



Statistical analysis of stochastic processes in social networks

Natalia Markovich*

V.A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences, Moscow, Russia -
nat.markovich@gmail.com

Social networks are an important part of human activity. The statistical analysis and risk assessment in complex social networks has to take into account extremes arising as nodes with large degrees or large PageRanks. These large nodes surrounded by smaller nodes build clusters of interconnected objects. The cluster structure of the network is caused by dependence between nodes that have appeared due to social relationships and common interests. A study of the stochastic properties of these clusters may improve the dissemination of information and advertisements. Clusters may be determined as conglomerates of nodes whose degrees exceed a high threshold. To study clusters determined by the exceedances of the degree metric, we apply results of extreme value theory. Namely, the asymptotically equivalent distributions of cluster and inter-cluster sizes obtained in Markovich (2014) and the extremal index that determines the dependence measure of extremes are used. Sampling techniques like a PageRank random walk, and Metropolis-Hastings Markov chains are often used to collect information about characteristics of nodes such as their influence and popularity. To compare sampling techniques, we propose to estimate the first hitting time, i.e. the minimal time to reach a large node. To this end, we use the author's recent results with regard to the hitting time of threshold exceedances and its asymptotically equivalent geometric-like distribution and the limit mean. Secondly, we look deeper into the PageRank random walk that can be considered as a specific branching process. Jelenković and Olvera-Cravioto (2010), and Volkovich, Litvak (2010) proved that the stationary distribution of the PageRank process is heavy-tailed, regularly varying. We study extremal properties of the PageRank process. Particularly, we represent the extremal index of the PageRank process obtained in Avrachenkov, Markovich and Sreedharan (2014) by means of copulas. The exposition is provided by illustrations arising from the statistical analysis of real data that have been obtained in social networks.

Keywords: Cluster of exceedances; extremal index; first hitting time; PageRank.