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

Exercise Sheet 02

Published: November 10, 2022
Due: November 17, 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. **Mixing Patterns in Networks**

   (a) Prove that the degree assortativity coefficient corresponds to the fraction of the covariance $cov(d_i, d_j)$ of degrees of connected node pairs i, j and the variance $var(d_i)$ of the degree of nodes i, which are endpoints of randomly chosen edges. **3P**
   Hint: Start your proof from a definition of the average degree of nodes i, which are endpoints of randomly chosen edges (i, j) in an undirected network.

   (b) Give examples for networks with positive, negative, and zero degree assortativity. Demonstrate the correctness of your solution in `pathpy`. **1P**

   (c) Consider one of the empirical networks used in notebook 04-04 of the practice sessions[1]. Use the code from the practice session to generate a set of random microstates generated by the $G(n, p)$ random graph model, where the model parameters are fitted to the empirical network. Compute the degree assortativity for each microstate and plot the distribution of degree assortativity across all random microstates. Use this random graph ensemble to argue whether the degree assortativity that has been observed in the empirical network is significantly different from that in a corresponding random graph. **2P**

2. **Random Graph Models**

   (a) Derive an expression for the degree distribution of a microstate generated by the $G(n, p)$ model for random undirected networks with self loops. **2P**

   (b) In the lecture you have seen that the *triadic closure* of a network can be computed by averaging the local clustering coefficients of all nodes. Another, non-equivalent, way to compute the triadic closure of a network is given by the so called *global clustering coefficient*. This measure is defined as the fraction between the number T of closed triads and the total number of triads M (closed or not) in a network, i.e. $C = \frac{T}{M}$. **2P**
   Consider a $G(n, p)$ network with mean degree $c = np$. Using the independence of edges in a $G(n, p)$ compute the expected global clustering coefficient in the limit of large n.

[1] see `https://gitlab.informatik.uni-wuerzburg.de/ml4nets_notebooks/2022_wise_sna_notebooks`