PREFACE

Link prediction in complex networks, such as social, biological, and citation ones, has garnered increased attention from the research community. The application of link prediction is also manifold, such as anomaly detection, recommendation systems, virus spread simulation, etc. Currently, the standard paradigms of topological information-based link prediction on simple graphs are being extended to their more complex counterparts, such as multiplex and bipartite networks and knowledge graphs. This shift to dispersed information graphs allows us to research quasi-local similarity-based link prediction methods, which attempt to create a trade-off between the constrained information and simplicity of local similarity-based approaches and the thorough but computationally complex global ones.

Local similarity-based link prediction approaches do not consider the relative relevance of edges. Contemporary edge relevance quantification approaches are based on computing complex edge-based centralities. These centralities consider the number of shortest paths, simple paths, and edge's contribution to overall graph connectivity. Such approaches fail to consider the relative sphere of influence of the edge itself, which is much more constrained than the whole graph according to the 3 Degree of Influence theory. In order to address this issue and provide relevant quasi-local extensions to local similarity-based approaches, a new approach is proposed called *ELP*, a link prediction method based on the Ego perspective. Although some path-based approaches can deal with some nodes' commutative effect at some point, they are not designed to infer the total community effect of all local nodes on a specific link depending on the node proximity. Our proposed approach utilizes Ego regions to calculate edge relevance and uses this information to extend some traditional local similarity-based approaches. The best-proposed extension provides excellent results in the Accuracy score metric, which measures only the total number of correct predictions of an algorithm.

ELP aggregates different Ego regions of nodes to quantify edge relevance. This process can be visualized for a small graph by aggregating several graphs with the same node set with different edge weights. A correlation to multiplex networks with similar node sets in different layers of edge sets becomes evident. Hence in the remaining thesis,

experimentation with the concept of quasi-local similarity-based link prediction on an aggregated multiplex network, along with strategies to apply this link prediction on individual layers of the multiplex networks, is presented. Essentially the task of link prediction on a multiplex network generates layer-wise edge probabilities for non-existent edges in that particular layer.

The phenomenon of six degrees of separation on this aggregated multiplex network is studied and a novel link prediction method for multiplex networks called $HOPLP-MUL$ is proposed. Considering the aggregated structure of all the layers of a multiplex network, a complete overview of the network can be achieved, which is impossible to achieve using any single layer itself. In this proposal, it is theorized that this summarized graph (overview) provides us with an opportunity to determine the regional influence of nodes to greater certainty, and we can exploit this for more accurate link prediction. To begin, we use an aggregation model that combines information from many layers into a single summary weighted static network while accounting for the relative density of the layers. Then, we propose an algorithm $HOPLP-MUL$ which iteratively calculates link likelihoods taking longer paths between nodes into account. We also incorporate the concept of layer ranking based on densities as well as the dampening effect of longer paths on information flow.

In $HOPLP-MUL$, the relevant region for link prediction extends to two hops from the node itself, which does not incorporate much information about the global structure of the graph. To improve our link prediction accuracy, another method called $MNERLP-MUL$ is presented, which attempts to incorporate more global information in the quasi-local similarity-based approach using edge relevance as local information and node centrality-based relevance as global information. The experiments demonstrate that global node centrality-based information has more weightage to the final link prediction than the local edge relevance-based information.

Community structures are also an integral part of social network analysis. Creating communities can be considered a quasi-local information optimization operation, balancing the local information, which is intra-community separation with global inter-community separation. In the context of multiplex networks, even if the links between nodes may change depending on the specific layer, the nodes represent the same entities and share some fundamental structures (communities). Some communities may

change from layer to layer, but there exist more rigid communities (superimposed on the whole network structure) whose influence is felt across layers. For the $CLP-MUL$ community-based link prediction in multiplex networks proposal, it is theorized that nodes belonging to such a community have a greater likelihood of having connections between them (in various layers) and that this may be exploited for more accurate link prediction. This method uses information diffusion for label propagation to determine the regions of influence (rigid communities/clusters) of different central nodes. Finally, these clusters are used for calculating intra-cluster and inter-cluster similarity between node pairs for link prediction.