

ABSTRACT. This paper aims to study several link prediction algorithms as the tools for detecting missing and spurious links in noisy complex networks. Chapter 1 summarizes the available link prediction approaches for undirected networks and directed networks. Chapter 2 proposes a method to assess the likelihood of the existence of links in directed networks using the observed interactions in the networks of interest. The method, extended from the Hierarchical Random Graph (HRG) algorithm for undirected networks introduced by Clauset et al. [9], combines the Maximum Likelihood Estimation and the Markov Chain Monte Carlo methods to infer the hierarchical organization of a network, and consequently, the interactions that truly exist in the network. Chapter 3 documents a simulation study, in which we studied the ability to de-noise networks of the Common Neighbor (CN) score, the Adamic-Adar (AA) score, and the HRG algorithm in the undirected network setting, and also investigated the performance of our proposed model, benchmarked against those of two other scores, which could be considered the directed-network versions of CN and AA. The simulation results showed that the HRG method and our method generally outperformed the remaining scores. We applied our algorithm to three human transcription factor regulatory networks reported by Neph et al. [32], carried out an exploratory comparison study of the original noisy networks and the de-noised networks, and report our findings in Chapter 4. The program for the proposed method was developed based on the C++ source code provided by Clauset et al. We also fixed a logic error in this source implementation to obtain the correct likelihoods of existence for the network links.