Knowledge-Based Systems 87 (2015) 16–25

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Knowledge discovery in social networks by using a logic-based treatment of implications ☆

Pablo Cordero, Manuel Enciso, Angel Mora, Manuel Ojeda-Aciego *, Carlos Rossi

Universidad de Málaga, Spain

ARTICLE INFO

Article history:
Received 27 October 2014
Received in revised form 15 July 2015
Accepted 16 July 2015
Available online 21 July 2015

Keywords:
Formal concept analysis
Logic-based systems
Implicational systems
Social network analysis

ABSTRACT

This work can be seen as a contribution to the area of social network analysis. By considering Formal Concept Analysis (FCA) as the underlying formalizing tool, we use logic-based techniques in order to offer novel solutions to identify user's influence in a social network. We propose the use of the Simplification Logic $\mathbf{SL}_{FD}$ for attribute implications as the core of an automated method to build a structure containing the complete set of influences among users.

1. Introduction

Social influence is being studied as a research topic since long ago. The outstanding role of social networks in our life has intensified those studies in recent years. Researchers have dealt with several problems, such as the identification of opinion leaders, the behavior of trust networks and the effect of social influence in recommender systems. Social networks can be considered as a multidisciplinary research field, since they are being analyzed in the areas of Sociology [45,7], Marketing [21] or Computing [40], among others.

This paper focuses on the construction of a framework to build the influence network among individuals in a social network [27]. There are many works in the literature dealing with this problem: some of them apply link-based techniques [41], analyzing the connections of each individual (or node) to identify those who occupy the central positions of the network; other authors [10] analyze the socioeconomic features of each individual to deduce whether they are more or less influential; some recent works follow a hybrid approach [32].

One aspect to consider in the study of influence networks is that they usually present a group structure. This structure arises from the very nature of human relationships [44,39] (in most cases, one is influenced consciously or unconsciously to be accepted into a social group). The group-based behavior of traditional social networks is also given in the virtual ones. Some works [48] prove that social networks have a small-world network structure, i.e. they are highly clustered and with short distance between randomly selected users. This feature is relevant in the computational analysis of social networks, since it allows to reduce the search space in the execution of algorithms. Therefore, it is important not only to identify opinion leaders in the networks of influence, but the identification of the groups present in these networks [4].

Many of the works in the literature are based on the definition of some kind of metrics [32,26,8,10]. This way it is possible to associate a measure to the network topology, to individual features, or to the actions they do with respect to other network members. For example, in [32] the authors consider actions such as posting, viewing, replying or forwarding in an online learning network to define indicators related to expertise, novelty, influence, activity, longevity or centrality features. The values of these indicators determine the degree of influence of each member of the network.

In [39] a formal semantics of social influence was given. The authors characterize different types of social influence by combining four descriptors (positive, negative, active and passive). From these types they differentiate five subtypes that are mapped to natural language terms. This way, they construct an ontological theory of process that occupy spatio-temporal regions. The theory is expressed by means of a higher order predicate language. These authors highlighted the need for an axiom system and automatic

☆ Partially supported by Grants TIN2011-28084, TIN12-39353-C04-01 and TIN2014-59471-P of the Science and Innovation Ministry of Spain.

* Corresponding author.
E-mail addresses: pcordero@uma.es (P. Cordero), enciso@uma.es (M. Enciso), amora@uma.es (A. Mora), aciego@uma.es (M. Ojeda-Aciego), rossi@uma.es (C. Rossi).

http://dx.doi.org/10.1016/j.knosys.2015.07.018
0950-7051/

reasoning mechanisms on influence networks, but it was cataloged as future work.

The development of these reasoning mechanisms is precisely one of the contributions of this paper. Our main goal is to define a knowledge-based model of influence networks using formal concept analysis and a logic-guided method for obtaining a structure of the user's influence and the most relevant pieces of information, not needing to consider any metrics. Our method constructs the complete influence map of a network and allows for efficiently querying which individuals influence on a given one and, vice versa, which ones are influenced by somebody else. Besides, our results allow the identification of the influence groups of the network and who are opinion leaders in these groups.

Formal concept analysis (FCA) [46,19] has become a useful framework both in theoretical and applied areas. FCA has a solid mathematical foundation, and is built on lattice theory and Galois connections. FCA allows for mathematically formalizing the notion of concepts (a general idea that corresponds to some kind of entity and that can be characterized by some essential features of the class). The mechanism employed to extract the so-called concepts from a dataset is purely mathematical, and can be described in a completely abstract way within the ordered theory of lattices. It is important to recall that the concept lattice allows us to capture all the implicit knowledge which can be deduced from a context, which can be alternatively described as a set of implications.

