Chapter 10
Social Media Analytics: An Application of Data Mining

Sunil Kr Pandey
Institute of Technology and Science, India

Vineet Kansal
Institute of Technology and Science, India

ABSTRACT

Many popular online social networks such as Twitter, LinkedIn, and Facebook have become increasingly popular. In addition, a number of multimedia networks such as Flickr have also seen an increasing level of popularity in recent years. Many such social networks are extremely rich in content, and contain tremendous amount of content and linkage data which can be leveraged for analysis. The linkage data is essentially the graph structure of the social network and the communications between entities; whereas the content data contains the text, images and other multimedia data in the network. The growth of the usage and penetration of social media in the recent years has been enormous and unprecedented. This significant increase in its usage and increased number of users, there has been trend of a substantial increase in the volume of information generated by users of social media. Irrespective of primary domain in which organization is operating in to, whether it is insurance sector, social media (including facebook, twitter etc), medical science, banking etc. Virtually a large number of varying nature and services of organizations are making significant investments in social media. But it is also true that many are not systematically analyzing the valuable information that is resulting from their investments. This chapter aims at providing a data-centric view of online social networks; a topic which has been missing from much of the literature and to draw unanswered research issues which can be further explored to strengthen this area.

DOI: 10.4018/978-1-4666-4213-3.ch010
INTRODUCTION

The explosion of the digital data generated as a by-product of the increasing adoption of social media means that the social sciences are flooding with data that promises to revolutionise research, but which the research community is presently not equipped to exploit. While the sheer volume of such data presents challenges for the social sciences, such data is now being routinely analysed by industry for its own purposes. Where, in the past, academic social science was an obligatory point of passage for those wanting to learn about social phenomena, there is now a danger that social scientific research is simply bypassed by powerful actors with access to vast datasets. The amount of data in our world has been exploding, and analyzing large data sets—so-called big data—will become a key basis of competition, underpinning new waves of productivity growth, innovation, and consumer surplus. (Research reports from MGI and McKinsey’s Business Technology). The increasing volume and detail of information captured by enterprises, the rise of multimedia, social media, and the Internet of Things will fuel exponential growth in data for the foreseeable future (Manyika, Chui, Brown, Bughin, Dobbs, Roxburgh, & Byers, 2011).

Regardless of where one looks at, one can see an explosion in the use of social media. Online communities have developed that focus on both personal and professional lives. Groups have been formed that focus on every potential area of interest, including food, sports, music, parenting, scrapbooking, and actuarial issues. It is estimated that there are over 900 social media sites on the internet. Some of the more popular platforms are Facebook, Twitter (Mosley, 2012), LinkedIn, Google Plus, and YouTube. To help understand the explosion in the use of social media, consider the statistics in Figure 1 which were compiled (Brown, 2011) at www.dannybrown.me.

Table 1 depicts the exponential growth in the user community joining the social media networks and amount of data that is being produced may be just imagined. The majority of the population is using social media in some form or another. In other words the substantial increase in the use of social media, there is a significant amount of information that is being generated. As seen in the same sources referenced above, the volume of this content is staggering:

- More than 30 billion pieces of content are shared each month on Facebook.
- Every minute, 24 hours of video is uploaded to YouTube.
- As of December 2010, the average number of tweets sent per day was 110 million.
- There are currently 133 million blogs listed on leading blog directory Technorati.

So not only are people joining and accessing social media sites, but they are also spending time engaging in social media and creating a significant amount of content. As a result of this time spent on social media and the information being generated, businesses have taken notice and are attempting to leverage the power of social media to help them succeed.

The growth of the “Social Web” and the corresponding rise in available “emotional text” (through on-line social network platforms such as Facebook and blogging platforms such as BlogSpot) over the past few years has led to an increased interest in sentiment analysis (Surma & Furmanek, 2011) and this is depicted in Figure 2. Research that makes use of such analysis primar-