Review on Pairwise Link Prediction [1]

Neelu Verma (MP19AI002)
verma.14@iitj.ac.in
Department of computer Science & Engineering, IIT Jodhpur

1 Introduction

Link predictions is a common method aims at predicting missing links in the network which predict whether there will be links between the two node based on existing link information and some attribute information. An approach for link prediction is restricted to predict a link between two nodes with the use of topology of the network. But there can be higher-order structures include more than pairs of nodes like clique. To predict links over higher-order structure for new triangles or clique. Paper proposed an approach of pair-wise Link prediction [1].

2 Approach/Methodology

The framework that is proposed for pair-wise link prediction includes three methods. A local similarity measure for pair-wise prediction and two diffusion based methods.

1. local similarity measure compute the similarity an edge and a node. They proposed three similarity measures this purpose.
   - Jaccard Similarity.
     $$JS(w,(u,v)) = \frac{|\Gamma(w) \cap \Gamma((u,v))|}{|\Gamma(w) \cup \Gamma((u,v))|}$$
   - Adamic–Adar.
     $$AA(w,(u,v)) = \sum_{z \in \Gamma(w) \cap \Gamma((u,v))} \frac{1}{\log|\Gamma(z)|}$$
   - Preferential Attachment.
     $$PA(w,(u,v)) = |\Gamma(w)| \cdot |\Gamma((u,v))|$$

2. Next two methods are based on PageRank is Pair-seeded PageRank and triangle reinforced PageRank.The Pair-seeded PageRank method is using markov chain to flow information from seed node to other node and triangle reinforced PageRank uses weights on the edges for a weighted graph and high degree of nodes for unweighted graphs that contains each edge.
   for edge (u,v) the pair - seeded PageRank equation is
   $(I - \alpha P)x_u = (1-\alpha)e_u$
   $(I - \alpha P)x_v = (1-\alpha)e_v$.

Adding the above two equations yields
$(I-\alpha P)(x_u+x_v)=(1-\alpha)(e_u+e_v)$
$(I-\alpha P)(x_u+x_v)=(1-\alpha)(2e_{u,v})$
$\frac{1}{2}(I-\alpha P)(x_u+x_v)=(1-\alpha)e_{u,v}$
$(I-\alpha P)x=(1-\alpha)e_{u,v}$.

3 Pros and Cons:

1. The approach proposed in paper is outperformed method over a local link prediction approach on various types of datasets.
2. The multiple - seeding approach used for PageRank proposed in this paper for PageRank is an outperforming approach over standard single seeded counterparts.
3. The computational efficiency is not so good and it can be improved further.
4. There is limited sub-structure in the space of higher-order structure prediction.
5. This paper have no approach of predicting a edge for single node.

4 Results

Number of experiments were performed on the synthetic and on real world graphs. So the experiment mostly focus on predicting two edges connected to the endpoints. They make prediction on a seed edge by using success probability measures (SP). Prediction sets have top k nodes which have height similarity nodes. So the prediction is in such a way that node can form a triangle with the seed edge. 500 random experiment have done for each graph. SP taking 0 or 1 value.An ROC curve is used for AUC score. Leave out edges triangle out and hold out-cross validation used for more accurate result.

5 Observation

Finding neighbourhood of an edge has a large impact on the empirical performance. So there should be an approach which is more consistent the result and should have less sensitive to the type of graph.

References

[1] Huda Nassar, Austin R. Benson, and David F. Gleich. Pairwise link prediction, 2019.
[2] https://www.cs.cornell.edu/~arb/papers/pairwise-SNAM-2020.pdf