Revealing the Predictability of Intrinsic Structure in Networks

Jiachen Sun[1], *Ling Feng*[2], *Jiarong Xie*[3], *Xiao Ma*[3],*Yanqing Hu*[3]

(1) School of Electronics and Information Technology, Sun Yat-sen University, Guangzhou 510006, China, sunjch6@mail2.sysu.edu.cn.

*(2) Institute of High Performance Computing, A*STAR, 138632 Singapore*

(3) School of Data and Computer Science, Sun Yat-sen University, Guangzhou 510006, China

Structure prediction in networks is among the most important and widely studied problems in network science and machine learning, which has enormous contribution to various fields from biology, recommendation systems to social media. Despite the significant progress in prediction algorithms, the fundamental predictability of structures remains unclear as networks' complex underlying formation dynamics are usually unobserved. Hence to date there has been a lack of theoretical guidance on the practical development of algorithms for their absolute performance. Here, for the first time, we find that the shortest compression length for a network structure increases linearly with the structure predictability. We observe such linear relation from networks across different domains including biological, social, economic and infrastructure networks, hinting at a possible universal class among empirical networks. In addition, our finding leads to analytical results for maximum prediction accuracy adjustable for different requirements and allows the estimation of the network dataset potential commercial values through the size of the compressed network data file.

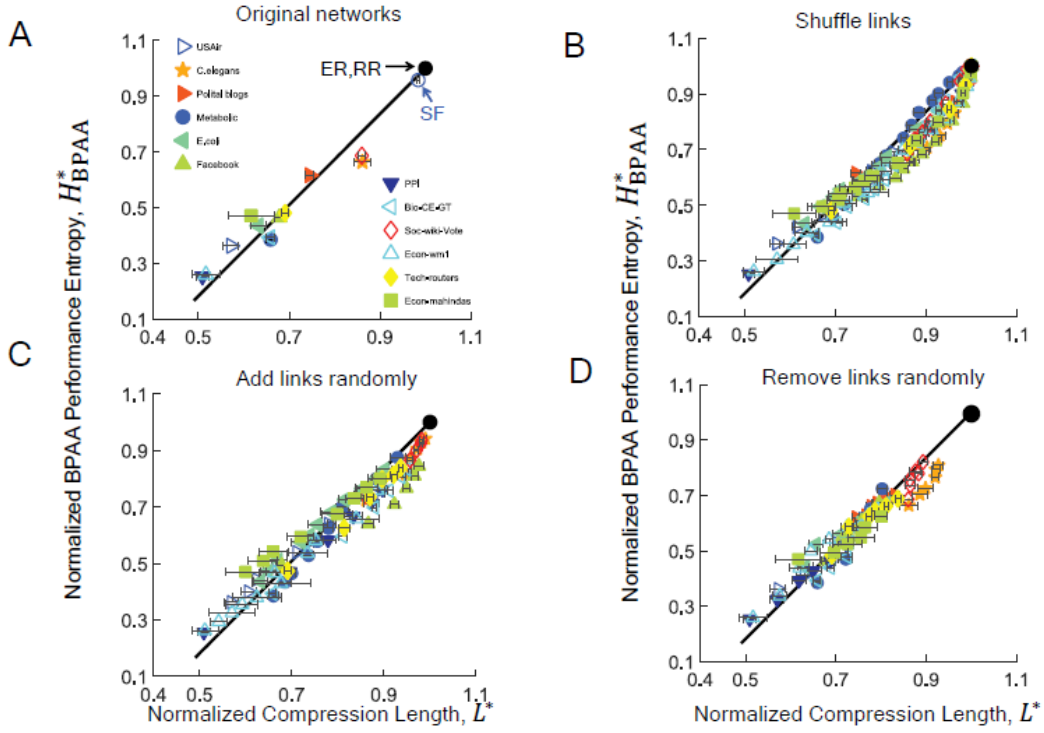

Figure 1: Network Prediction Limit vs. Shortest Compression Length on Empirical Networks.

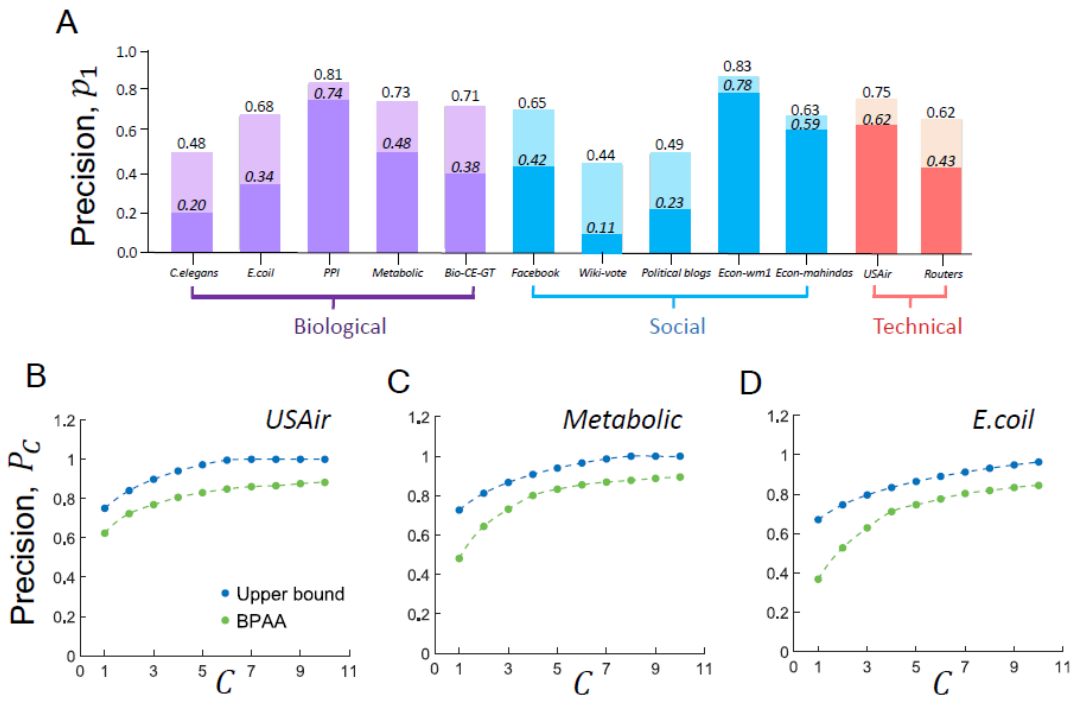

Figure 2: Upper Bound of Link Prediction Precision vs. Best Performance Algorithm Available (BPAA)'s Precision.