Chapter 12
Analyzing Persian Social Networks:
An Empirical Study

Leila Esmaeili
University of Qom, Iran

Mahdi Nasiri
Iran University of Science and Technology, Iran

Behrouz Minaei-Bidgoli
Iran University of Science and Technology, Iran

ABSTRACT

Analysis of data in social networks is very important for researchers, sociologists, and academics. Given the size and diversity of web data in a Web 2.0 environment, analyzing this data has been a challenge. Since data act as inputs in such projects, the accuracy of the output is directly related to the input. Good data allows for extraction of valuable knowledge. In this article, the authors present their experiences with preparation and preprocessing of data in a Persian social network. The authors also report on the analysis of the data and findings.

INTRODUCTION

In recent years there has been a growing interest in web mining using scientific, social, political, and economic techniques. Organizations in different countries invest in social network analyses for different reasons. Analyses are mainly based on three methods: content analysis, structure analysis, and usage analysis. The type of analysis determines the manner in which researchers collect and prepare their data sets.

General web 2.0-based information systems, due to their free and interactive natures generate data that is not appropriate for web mining and

knowledge discovery. A large amount of data generated is textual in nature and needs to be preprocessed before knowledge discovery. Moreover, systems change over time, and the amount and type of data generate changes. Therefore, unstructured data, which is not in the right format has to be transformed into structured, usable data that can be used in web mining studies. This process of conversion is complex and time consuming. In general, preprocessing techniques, if performed before data mining, can significantly improve the mining process, and reduce processing time.

In most web mining, data mining, text mining, and social network analysis studies and projects, data preprocessing is considered an important stage. But many of these studies are conducted without preprocessing due to the difficulty in collecting large amounts of data. Sometimes studies are conducted using smaller data sets to make preprocessing faster.

In our analysis of Persian social networks we did not come across any study that included complete preprocessing of data. Most of the research was conducted using blogs. Studies were based on data that had been collected by other researchers or in some cases Persian data was translated into English and then preprocessed (Sheykh Esmaili, Jamali, Neshati, Abolhassani, & Soltan-Zadeh, 2006; Sahebi, Oroumchian, & Khosravi, 2008). Esmaeili et al. (2011) used data stored in Parsi-yar Persian social network database to personalize recommended groups to users of the social network (Esmaeili, Nasiri, & Minaei-Bidgoli, 2011). The studied database consisted of content data and to some extent structured data. The data set was a raw one, which was analyzed for the first time. Complete preprocessing of a large volume of data in a Persian social network for the first time, classification of textual features are some of the study's strengths. In this study, we elaborate on some preprocessing experiments and provide details of statistical and network analysis of the data set.

The data set employed in our attempt included data from a Persian social network called Parsi-yar. Parsi-yar contained activities for 5 years and 6 months for 78467 users, 3359 groups within 19 categories, and 275 groups without a specified category (Table 1). Data could be classified into three categories: user information, group information, and other information. The category, other information, included user interactions in the network, their public and private messages, users' comments on messages, their friends, and their groups.

The organization of the rest of the paper is as follows. In the next section we explain the framework of data processing. This is followed by statistical analysis of research data and by network analysis of the data set. In the final section we provide a conclusion and areas of future research.

Table 1. Subjective classification of groups (active and inactive groups)

Category	Num. of groups	Category	Num. of groups	Category	Num. of groups
Revolution	100	Sport	167	History	37
Social	716	Entertainment	332	Literature	117
Sciences	191	Game	75	Morality and spirituality	129
Youth	72	Familial	28	News	71
Geography	41	Art	182	Hygiene	30
Buy and sell	31	Religion	335	Computer	352
Business	78				

Related Content

Augmented Reality Indoor Navigation Using Handheld Devices

Angelin Gladstonand Aadharshika Duraisamy (2019). *International Journal of Virtual and Augmented Reality (pp. 1-17).*

www.irma-international.org/article/augmented-reality-indoor-navigation-using-handheld-devices/228943

Strategies for Virtual Work

Paul Jacksonand Jane E. Klobas (2008). *Encyclopedia of Networked and Virtual Organizations (pp. 1535-1541).*

www.irma-international.org/chapter/strategies-virtual-work/17789

Community and Gender in the Virtual Classroom

Alred P. Rovaiand Jason D. Baker (2008). *Virtual Technologies: Concepts, Methodologies, Tools, and Applications (pp. 1451-1458).*

www.irma-international.org/chapter/community-gender-virtual-classroom/30996

Preparing for the Forthcoming Industrial Revolution: Beyond Virtual Worlds Technologies for Competence Development and Learning

Albena Antonova (2017). *International Journal of Virtual and Augmented Reality (pp. 16-28).*

www.irma-international.org/article/preparing-for-the-forthcoming-industrial-revolution/169932

The Role of Mechanics in Gamification: An Interdisciplinary Perspective

Miralem Helmefalk, Siw Lundqvistand Leif Marcusson (2019). *International Journal of Virtual and Augmented Reality (pp. 18-41).*

www.irma-international.org/article/the-role-of-mechanics-in-gamification/228944