The works related to FCA cover from data analysis, information retrieval, knowledge representation, etc. A survey about the more important applications of FCA is presented in [35], in particular, FCA has been used in the study of social networks as, for instance, trying to extract knowledge about the social relations among people, the common interests, activities, backgrounds, or real-life connections. The relationships and links between data sets on such networks can be analyzed in the domain of mathematical ontologies where the data representation is done in a hierarchical manner based on formal concepts. Thus, in [28] an ontology-based technique is proposed for a more fine-grained sentiment analysis of Twitter posts. Concerning social networks, FCA has been considered a great tool to study social communities [12]; in [5] the authors consider objects as members and their attributes as their contacts and build the formal context on which the concept lattice is generated. This lattice is then used to calculate statistics, which the authors call Conceptual Relatedness and Closeness, about every member of the social network. Recently, some authors [6] highlighted the importance of including social information in recommender systems, an area in which the authors have applied FCA to define a contextual pre-filtering process [30].

We must remark that in our approach we make a topic-based study of influence relationships; we build a model that tells us about who influences a given individual, and what piece of information that influence is based on. Specifically, we explain the whole approach on the basis of a small example based on Twitter. Other authors [32] also used this topic-based approach, although they apply a metric-based analysis of influence networks and strictly not a knowledge-based one. One intended application of our approach is to build a "social distillation" mechanism: continued use of any social network (and Twitter in particular) naturally increases the number of followed users or information sources. This overwhelms the followers with a wealth of information, a common situation for individual users and especially for community managers of companies and institutions. The mechanism proposed in this paper can solve this information overwhelming by identifying sources of "redundant" information that the user may stop following. It must be remarked that our mechanism is not intended as an interactive process, but for an on-demand (or scheduled) "social distillation" on a given set of topics.

The structure of the paper is as follows: in Section 2, the preliminary definitions related to FCA and the Simplification Logic are given; then, in Section 3, the closed sets are recalled, together with an algorithm to obtain the minimal generators associated to a given closure system; later, a small application example is carefully explained in order to show how the process of knowledge discovery works and, as a logical continuation, some general pointers to other fields of application are discussed. The final section presents some conclusions and prospects for future work.

2. Preliminaries

2.1. Formal concept analysis

As stated in the introduction, the main theoretical tool that we use in this paper arises from the area of Formal Concept Analysis (FCA). This area originated in the late eighties with the seminal works of Ganter and Wille [19] and, since then, has proven to be a useful framework on which several applications have been developed; for instance, one can see papers ranging from ontology merging [9] and resolution of fuzzy or multi-adjoint relational equations [1,14], to applications to the Semantic Web by using the notion of concept similarity or rough sets [18], and from noise control in document classification [31] to the study of fuzzy databases, in areas such as functional dependencies [33], or even linguistics [16].

Roughly speaking, the main idea underlying FCA is the extraction of implicit knowledge from a given data table. The mechanism used to extract the so-called concepts from a table is purely mathematical and can be described in a completely abstract way within the ordered theory of lattices. The background about FCA, including the following definitions, can be found in [19].

Definition 1. A *formal context* $\mathbf{K}$ is a triplet (G, M, I), where G and M are nonempty sets (representing, respectively, the set of *objects* and the set of *attributes*), and I is a relation in $G \times M$, the *incidence relation*, stating which attributes hold for any object.

Example 1. Table 1 depicts the formal context representing the use of several social networks (attributes) by age (objects). We introduce this context for illustrative purposes, since it will be used as a running example in this section.[1]

Given a formal context, essentially a data table with the relationship between objects and attributes, the concept-forming operators (also called derivation operators) are defined as follows:

Definition 2. Given a formal context $\mathbf{K} = (G, M, I)$, the *concept-forming operators* are mappings $' : 2^G \to 2^M$ and $' : 2^M \to 2^G$ defined, for all $B \subseteq G$ and $A \subseteq M$, as follows:

$$B' = \{a \in M \mid bIa \text{ for all } b \in B\}$$
$$A' = \{b \in G \mid bIa \text{ for all } a \in A\}$$

Intuitively, given a set of objects B, the corresponding concept-forming operator builds the set of attributes which satisfy all the objects in B, and vice versa. Furthermore, the pair of concept-forming operators forms a so-called *Galois connection* between the powersets of objects and attributes.

Example 2. The application of these operators to the formal context depicted in Table 1 provides, among others, the following results:

[1] This formal context is based on the use of these social networks in the UK on May 2014. https://www.onestop-webshop.co.uk/blog/social-media-connecting-communicating-people/.