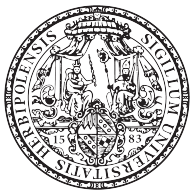

Machine Learning for Complex Networks
SoSe 2023

Prof. Dr. Ingo Scholtes
Chair of Informatics XV
University of Würzburg

Exercise Sheet 06

Published: June 7, 2022
Due: June 15, 2022
Total points: 10

Please upload your solutions to WueCampus as a scanned document (image format or pdf), a typesetted PDF document, and/or as a `jupyter` notebook.

1. Topology-based Similarity Scores

   (a) Create a synthetic network with heterogeneous degrees for which Adamic-Adar index based link prediction has larger predictive power (in terms of AUC) than Szymkiewicz-Simpson coefficient based link prediction. [2P]

   (b) Consider the cosine similarity score between two nodes v and w in an undirected network with no self-loops or multi-edges. Prove that it corresponds to the number of common neighbors normalized by the geometric mean of the degrees d_v and d_w. [2P]

   (c) Consider a random network with heterogeneous node degrees. Compute the Katz index and the Leicht-Holme-Newman index for all node pairs and plot the correlation between node similarities and the product of the node degrees for all node pairs. What do you observe? [2P]

   (d) Consider the probabilistic generative model to generate Watts-Strogatz networks with different parameters of the rewiring probability p. This model is implemented in pathpy and described in the script Statistical Network Analysis. Generate networks for different rewiring probabilities and test the AUC of different similarity scores for link prediction. How does the parameter p influence the performance of different scores? [2P]

2. Generative Models

   (a) Consider the Stochastic Block Model and explain how we can use it for link prediction in networks that exhibit a community structure. Implement and evaluate your idea in synthetic networks with community structure that are generated using the stochastic block model. Compare the performance of your approach to that of neighborhood and walk-based similarity scores. [2P]