profile, Album* and *photo*, he also visits the page *home* and stays connected for more than 2 h, with a confidence rate of 100 per cent.

It indicates the actions (visited pages) did by users. The information related to the user connection time (>2 h), corresponds to temporal information that is represented as a characteristic of user behavior, which is defined by a conceptual model represented through a formal context.

2.2 Minimal set of implications

Definition 2.6. Two implication sets of M are considered equivalent if they have the same closure. The closure is the set of all sets of elements formed by concatenation of zero or more elements from the original set. Therefore, if L and K are equivalent then K is a set cover of L. Thus, a minimum set cover of an implication set L corresponds to the smallest set among the remaining sets equivalent to L.

Guigues and Duquenne (1986) proved that for all context with a finite attribute set there is a complete and non-redundant set of implications.

A set $P \subseteq M$ is called pseudo-intent if and only if $P \neq P'$ and $Q' \subseteq P$ for every pseudo-intent $Q \subseteq P, Q \neq P$ (Ganter and Wille, 2012).

The implication set of the form $P \rightarrow (P' \setminus P)$, where P is pseudo closed, is called *stem base* or *Duquenne-Guigues base* (Ganter and Wille, 2012). A stem base is a minimum and complete set among the sets that are equivalent to L. The set is also non-redundant; therefore, it is not possible to remove implications from this set.

Some algorithms have been proposed to extract the stem base. Examples are the NextClosure (Guigues and Duquenne, 1986), LinClosure (Beeri and Bernstein, 1979), and Wild's Closure (Bazhanov and Obiedkov, 2011). In Bazhanov and Obiedkov (2011) the algorithms were compared and NextClosure was performed faster in most tests. This implications set is special because it is able to represent all knowledge of relationships between attributes, which describes users' activity patterns, using a reduced form.

Given a formal context (G,M,I), after applying the NextClosure algorithm to generate a stem base L, it is possible to obtain a set of implications with the following characteristics:

- a subset of implications that represent the specific features of the network: *type* $\alpha = \{P \rightarrow R \in L \mid Q \rightarrow R \text{ iff } Q = P \; \forall P,Q,R \subseteq \mathrm{P}(M)\}$;
- a subset of implications with the same conclusions that define the basic structure and the relationships present in the network: *type* $\beta = \{P \rightarrow R, S \rightarrow Q \in L \mid R = Q; \forall P,R,S,Q \subseteq \mathrm{P}(M)\}$; and
- a subset of transitive implications in which the conclusions of an implication correspond to the premise of other ones: *type* $\gamma = \{P \rightarrow Q, Q \rightarrow S \in L \mid P \rightarrow S, \forall P,Q,S \subseteq \mathrm{P}(M)\}$.

Note that *type* α, *type* β, *type* γ implications are extracted from the stem base which is sound, complete and non-redundant and it is capable to represent all SN.

The *type* α set corresponds to SN patterns for which there are no other activities (behaviors) that lead to the same conclusion. The patterns are required to represent the complete network structure. The *type* β set represents different SN patterns that result in the same activity – the most influential activities in the network. The *type* $-\gamma$ set defines network activities that exhibit transitive behavior. Note that, as we work with a minimal and non-redundant base, these implications are exclusive in the stem base. Given A, B and $C \subseteq$ P(M), if $A \rightarrow B$ and $B \rightarrow C$ (*type* γ implication), then there is no $A \rightarrow C$ (*type* α) and therefore there is not a *type* β implication with the conclusion in C.

It is possible to obtain a set of implications that is a result of the intersection of different bases for different contexts. It permits the evaluation of conservative behaviors over time. To identify conservative implication over time, the intersections of minimum sets of the formal contexts can be determined. The intersections allow us to identify structures that are more conservative. It is important to note that, we have considered as conservative behavior the implications that were repeated during the period studied.

3. Methodology for social network analysis

In this section, we describe the methodological procedures adopted to obtain the minimal set of implications, or stem base, of a social two-mode network for relationships and shared information.

3.1 Data source

For this work, the considered SN corresponds to Orkut relationship network, which was, in the last years, popular in Brazil. It is important to mention that, the adopted procedures in our approach can be applied to other SNs or websites, where the goal is to extract information about the access patterns by means of a minimal set of implications.

