

John Lin

COMP790-166

2022-04-06

Reading Question #1

Paper: Refining Network Alignment to Improve Matched Neighborhood Consistency

1. Heimann et al. use matched neighborhood consistency (MNC) to improve alignment accuracy between nodes in disparate graphs, which may be helpful in unsupervised network alignments. The authors apply their method (RefiNA) across a variety of different datasets and alignment methods, showing accuracy improves even when graphs are topologically diverse.

2. The main contributions of these authors were:

- development of RefiNA (a seemingly simple, computationally efficient algorithm which improves network alignment on the premise of iteratively improving MNC)
- theoretical proof that MNC is correlated with alignment accuracy
- series of experiments showcasing that RefiNA improves alignment accuracy up to 90% across different alignment methods and can resolve matchings in datasets with 5 times as much noise

3. The authors tested five different network alignment methods (NetAlign, FINAL, REGAL, CONE-Align, and MAGNA) with and without RefiNA across six different datasets (ranging from biological to social contexts) as shown in Table 2 in the paper. They tested several different experimental conditions. First, they show that RefiNA improves accuracy and MNC with simulated noise and real noise (PPI datasets). Additionally, they attempt to show how RefiNA is helpful even when matchings are unknown, in which case, the authors use "Normalized Overlap and Largest Conserved Connected Component" to describe conservation of "edges and large substructures", respectively. Baseline models followed suggestions from literature, while the authors detail specific default hyperparameters in section 5.1.3. They looked at both dense and sparse implementations of RefiNA.

4. The authors were not the first to try to improve network alignments. They reference several existing network alignment algorithms such as NetAlign and FINAL. They specifically draw comparisons to graph neural matching networks models, which try to ensure "matching consistency between neighboring nodes". Comparably, they state that RefiNA is "faster and simpler" and does not need a supervised training environment. While there is some technical comparison to graph neural matching networks models, the authors don't seem to objectively evaluate efficiency against this specific type of model. However, they do assess "computational overhead" when added to methods: NetAlign, FINAL, REGAL, CONE-Align, and MAGNA.

5. The authors do seem to provide evidence to suggest that RefiNA does improve network alignment under certain conditions and with some observations. In Figure 2, they show greater alignment accuracies across all datasets and methods. Greater noise seems to decrease accuracy, but they claim there is an "appreciable" benefit with certain levels of noise for each

method. Admittedly, I didn't fully understand how they determined what noise percentage level was the cutoff for an "appreciable" benefit. There is a similar statement made when comparing computational overhead as shown in Figure 3. Expectedly, RefiNA adds time for both dense and sparse implementations. However, they state it is "reasonable overhead compared to the initial alignment time". It seems that the time increases more than twice in some cases... It would be interesting to see if there are any statistical differences between the bar charts in Figures 2 and 3. Perhaps I didn't fully understand, but I was also interested in their choice of hyperparameters (token match score, number of iterations, etc.) and how they chose their defaults. This being said, the authors do state limitations of their approach, such as how performance seems best with high degree nodes and how isomorphic graphs can result in misalignment with  $MNC=1$ .

6. It would be interesting to see how RefiNA assists in other biological contexts other than the PPI-H and PPI-Y datasets. For instance, it may be interesting to see how RefiNA can improve alignments in gene and protein interaction networks. Perhaps it could assist in detection of novel mechanisms in this manner. It may also shed more light on how much noise is tolerable in these contexts versus others.