Homework 3 Solutions

The average accuracy among users X:

Number of common friends: 10.66%
Jaccard's Index: 6.38%
Adamic/Adar Index: 9.44%

The average rank among newly formed edges:

Number of common friends: 2.54
Jaccard's Index: 2.19
Adamic/Adar Index: 2.37

Bonus Question:

I will consider using the "nested biased random walk" algorithm to pursue a better accuracy performance.

1. The basic idea is to calculate the probability that a walker jumps from node j to node i as described below.

   On an undirected graph, a walker takes a step from the current node, j, to node i. Assuming that each node has an attribute α_i, the probability of jumping from node j to i is given by:

   $$T_{ij}^{\alpha} = \frac{\alpha_i A_{ij}}{\sum_k \alpha_k A_{kj}},$$

2. The attribute α_i is equal to the (number of current collaborations between nodes j and i) divided by the (number of all collaborations between node j and all its neighbors). This represents how likely the node j is connected to node I given current data.
3. Then we can calculate the probability from node X to j to i using the Bayes chain rule. Firstly, calculate the probability of jumping from node X to j then multiply with the probability of jumping from node j to i. The product of the two probability can be counted as the score of the predicted new collaboration.

This algorithm is supposed to have a better performance over the provided functions, because it takes into consideration of the existing number of collaborations between same nodes instead of just count the number of neighboring nodes/common friends. It's reasonable to assume that the authors who collaborate with each other more frequently may belong to the same organizations or have similar interested topics, thus are more likely to have future collaborations. Meanwhile, this algorithm also takes into consideration the impact of the current collaborations by calculating the probability of jumping from user X to all its existing neighbors.