Keynote Speech

Exploring Very Large Data Sets from Online Social Networks

Virgilio Almeida
Federal University of Minas Gerais
Antônio Carlos, 6627 - Pampulha 31270-010 - Belo Horizonte,
MG, Brazil
+55-31-3409-5860
virgilio@dcc.ufmg.br

ABSTRACT

The explosion in the volume of digital data currently available in social networks has created new opportunities for scientific discoveries in the realm of social media. In particular, I show our recent progress in user preference understanding, data mining, summarization and explorative analysis of very large data sets. In information networks where users send messages to one another, the issue of information overload naturally arises: which are the most important messages? Based on a very large dataset with more 54 million user accounts and with all tweets ever posted by the collected users - more than 1.7 billion tweets, I discuss the problem of understanding the importance of messages in Twitter

In another work based on large-scale crawls of over 27 million user profiles that represented nearly 50% of the entire network in 2011, I show a detailed analysis of the Google+ social network. I discuss the key differences and similarities with other popular networks like Facebook and Twitter, in order to determine whether Google+ is a new paradigm or yet another social network.

Categories and Subject Descriptors: J.0 General

Keywords:
Big data, Content analysis, Data mining, Social networks, User preferences.