The access data was collected during the month of March 2009. The data was provided by a local internet service provider and were made anonymous appropriately. The initial database includes traffic logs from the service provider that recorded user session transactions. The database contains a total of 3,274,054 distinct sessions, 48,743 distinct users, and 6,319,333 accesses. The considered web sessions correspond to the accesses of Orkut pages. In this paper, a session is defined as the set of transactions made during a stipulated connection time in the internet service provider. More information about this data source is described by Goncalves *et al.* (2010).

From the analyzed URL domains based exclusively on Orkut, it was possible to identify the visited pages of each user, and to associate these pages for the several functionalities offered by the network. Therefore, it was possible to characterize the user activity based on the web pages accessed. The functionalities of Orkut network are describing as following:

- Home: refers to Orkut homepage;
- Profile: page containing the profile of a friend or other Orkut user;
- Album: users' album pages;
- Photo: photo pages;
- Scraps: message sending pages;
- Communities: community pages;
- Videos: posting of videos or accessing videos in profiles;
- Add Friend: page used to add friends or view friend requests;
- Friends: pages and profiles belonging to friends;
- Testimonials: friends' testimonial pages or testimonials written by friends;
- Logout: Orkut logout page;

- Settings: page containing account, profile and security settings; and
- Connection time: duration of a session.

3.2 Formal context modeling

To evaluate our proposal, four formal contexts were generated. Each formal context (G,M,I) was built from the Orkut navigation data that was generated by users. In this way, the objects G of each formal context represents the users identified by a *User Id*.

The formal contexts contain 13 attributes M and one of them is multivalued. The 12 first attributes of the context refer to the pages accessed by the user. The attributes correspond to the functionalities presented in the previous section. The attributes for the video and settings functionalities were removed because of the presenting low-access frequencies. The last attribute, connection time, is multivalued and corresponds to the initial and final time points for each transaction. This attribute refers to the total duration of a session (per user), having four options to select from: less than 30 min, between 30 min and 1 h, between 1 and 2 h and more than 2 h.

Finally, $i \in I$ represents that one user $g \in G$ has an attribute $m \in M$.

Table II shows an example of a formal context built from the Orkut users access data. Only 5 of the 47,029 objects (users) of the context are shown.

The experiments were conducted considering the four weeks of the month. Therefore, four formal contexts, Week 1 to Week 4, were generated. Table III provides it in detail.

3.3 Stem base applied to SNA

For the extraction of stem base from the formal contexts, the Conexp tool was used (Yevtushenko, 2000)[2]. The main reason to use this tool is that it is an implemented and validated version of NextClosure algorithm (Guigues and Duquenne, 1986).

Table IV shows the number of implications extracted from each formal context (W1 to W4) by the NextClosure algorithm. The table shows three examples of implications for each week. For example, the implication $\{a, c, e\} \rightarrow \{d\}$ should be understood as: "if the user accessed the pages *home*, *album* and *scraps* then the user also accessed the page *photo*" with a confidence rate of 100 per cent. It can reveal that every time users access the premise functionalities; they also access the conclusion functionalities.

It is important to remember that the obtained implications correspond to a minimal set of implications and are able to represent all user activities into the Orkut pages. This representation is limited by the formal context built (4 different formal contexts for each week to be analyzed), which relates the activity of users with the visited pages. Therefore, it is important to consider the construction of representative formal contexts for the SN is being analyzed.

