Analyzing the Effect of Link Recommenders in Degree Inequality

Elif Ece Erdem[1], Fernando P. Santos[1],

1. Informatics Institute, University of Amsterdam, Netherlands

Algorithmic link-recommendations affect the topology of online social networks by impacting the creation of new edges [1]. Although these algorithms are effective in achieving short-term predictive accuracy, they often exacerbate inequalities by amplifying the popularity of well-connected users, overlooking niche or underrepresented groups [2]. Which link-recommendation algorithms can offer the best compromise between performance and degree equity?

To explore this question, we evaluated various link recommendation methods, spanning structural similarity-based approaches (Common Neighbors, Adamic Adar, Resource Allocation, Cosine, Jaccard, PageRank) and latent factor-based models (DeepWalk, Node2vec, Graph Convolution Networks (GCN), GraphSAGE, Graph AutoEncoder (GAE), Graph Variational AutoEncoder (GVAE)) using two distinct datasets: Twitter [3] (directed) and Facebook [4] (undirected). The Twitter dataset is split according to the link creation time, with interactions before July 9, 2015, serving as the training set and interactions after this date as the test set. For the Facebook dataset, edges are split via degree-preserving randomization [5], maintaining the degree distribution and avoiding disconnected components in the training graph.

In our workflow, models were trained or similarity scores calculated on the training dataset, and 10 recommendations per user were evaluated using Precision@10. To assess fairness, we measured the Gini index of degree distributions for two resulting networks: (1) the predicted network (all recommended links) and (2) the accepted network (the intersection of predicted links and ground truth). As both showed similar trends, we report results for predicted networks.

Figure 1 shows the relationship between Precision@10 and equality, as measured by the complement of Gini index $(1 - gini)$, for different link recommendation models on the Twitter and Facebook datasets. In both datasets, structural similarity-based methods, like Adamic-Adar and Common Neighbors, demonstrated lower equality (lower $(1 - gini)$) due to their tendency to favor highly connected nodes disproportionately. These methods also exhibited lower predictive performance. Conversely, latent factor-based approaches, particularly graph-based methods like GCN, GAE, and VGAE, achieved a better balance. On the Twitter dataset, GCN, GAE, and VGAE showed moderate precision with higher equality, while on the Facebook dataset, GAE and VGAE displayed similar trends, achieving a balance between predictive accuracy and equitable link distributions.

Our findings highlight the critical trade-offs between performance and fairness, emphasizing the need for more balanced and inclusive approaches to recommendation systems.

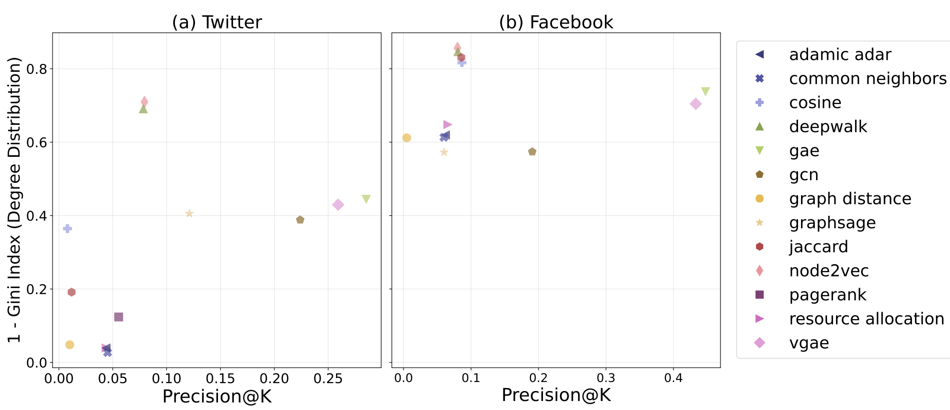

Figure 1: Gini Index vs Precision@K for all predicted links created by different models

[1] Su, J., Sharma, A., & Goel, S. (2016). The Effect of Recommendations on Network Structure. Proceedings of the 25th International Conference on World Wide Web.

[2] Stoica, A., Riederer, C.J., & Chaintreau, A. (2018). Algorithmic Glass Ceiling in Social Networks: The effects of social recommendations on network diversity. Proceedings of the 2018 World Wide Web Conference.

[3] `https://github.com/ir-uam/RELISON/tree/master/data`

[4] `https://networkrepository.com/socfb-Amherst41.php`

[5] Liu, Y., Slotine, J.E., & Barabási, A. (2012). Control Centrality and Hierarchical Structure in Complex Networks.PLoS ONE,7.