ANALYSIS OF THE SPREAD OF INFORMATION ON TWITTER AND ITS APPLICATION TO INTERNET CONTENT DISTRIBUTION

AMIT RUHELA

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY DELHI

MAY 2016

ANALYSIS OF THE SPREAD OF INFORMATION ON TWITTER AND ITS APPLICATION TO INTERNET CONTENT DISTRIBUTION

by

AMIT RUHELA

Department of Computer Science and Engineering

Submitted

in fulfillment of the requirements of the degree of

Doctor of Philosophy

to the

Indian Institute of Technology Delhi

May 2016

To my family for their love, support and patience

Certificate

This is to certify that the thesis titled **"ANALYSIS OF THE SPREAD OF INFORMATION ON TWITTER AND ITS APPLICATION TO INTERNET CONTENT DISTRIBUTION"** being submitted by Mr. **Amit Ruhela** to the **Indian Institute of Technology Delhi**, for the award of the degree of **Doctor of Philosophy**, is a record of bona-fide research work carried out by him under my supervision. In my opinion, the thesis has reached the standards fulfilling the requirements of the regulations relating to the degree.

The results contained in this thesis have not been submitted to any other university or institute for the award of any degree or diploma.

Dr. Aaditeshwar Seth
Assistant Professor
Department of Computer Science and Engineering
Indian Institute of Technology Delhi, New Delhi 110 016

Date :
New Delhi

Acknowledgments

I would like to take this opportunity to thank everyone who contributed to the success of my PhD dissertation. First, I would like to thank my supervisor Prof. Aaditeshwar Seth, who with his insightful suggestions, technical guidance and constructive feedback made my PhD experiences a unique one. He gave me complete freedom in my research work. Moreover, he always supported me in the technical and administrative matters that considerably simplified my PhD journey.

I express my gratitude for Prof. Pankaj Jalote who encouraged me to do a PhD and supported me in the initial years. I am also thankful to Dr. Anirban Mahanti and Prof. Amitabha Baghchi who helped me throughout my research work in IIT Delhi by providing day-to-day suggestions, feedbacks, and encouragement. I would also like to thank Prof. Vinay Joseph Riberio, Prof. S. Arun Kumar and Dr. Amit A. Nanavati, who gave useful feedbacks in my SRC presentations.

I would also like to thank Prof. Naveen Garg, Prof. Preeti Ranjan Panda and Prof. Subodh Kumar, who were PhD coordinators at various points of time. The PhD seminar series initiated by them in our department was extremely helpful in getting diversified outlooks on my work.

I am grateful to my seniors Neeraj Goyal, Anant Vishnoi, Vaibhav Jain and Vikram Goyal for their encouragement and guidance during PhD. I would like to express my gratitude to Dr. Pankaj Dalela and Dr. Ashu Verma with whom I had many fruitful informal discussions on various matters related to PhD. They cannot be thanked enough for the emotional support during the troughs of my PhD life.

Discussions with my PhD colleagues Rudra Mohan Tripathi, Chinmay Narayan, Shibashis Guha, Yamuna Prasad, Namita Sharma, Tarun Beri and Dipanjan Chakraborty helped me at various points. Long discussions with Rudra Mohan Tripathi about different research possibilities in Online Social Networks laid a foundation in my research aptitude. Along with the colleagues, I am also thankful to numerous unknown reviewers of my conference papers for providing their feedback about my research ideas.

I would like to thank members of the lab staff Vandana Ahulwalia, S S Negi, M. Rathinam, Rajesh Kumar, Arun Kumar and Suresh Kumar, who made available their support in number of ways. I am also thankful to office staff and CSE department for helping me in various ways.

I express my deepest regards to Shri Vipin Tyagi, Shri Praveen Mathur, Dr. Anand Srivastav, Ravi Gupta and Yogesh Kumar who are from C-DOT where I work full time. They have always provided me the necessary support to pursue PhD alongside the official responsibilities. I am also thankful to my colleagues Priyatosh Mandal, Abhay Gupta, Sumit Saharana, Ajay Kumar, Chetan Selwal, Pramod Jhaldiyal and Abhimanyu Popli for supporting me both technically as well as morally throughout my PhD journey.

I dedicate this thesis to my father Late Sh. Sumer Singh Ruhela whose support and encouragement has always been an inspiration to me. Even during last days of his life, my father was extremely caring and enthusiastic for my research work. The endless love, immense patience and moral support of my mother (Smt. Krishna Ruhela), wife (Pooja Ruhela) and daughter (Dhani) throughout my PhD life cannot be expressed in words. I am thankful to my sisters and their in-laws who have supported me a lot in fulfilling various social obligations.

Last but not the least, I owe my deepest gratitude to the Almighty, who makes everything happen, who brings the ideas in one's mind, creates an environment to cherish those ideas, and motivates one to implement the ideas.

May 2016 Amit Ruhela

Abstract

With rapidly growing traffic volumes on the Internet, service providers are finding it increasingly hard to manage delivery of content to end users. We observe that a large portion of the traffic is contributed and accessed by Online Social Networking (OSN) websites, and we attempt to leverage this fact to build better content placement and caching strategies for Content Delivery Networks (CDNs). Through a large measurement study of OSN datasets, we show that highly popular topics cross regional geographic boundaries, and time-zone differences can be used to potentially improve content pre-fetching policies. Through an event-based analysis, we show that topic popularity peaks when multiple small components in the social graph coalesce together to form a single giant component. We also show that to track topic trends, it is sufficient to track a small percentile of the most popular users rather than track all users. Finally, we apply these insights to design OSN-aided caching policies in CDNs, but we discover that simple LRU caching does quite well unless cache sizes are very small. This is a significant negative result that can help guide other researchers interested in applying OSN signals to improve Internet performance. Our contributions also go beyond just scoping the potential of using OSN signals to aid content distribution – our insights can find applications in the placement of advertisements on OSN websites, information recommendation, and other areas of work we hope to explore in the future.

Contents

List of Figures

List of Tables