Figure 2 shows a visual representation of the minimal sets of implications for the contexts W1 to W4. For a better visualization, a graph has been created. Where each node represents premises and conclusions, and the edges represent incidences. These representations were built using the Gephi[3] tool, considering the premises, conclusion and incidence of each implication. Minimal implications subset generated can be observed, which is important for the identification of users minimal behavior. To observe the substructures generated the Clustering Chinese Whispers (Biemann, 2006) algorithm has been applied to obtain clusters. There are 10 highlighted premises described below:

(1) P1: {Profile, Friends, > 2 h};

(2) P2: {Photo, Scrap, Friends};

Table II. Part of the formal context generated from user access data taken from the *Orkut SN*

User-Id	Home (a)	Profile (b)	Album (c)	Photo (d)	Scraps (e)	Community (f)	Add friend (h)	Friends (i)	Testimonial (j)	Logout (p)	<30 min (t)	>30 min and < 1 h (u)	>1 and < 2 h (v)	>2 h (x)
User 1					x						x			
User 2					x						x			
User 3		x									x			
User 4		x									x			
User 5				x							x			

(3) P5: {Add Friend, Logout};
(4) P6: {Testimonial, Logout};
(5) P7: {Home, Scraps, Friends};
(6) P8: {Home, Album, Photo}; and
(7) P9: {Album, Photo, > 2 h}.

Conclusions C1 to C5 were also highlighted, representing:

- C1: {Home};
- C2: {Home, > 2 h};
- C3: {Home, Profile, Album, Photo, Scraps, Add Friend, Friends, > 2 h};
- C4: {Home, Profile, Album, Photo, Scraps, Community, Add Friend, Friends, > 2 h}; and
- C5: {> 2 h}.

The graph supports us to view the distribution of premises and their conclusions. In this case, the Figure 2 shows the existence of implications *type* β with the same conclusion [e.g. C1, C2, C5 for context W1, in Figure 2(a); C1, C2, C4, C5 for context W3, in Figure 2(c)], coming from different sets of premises. These groupings correspond to the main implications that define the basic representative and generating structure of an SN. We can also observe that there are implications *type* α which have only single premise implying in one conclusion.

Applying the FCA in a social context, it is possible to identify the most influential pages for user behavior and associate accessed pages with time spent at the network. These pages

Week	Access period
W1	March 1st to March 7th
W2	March 8th to March 14th
W3	March 15th to March 21th
W4	March 22th to March 28th

Table III. Formal contexts and their characteristics access period

Context	Stem base size	Extracted implications (e.g.)
W1	73	{b, e, i} → {a, x}
		{a, c, e} → {d}
		{b, i, x} → {a}
W2	70	{b, d, e} → {x}
		{d, h} → {a, e, x}
		{i,p} → {x}
		{a,b,d,x} → {i}
W3	72	{b, d, i} → {a, x}
		{d, e, i} → {a, x}
W4	57	{a, c, d} → {x}
		{a, c, x} → {d}
		{c, d, x} → {a}

Table IV. Implications extracted by the NextClosure algorithm

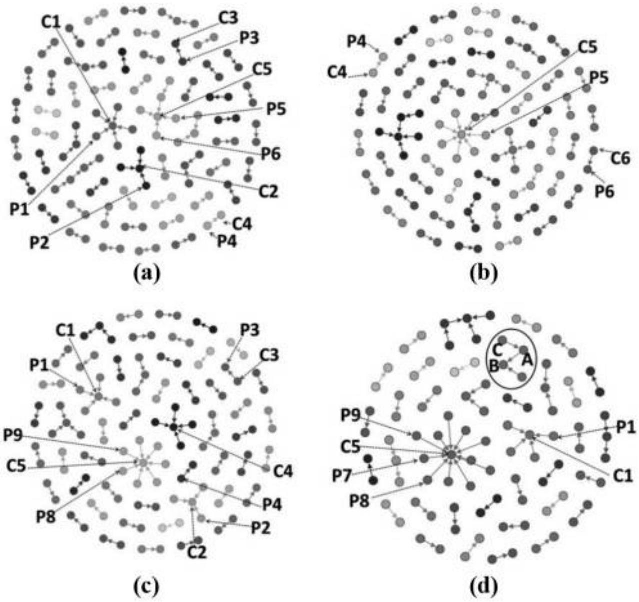

Figure 2.
Minimal implications for formal contexts W1 to W4

Notes: (a) W1; (b) W2; (c) W3; and (d) W4

were identified by nodes (conclusions and premises) with a higher degree of incidence. For example, as presented in Figure 2(d), it is possible to observe the following implications with the same conclusion: P7→C5, P8→C5 and P9→C5. The conclusion C5 is related to the users who spent more than 2 h in Orkut. Therefore, when the user accesses the pages *Home, Scraps*, and *Friends* they spent 2 h or more connected in the SN. The connection time has considerable importance to be monitored by server's administrators. We observed that the initial pages that the which was accessed by the user, it can be inferred how long the user will stay connected, leading to a possible prediction of overload on the network and to act on this prediction to maintain the desired quality of service.

