Physica A 436 (2015) 443–461

Contents lists available at ScienceDirect

Physica A

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

Interests diffusion in social networks

Gregorio D'Agostino [a,c], Fulvio D'Antonio [b], Antonio De Nicola [a,d,*], Salvatore Tucci [d]

[a] ENEA: Italian National Agency for New Technologies, Energy and Sustainable Economic Development, Rome, Italy
[b] Università Politecnica delle Marche, Italy
[c] Center for Polymer Studies, Boston University, Boston, Massachusetts, United States
[d] University of Tor Vergata, Rome, Italy

HIGHLIGHTS

- We provide a model for propagation of interests on social networks.
- We provide an explicit framework to handle semantic social networks.
- We provide a software application, to assess the theory and to measure individual features.
- We analyse the DBLP dataset that represents an exhaustive repository for Computer Science.

ARTICLE INFO

Article history:
Received 7 January 2015
Received in revised form 4 March 2015
Available online 13 May 2015

Keywords:
Social networks
Semantic analysis
Interest propagation
Complex systems

ABSTRACT

We provide a model for diffusion of interests in Social Networks (SNs). We demonstrate that the topology of the SN plays a crucial role in the dynamics of the individual interests. Understanding cultural phenomena on SNs and exploiting the implicit knowledge about their members is attracting the interest of different research communities both from the academic and the business side. The community of complexity science is devoting significant efforts to define laws, models, and theories, which, based on acquired knowledge, are able to predict future observations (e.g. success of a product). In the mean time, the semantic web community aims at engineering a new generation of advanced services by defining constructs, models and methods, adding a semantic layer to SNs. In this context, a leapfrog is expected to come from a hybrid approach merging the disciplines above. Along this line, this work focuses on the propagation of individual interests in social networks. The proposed framework consists of the following main components: a method to gather information about the members of the social networks; methods to perform some semantic analysis of the Domain of Interest; a procedure to infer members' interests; and an interests evolution theory to predict how the interests propagate in the network. As a result, one achieves an analytic tool to measure individual features, such as members' susceptibilities and authorities. Although the approach applies to any type of social network, here it is has been tested against the computer science research community. The DBLP (Digital Bibliography and Library Project) database has been elected as test-case since it provides the most comprehensive list of scientific production in this field.

* Corresponding author at: ENEA: Italian National Agency for New Technologies, Energy and Sustainable Economic Development, Rome, Italy.
E-mail address: antonio.denicola@enea.it (A. De Nicola).

http://dx.doi.org/10.1016/j.physa.2015.05.062
0378-4371/

1. Introduction

Social networking platforms (SNPs) collect a huge amount of information (and hence implicit knowledge) about their members and different domains of interest. Such knowledge concerns interests, friends, best practices, activities and other facts of life. General purpose SNPs (e.g., Facebook, Twitter) contain information on several domains (e.g., music, movies, literature, travels) whereas domain-specific SNPs, e.g., anobii and LinkedIn collect information on specific topics (e.g., books for anobii and job and careers for LinkedIn).

According to McKinsey industry report of 2011 [1], the total volume of worldwide dispersed data is increasing at a rate of about 50% per year, that is around a 40-times growth in ten years. Data storage is becoming almost free since hardware devices are almost inexpensive and SNP companies gain added value by gathering data about members [2]. Management and analysis of big data involved in social networks are among the most effective activities in the scope of Data Science [2].

Organizing and managing information conveyed by SNPs in order to extract knowledge about their members is leading the market to a new generation of services focused on specific users' needs. There is a big opportunity for a paradigm shift from decisions based on "gut feelings" to decisions based on data analysis. Advanced applications exploiting social network knowledge can generate value in different sectors, such as, security, politics, business, and "social good" [3].

In this paper we study temporal evolution of people's interests in social networks. We provide evidence that the fundamental mechanism driving the propagation of interests is a diffusion and, hence, the topology of SN plays a critical role. The objective is to test the basic mechanisms and to estimate some individual features such as susceptibility and authority; that is measuring the tendency of a person to be influenced by her/his connections and her/his tendency to influence others. Estimation of these human characteristics can be used, in the long run, as a basis for the development of advanced marketing services targeted to specific individuals.

A *social network* (SN) consists of a community of "members" linked together with some kind of relationships (e.g., friendship, coauthorship, co-working). A SN is a virtual artefact originated from human activities. Developing a service leveraging on SN knowledge requires a hybrid approach based on both engineering and natural science techniques and methodologies [4]. From this perspective, we may study the temporal evolution of people's interests as a dynamic phenomenon arising in an anthropic system. Our hypothesis is that this phenomenon results from the combined action of several factors: people connections, general trends, pre-existing interests and both the attitudes of people to be influenced by or to influence others. Furthermore, we deem that, given an application domain, temporal evolution of interests depends on the topics, since people can be susceptible to some specific information more than to others: e.g., American people are usually interested on the Super Bowl rather than Europeans; whereas it is the other way around for the final of the Champions League.

The interest propagation phenomenon in social networks has been already studied by different disciplines [5] through different approaches: data mining, complexity science, semantic, and social science.

In Refs. [6,7], the authors propose a data mining approach to estimate the propagation of events (e.g. threads) and the identification of influential members. Most of the efforts in the data mining community have been devoted to define progressive models. In such models, once a node (member) becomes active (interested in a topic), it remains active. The most important propagation models are the Independent Cascade Model (ICM) and the Linear Threshold Model (LTM). Both of the previous models were first introduced in Ref. [8]. The key characteristic of ICM is that diffusion events along every arc in the social network are mutually independent; while the key characteristic of LTM is that members change their behaviour if they are exposed to multiple independent sources. Another data mining approach was presented in Ref. [9]. Here the authors propose models and algorithms to learn influence probabilities parameters from a "social graph" and a log of actions by the users.

Complexity science includes the study of complex networks [10,11]. Among the phenomena treated by this discipline, epidemics [12,5] studies the spread of viral processes in networks. The complexity science is mainly focusing on human infectious diseases and software malware spread. However there is a growing interest in studying topics diffusion in social networks [13], social dynamics [14,15] or even non consensus dynamics [16].

Merging the topological and semantic analysis of social networks represents a new and potentially fruitful research field which is providing promising results [17–21]. Our work shares the use of a semantic conceptual representation of a Domain of Interest [22] in the social network context with the formers.

A social science approach is presented in Ref. [23]. There, the authors describe an experiment performed on Facebook to estimate influential and susceptible members of social networks with respect to some social features, such as age and sex. Another interesting issue considered by the social science community is homophily (i.e., the tendency for individuals to choose friends with similar tastes and preferences) [24,25]. Our work does not deal with such issues.

In the context of social science, the concept of "meme" [26] is acquiring a growing attention representing the elementary brick for the evolution of culture and behaviour in the human communities. As such the meme is different from our concept of interest; however some authors [27] have employed the term as synonym of our concept of interest.

We treat the SN as a physical system and we model interests dynamics as a diffusion process. Like a physical system, a thermodynamic equilibrium is reached after a certain time period when no heat source is applied. Similarly in SNs, arising of new topics can be considered as a heat source that hinders the equilibrium of interests thus preventing all people to be interested in the same topics.

Our approach is based on the analysis of social network's connections and the temporal evolution of the interests of its members. We have defined a general Markov evolution process and we have tested it on a co-authorships network in