University of Fribourg / Faculty of Science and Medicine / Department of Physics

Ranking with randomness
(Analysis on sports data)

Yuhao Zhou

Ranking is not only everywhere in our daily life but also one of the most basic and essential topics in scientific research. People keep pursuing perfect ranking, which is extremely important for many fields such as business, military, medical, science, and so on. However, the uncertainty of the natural world is the biggest barrier to our progress towards this goal. For example, data-driven ranking algorithms usually require us to interact and observe the real world. Sampling errors are inevitably hidden in this process and final observation data. This uncertainty of data quality will seriously affect the ranking performance of algorithms. Another example is that we live in a dynamic world, and the evolution of data on the time dimension is another important reason for uncertainty. These data uncertainties and biases put forward higher requirements for ranking algorithms.

Therefore, in recent years, studying the stability of algorithms has become a hot scientific topic. A stable algorithm means that it can maintain the original ranking accuracy when facing various uncertain environments. For example, some studies have shown that there is a strong correlation between the ranking stability and the network structure. In heterogeneous networks, ranking performance is usually more stable. Moreover, compared with the bottom part of the ranking, the top part is more resistant to randomness. Therefore, there are super-stable node groups in heterogeneous networks. Despite the results about ranking stability, the ranking performance is not always consistent with the stability. Strong stability can not guarantee a best-ranking performance in all scenarios, and algorithms with poor stability could also have their optimal scopes of applications.

Ranking in sports is an important topic that attracts much attention. Moreover, it is the most natural choice for our study due to the inherent uncertainty and complexity of sports data. In this paper, we study the influence of randomness on the performance of different ranking algorithms in detail with the background of the sports results network. We demonstrate here, for the first time, that this sensitivity combined with the natural randomness of sport greatly limits fully-global algorithms' (like PageRank) performance on results from a sports tournament. Compared with our ordinary cognition that improving information utilization is the key to better performance, our results show that appropriately reducing the utilization can enhance ranking robustness and expand the application scope of algorithms. The results on both the tournament and Association of Tennis Professionals (ATP) systems confirm that our conclusion is independent of network structure. We also prove that our robustness enhancement method can be combined with other anti-bias methods to obtain a fairer and more perfect algorithm. Besides providing specific results of ranking on sports results data, our work also highlights the need to carefully assess the actual performance and limitations of network metrics before using them. This need is exacerbated by the complexity and uncertainty of systems that produce the data, which makes it difficult to judge ex-ante if an algorithm is a good match for the data. Our models and research methods also provide a powerful tool for testing the performance of other ranking algorithms and finding the optimal algorithm in the different applications.

Jury:
Prof. Dr. Yi-Cheng Zhang (thesis supervisor)
Prof. Dr. Claudio J. Tessone (external co-examiner)
Prof. Dr. Linyuan Lü (external co-examiner)
Prof. Dr. Philipp Werner (president of the jury)