From another analysis showed in Figure 2, it is possible to observe that, over the weeks, premises with different sets of attributes imply the same conclusion. It can also be observed that the same premise implies different conclusions over the weeks. These implications demonstrated variations on users' access over the weeks, explaining different browsing patterns. This can be seen in W1 [Figure 2(a)] on which P5 and P6 imply in C5, in W2 [Figure 2(b)] only P5 imply in C5 and P6 is a part of another implication. In W3 [Figure 2(c)] appear the premises P8 and P9 implying in C5, in W4 [Figure 2(d)] the implications P8 → C5 and P9 → C5 are kept and P7 imply in C5.

The conservative behavior could also be seen from the visual representation of implications. Some intersections previously analyzed also were highlighted in Figure 2, which represent the contexts W1 to W4. The implication P1 → C1 can be seen in Figure 2(a), 2(c) and 2(d). The implications P2 → C2 and P3 → C3 can be seen in Figure 2(a) and 2(c). The implication P5 → C5 can also be seen in Figure 2(a) and 2(b).

It is important to emphasize that the existence of conservative implications, obtained by the intersection of minimal sets for different data samples (weekly contexts), is expected when there is a repetition of patterns for user behavior over time. When the intersection of such implication base is empty, it indicates differences or changes in the user behavior.

When analyzing the node incidence degree, we could identify the most important or influential nodes. In this case, a bigger degree node represents the most common accesses which are different sets of pages produce the same behavior. For example, in week W4 [Figure 2(d)], the conclusion node C5 has an in-degree equal 11. In other words, there are 11 premises which imply in conclusion C5. Therefore, there are different types of web navigation that lead the user to keep connected for 2 h or more. The same occurred in week W1 [Figure 2(a)] which has five different premises lead the user to access the Orkut's *home* page.

Our approach could be also applied to analyze all users' access to internet service provider. In this case, the substructures are used to identify some user behavior types. For example, an improper behavior can be identified by the set of pages, which lead a user to access a malicious website. Another example is the network monitoring that it could be used to evaluate the nodes with higher in-degree representing the pages that can cause bottlenecks in the network. Thus, the managers would be able to do some adjustments on the network with the purpose of maintaining the quality of the services provided.

4. Conclusions and future works

In this paper, we presented an FCA-based approach to obtain a minimal structure of access behavior for two-mode SN. Specifically, we consider a minimal set of implications to the Orkut network to explore the users' pattern of navigation into Orkut functionalities. In this case, firstly, we received the real traffic data from a residential broadband provider and selected only data related to the SN considered. Next, we transformed these data into a formal context and applied the NextClosure algorithm to extract a minimal set of implications (stem base). Finally, we built a one-mode network using only the implications base, and we analyzed the relationships between premises and their respective conclusions. The main contributions of this paper are:

- An FCA-based approach to obtain the minimal set of implications (stem base) capable of representing the minimum structure of the users' behavior in an SN.
- We defined and analyzed three types of implications that form the sets of implications. These types of implications define substructures of the network (*type* α), the capacity of generation users' behaviors (*type* β), transitive (*type* γ) and conservative behavior when temporal aspect was considered.
- Our study also allowed us to analyze two-mode networks as one-mode networks applying the FCA and complex network theories to identify navigation patterns.

The experiments answer the following questions:

Q1. How do we find minimal substructures among SN functionalities?

Q2. How do we find behavior patterns and what do these behaviors represent?

For the first question, the experiments permitted that implications, represented as a complex network, enable the identification and visualization of minimal substructures, which could not be found in two-mode network representation. For the second question, the experiments demonstrated that relationships among premises and conclusions represent navigation behavior of the SN functionalities. We analyze the following behaviors: conservative, transitive, main functionalities and access time.

