Special Issue on Searching and Mining the Web and Social Networks

Nelly Litvak and Sebastiano Vigna

The past few decades have seen the rise of online social networks as a worldwide phenomenon with a high impact on our society. Beyond the obvious exposure phenomena, with obvious implications on security and privacy, people have started to become acquainted—even married!—in online social networks. In parallel, we have seen an enormous growth in terms of the number of published articles in computer science, mathematics and physics that study the organization of such networks. The availability of large free databases of friendships, collaborations and citations have made possible to study social networks at a scale and with a precision previously unknown.

This issue of *Internet Mathematics*, titled "Searching and Mining the Web and Social Networks," was born out of the interest of the editors in the problem of searching and analyzing not only the web, but also social networks in a broad sense. In particular, we aimed to publish a collection of articles that take a rigorous mathematical viewpoint on problems most important and common in network applications. The general topics represented in this special issue cover ranking of the nodes, network measurements, and adversarial behavior. Each of these topics has received a large attention in the literature. We believe however that the originality of the articles presented in this volume is in a high level of mathematical rigor.

All submitted articles have been thoroughly reviewed in accordance with the usual high standards of *Internet Mathematics*. Each article received reviews from at least two experts: one in the field of application, and one in the relevant branch of mathematics. Some high quality submissions have been rejected because they did not sufficiently satisfied either the criterion of practical importance for social networks or did not have a sufficient mathematical depth.

The first two articles deal with the problem of detecting interesting properties of the nodes using only the structure of a social network (i.e., the underlying graph). In the article "Axioms for Centrality" by Paolo Boldi and Sebastiano Vigna, the authors try to understand the inner workings of *centrality measures*, which are designed to identify which nodes in a social network are more important than others. The article "Toward Quantifying Vertex Similarity in Networks" by Charalampos E. Tsourakakis proposes (somewhat dually) new techniques to identify *similar* nodes in large networks.

The next two articles address statistical measurements in social networks, including the in-depth mathematical analysis of the proposed estimators. The article "Degree-Degree Dependencies in Random Graphs with Heavy-Tailed Degrees" by Remco van der Hofstad and Nelly Litvak is the first rigorous study of statistical estimators for correlations between degrees of neighboring nodes in general social networks, and in common random graph models of them. The article "Estimating Sizes of Social Networks via Biased Sampling" by Liran Katzir, Edo Liberty, Oren Somekh, and Ioana A. Cosma proposes a new random walk sampling techniques for estimating network size–a basic network characteristic that is often important and unavailable in practice.

Finally, the last two articles are devoted to identifying, predicting, and preventing adversarial behavior in web and social networks. The article "Communities, Random Walks, and Social Sybil Defense" by Lorenzo Alvisi, Allen Clement, Alessandro Epasto, Silvio Lattanzi, and Alessandro Panconesi addresses the question whether a sybil attack, when an adversary introduces fake nodes and links in a graph, can be identified, based solely on graph structure. To this end, they study analytically the change of the graph characteristics under a specified model of sybil attack. The article "The Classification Power of Web Features" by Miklós Erdélyi, András A. Benczúr, Bálint Daróczy, András Garzó, Tamás Kiss, and Dávid Siklósi analyzes in a rigorous experimental setting a wide range of signals used to detect *spam* pages. Both articles also provide an excellent review of their respective topics.

We would like to thank the authors of all submissions for their high quality contributions. It has been our pleasure to receive and handle the very interesting articles submitted to this volume. We hope that it will give rise to new fascinating research on the topic.

Nelly Litvak, Department of Applied Mathematics, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands (n.litvak@utwente.nl)

Sebastiano Vigna, Dipartimento di informatica, Università degli Studi di Milano, via Comelico 39, 20135 Milano, Italy (vigna@di.unimi.it)