Early identification of significant nodes in growing networks

M.S. Mariani, Z.-M. Ren, Yi-C. Zhang, M. Medo
University of Fribourg

In a world where we need to make many choices and our attention is inherently limited, we often rely on automated scoring and ranking algorithms to orient ourselves in the copious amount of available information. An important class of ranking algorithms is based on the network representation of many real datasets. Network-based ranking algorithms use network representations of the data to compute the inherent value, relevance, or importance of individual nodes in the system. Widely-used examples of ranking algorithms include ranking by degree (i.e., by the number of connections received by a node) and Google's PageRank. The latter is of particular interest for the physics community since it is built on statistical-physics concepts such as diffusion and random walk. Despite the broad use of network-based ranking algorithms, we still lack a clear understanding of which network properties determine whether a given algorithm will succeed or fail in identifying the most valuable nodes in the network.
In this presentation, we focus on the dependence of degree and Google's PageRank performance on network aging effects. We start by using a recent model of growing information networks to understand PageRanks bias by node age and show how this bias damages the algorithms performance. Motivated by the observed regularities in PageRank performance patterns, we introduce a new node-scoring metric for growing networks, called rescaled PageRank score, which combines PageRank score with the explicit requirement that node score is not biased by age. Model-data analysis shows that rescaled PageRank score markedly outperforms PageRank score for a broad range of model parameters. We analyze the network of citations between the 449937 papers published in American Physical Society journals between 1893 and 2009. We find that in contrast with classical PageRank score, the rescaled PageRank score allows us to compare papers of different age on the same scale. As a positive consequence, rescaled PageRank identifies much earlier the Milestone Letters of physics, which are papers of exceptional significance selected by the editors of the APS. Finally, we show that rescaled PageRank can also early identify Oscar-awarded movies in a movies' inspiration network, and significant expert-selected patents in the US patents citation network.

[l] M. Medo, G. Cimini, S. Gualdi, Phys. Rev. Lett. **107 (23)**, 238701 (2011).
[2] S. Brin, L. Page, Comp. networks ISDN **30 (1)**, 107 (1998).
[3] M. S. Mariani et al., Sci. Rep. **5**, 16181 (2016).
[4] Z.-M. Ren et al., arxiv **1703.07656**, 1703.07656 (2017).
[5] M. S. Mariani et al., J. Informetr. **10(4)**, 1207 (2016).