According to our analysis, the conservative behavior found by intersections between different time periods (formal contexts) showed that there are not many implications that repeat all of the weeks. These frequent implications have a strong relation with the time that the users stay connected to the network. We also did a critical analysis regarding the transitive behavior. We observed that this behavior could not be found only by analyzing the visual representation of implications because it is necessary to analyze when adding two premises that generate a valid implication. The main functionalities and access time analyzed were done applying the in-degree node theory. The results demonstrated there are specific relations among the set of pages, the main functionality is the *Orkut's Homepage*. Moreover, though most users have been connected for less than 2 h, there is a considerable amount of patterns produced by users who were connected for 2 h or more.

Within the scope of our new approach, we can assert: the set of implications is not unique; hence, it is possible to have several minimal sets of implications that represent the same activity of the network. However, it is possible to observe that when the implication base is obtained for different samples of the network activity, either for the same or different time periods, it is possible to find the intersection of implications which indicate that different samples have similar substructures generated.

For future research, we suggest improving the analysis related to access time, which can be done through a better discretization of access periods. The FCA-based approach should be applied to characterize other SNs by implications or apply the FCA and its implication to the SNs in different moments to evaluate behavioral shifts of the users. We also intend to explore techniques to reduce the size of the formal concept, concept lattice (Dias and Vieira, 2017, Kumar and Srinivas, 2010, Li *et al.*, 2017) and implications set (Kumar, 2012). Finally, we suggest also to explore the temporal evolution of the SN using implications based on the FCA.

Notes

1. Orkut: www.orkut.com/index.html
2. Conexp: http://conexp.sourceforge.net/
3. Gephi: https://gephi.org/

References

Aufaure, M.-A. and Le Grand, B. (2013), "Advances in FCA-based applications for social networks analysis", *International Journal of Conceptual Structures and Smart Applications*, Vol. 1 No. 1, pp. 73-89.

Bazhanov, K. and Obiedkov, S.A. (2011), "Comparing performance of algorithms for generating the Duquenne-Guigues basis", *CLA – Concept Lattices and Their Applications*, CEUR, Nancy, Vol. 959, pp. 43-57.

Beeri, C. and Bernstein, P.A. (1979), "Computational problems related to the design of normal form relational schemas", *ACM Transactions on Database Systems (TODS)*, Vol. 4 No. 1, pp. 30-59.

Benevenuto, F., Rodrigues, T., Cha, M. and Almeida, V. (2009), "Characterizing user behavior in online social networks", *Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement, ACM, Chicago, IL*, pp. 49-62.

Bianchini, D., Bianchini, D., De Antonellis, V., De Antonellis, V., Melchiori, M. and Melchiori, M. (2016), "The role of developers' social relationships in improving service selection", *International Journal of Web Information Systems*, Vol. 12 No. 4, pp. 477-503.

Biemann, C. (2006), "Chinese whispers: an efficient graph clustering algorithm and its application to natural language processing problems", *Proceedings of the First Workshop on Graph Based Methods for Natural Language Processing, Association for Computational Linguistics, Stroudsburg, PA*, pp. 73-80.

Breiger, R.L. (1974), "The duality of persons and groups", *Social Forces*, Vol. 53 No. 2, pp. 181-190.

Castellanos, A., Cigarrán, J. and Garcıa-Serrano, A. (2017), "Formal concept analysis for topic detection: a clustering quality experimental analysis", *Information Systems*, Vol. 66, pp. 24-42, ISSN 0306-4379, doi: https://doi.org/10.1016/j.is.2017.01.008

Cordero, P., Enciso, M., Mora, A., Ojeda-Aciego, M. and Rossi, C. (2015), "Knowledge discovery in social networks by using a logic-based treatment of implications", *Knowledge-Based Systems*, Vol. 87, pp. 16-25, ISSN 0950-7051, doi: https://doi.org/10.1016/j.knosys.2015.07.018

Cuvelier, E. and Aufaure, M.-A. (2011), "A buzz and e-reputation monitoring tool for Twitter based on galois lattices", in Andrews, S., Polovina, S., Hill, R. and Akhgar, B. (Eds), *Conceptual Structures for Discovering Knowledge*, Springer, Berlin, Heidelberg, pp. 91-103.

Dias, S.M. and Vieira, N.J. (2017), "A methodology for analysis of concept lattice reduction", *Information Sciences*, Vol. 396, pp. 202-217, ISSN 0020-0255, doi: https://doi.org/10.1016/j.ins.2017.02.037

Elloumi, S., Ferjani, F. and Jaoua, A. (2016), "Using minimal generators for composite isolated point extraction and conceptual binary relation coverage: application for extracting relevant textual features", *Information Sciences*, Vol. 336, pp. 129-144, ISSN 0020-0255, doi: 10.1016/j.ins.2015.12.013.

Freeman, L.C. (1996), "Cliques, galois lattices, and the structure of human social groups", *Social Networks*, Vol. 18 No. 3, pp. 173-187.

Freeman, L.C. (2003), "Finding social groups: a meta-analysis of the southern women data", *Dynamic Social Network Modeling and Analysis*, The National Academies Press, Washington, DC, pp. 39-97.

Freeman, L.C. (2005), "Graphic techniques for exploring social network data", in Carrington, P.J., Scott, J. and Wasserman, S. (Eds), *Models and Methods in Social Network Analysis*, Cambridge University Press, Cambridge, pp. 248-269, doi: 10.1017/CBO9780511811395.012.

Ganter, B., Stumme, G. and Wille, R. (2005), *Formal Concept Analysis: Foundations and Applications*, Springer Science & Business Media, New York, NY, Vol. 3626.

Ganter, B. and Wille, R. (2012), *Formal Concept Analysis: Mathematical Foundations*, Springer Science & Business Media, Secaucus, NJ.

Getoor, L. and Diehl, C.P. (2005), "Link mining: a survey", *ACM SIGKDD Explorations Newsletter*, Vol. 7 No. 2, pp. 3-12.

Goncalves, K.C., Marques-Neto, H.T. and de Almeida, J.M. (2010), "Caracterizacõao de padrões de acesso de usuários do Orkut", Anais do Simpósio Brasileiro de Sistemas Multimidia e Web (in Portuguese), *Webmedia*, *Brazil*, pp. 195-202.

Guigues, J.-L. and Duquenne, V. (1986), "Familles minimales d'implications informatives résultant d'un tableau de donńees binaires", *MathÉmatiques Et Sciences Humaines*, Vol. 95 No. 1, pp. 5-18.

Hao, F., Park, D.-S., Min, G., Jeong, Y.-S. and Park, J.-H. (2016), "k-cliques mining in dynamic social networks based on triadic formal concept analysis", *Neurocomputing*, Vol. 209, pp. 57-66, ISSN 0925-2312, doi: 10.1016/j.neucom.2015.10.141.

Jacomy, M., Venturini, T., Heymann, S. and Bastian, M. (2014), "Forceatlas2, a continuous graph layout algorithm for handy network visualization designed for the gephi software", *PloS One*, Vol. 9 No. 6, pp. 1-12.

Jonnalagadda, A. and Kuppusamy, L. (2016), "A survey on game theoretic models for community detection in social networks", *Social Network Analysis and Mining*, Vol. 6 No. 1, p. 83.

Jota Resende, G., De Moraes, N.R., Dias, S.M., Marques Neto, H.T. and Zarate, L.E. (2015), "Canonical computational models based on formal concept analysis for social network analysis and representation", *Web Services (ICWS), 2015 IEEE International Conference on*, IEEE, pp. 717-720.

Krajči, S. (2014), "Social network and formal concept analysis", Social Networks: A Framework of Computational Intelligence, *Springer International Publishing*, *Cham*, pp. 41-61.

Kumar, C.A. (2012), "Fuzzy clustering-based formal concept analysis for association rules mining", *Applied Artificial Intelligence*, Vol. 26 No. 3, pp. 274-301.

Kumar, C.A. and Srinivas, S. (2010), "Concept lattice reduction using fuzzy k-means clustering", *Expert Systems with Applications*, Vol. 37 No. 3, pp. 2696-2704.

Li, J., Kumar, C.A., Mei, C. and Wang, X. (2017), "Comparison of reduction in formal decision contexts", *International Journal of Approximate Reasoning*, Vol. 80, pp. 100-122, ISSN 0888-613X, doi: 10.1016/j.ijar.2016.08.007.

Li, J., Mei, C., Xu, W. and Qian, Y. (2015), "Concept learning via granular computing: a cognitive viewpoint", *Information Sciences*, Vol. 298, pp. 447-467, ISSN 0020-0255, doi: 10.1016/j.ins.2014.12.010.

Mouakher, A. and Yahia, S.B. (2016), "Quality cover: efficient binary relation coverage guided by induced knowledge quality", *Information Sciences*, Vols 355/356, pp. 58-73, ISSN 0020-0255, doi: 10.1016/j.ins.2016.03.009.

Oliver, K.A., de Vocht, F., Money, A. and Everett, M. (2015), "Identifying public health policymakers' sources of information: comparing survey and network analyses", *The European Journal of Public Health*, Vol. 25 No. 4, pp. 1-6.

Poelmans, J., Elzinga, P., Dedene, G., Viaene, S. and Kuznetsov, S.O. (2011), "A concept discovery approach for fighting human trafficking and forced prostitution", in Andrews, S., Polovina, S., Hill, R. and Akhgar, B. (Ed.), *Conceptual Structures for Discovering Knowledge*, Springer, Derby, pp. 201-214.

Poelmans, J., Ignatov, D.I., Kuznetsov, S.O. and Dedene, G. (2013), "Formal concept analysis in knowledge processing: a survey on applications", *Expert Systems with Applications*, Vol. 40 No. 16, pp. 6538-6560.

Rome, J.E. and Haralick, R.M. (2005), "Towards a formal concept analysis approach to exploring communities on the world wide web", *Formal Concept Analysis: Third International Conference, ICFCA 2005*, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 33-48.

Sander, T., Teh, P.L. and Sloka, B. (2017), "Your social network profile reveals you", *International Journal of Web Information Systems*, Vol. 13 No. 1.

Singh, P.K., Aswani Kumar, C. and Gani, A. (2016), "A comprehensive survey on formal concept analysis, its research trends and applications", *International Journal of Applied Mathematics and Computer Science*, Vol. 26 No. 2, pp. 495-516.

Smatana, M. and Butka, P. (2017), "Hierarchical topic modeling based on the combination of formal concept analysis and singular value decomposition", *Multimedia and Network Information Systems*, Springer, Cham, pp. 357-368, ISBN 978-3-319-43982-2.

Tamassia, R. (2013), *Handbook of Graph Drawing and Visualization*, CRC Press, Florida.

White, D.R. (1993), "Using galois lattices to represent network data", *Sociological Methodology*, Vol. 23 No. 1, pp. 127-146.

Wille, R. (1982), "Restructuring lattice theory: an approach based on hierarchies of concepts", in Rival, I. (Ed.), *Ordered Sets*, Springer, Dordrecht, pp. 445-470, ISBN 978-94-009-7798-3.

Yamamoto, S., Yamamoto, S., Wakabayashi, K., Wakabayashi, K., Kando, N., Kando, N., Satoh, T. and Satoh, T. (2016), "Twitter user tagging method based on burst time series", *International Journal of Web Information Systems*, Vol. 12 No. 3, pp. 292-311.

Yevtushenko, S.A. (2000), "System of data analysis concept explorer", *Proceedings of the 7th national conference on Artificial Intelligence KII-2000, Moscow*, pp. 127-134.

Zárate, L.E. and Dias, S.M. (2009), "Qualitative behavior rules for the cold rolling process extracted from trained ANN via the FCANN method", *Engineering Applications of Artificial Intelligence*, Vol. 22 Nos 4/5, pp. 718-731.

Zárate, L.E., Dias, S.M. and Song, M.A.J. (2008), "FCANN: a new approach for extraction and representation of knowledge from ANN trained via formal concept analysis", *Neurocomputing*, Vol. 71 Nos 13/15, pp. 2670-2684.

Corresponding author
Sérgio Dias can be contacted at: sergiomariano@gmail.com