Link Prediction and Ethical Implications

Alina Baciu, Sion Coolwijk, Zsofia Katona, and Anya Nikiforova

University of Amsterdam
Fundamentals of Data Science: Assignment 1
Group A5

Abstract. This paper discusses the significance of link prediction in social networks. Link prediction can serve to uncover missing links and forecast future connections within the dynamic networks and facilitate the spread of trends and behaviors in online environments. The essay explores the ethical implications of link prediction in the context of common neighbors and community-based analysis. We highlight potential ethical concerns related to model fairness, privacy, and transparency, offering strategies to enhance fairness and balance predictions. We propose improving data quality, re-weighting data, and adjusting scoring systems to address potential biases.

Keywords: social networks · fairness · link prediction.

1 Link Prediction Task

1.1 Feature engineering

For our particular approach, we made advantage of the provided attributes. The format in which they were given, made them connected to a respective node from our graph. The approach that we too was to test all the models that were classified as link prediction models in the NetworkX documentation [13].

To utilize some of these models, it was required for the graph to have a community. As seen in Figure 1 (nodes displaced using a Directed Force Algorithm) it is clear that the provided attributes have some sort of systematical distribution. This is the reason why we decided to parse these attributes into the graph to make them function as the community in the models where one is needed.

1.2 Data preparation

Our data preparation phase began with the import of provided edges using the read_edgelist module from NetworkX. This created our graph. To make the graph more aesthetically appealing, we used a Directed Force Algorithm. These steps were done to provide a clear graphical understanding of the data.

To gain more insights in the graph, we identified clusters in the graph with the greedy_modularity_communities module from NetworkX. We also identified the clusters for the provided attributes that were given. These clusters helped us even more in understanding the communities in the graph.

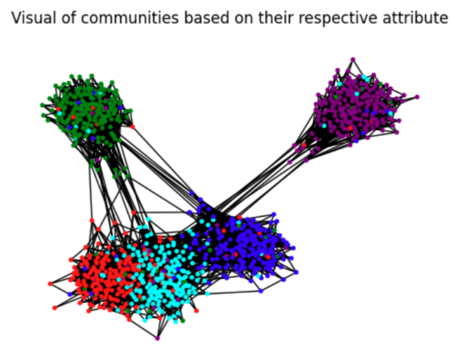

Fig. 1. This figure displays the provided attributes as communities in colors

We then proceeded to randomly remove 10% of the existing edges, equivalent to 437 edges, using the NumPy library.

After this edge removal process, we engaged in an exploration of various link prediction models available in the NetworkX library, as documented [13]. A custom function was developed for the purpose of systematically looping trough all these models and then immediately evaluating the results. The function added potential edges to our training dataset. This included positive examples for existing edges and negative examples for non-existent edges. We opted to include three times the number of existing edges in our training data to ensure comprehensive model training.

1.3 Model selection

Logistic regression, Decision Tree and Random Forest were the 3 selected supervised learning algorithms for this assignment. We decided to use the first 2 due to their fast training time and interpretability. Random Forest was also chosen as it can provide a higher accuracy than the other algorithms.

1.4 Model training and validation

The hyperparameters of the model were selected using the GridSearchCV technique, namely combining GridSearch and Cross validation on the training data. The process began by providing the grids with parameter values for all estimators: Logistic regression - 'C':[0.001,.009,0.01,.09,1,5,10,25]; Decision Tree - 'maxdepth': [1, 3, 5, 7, 10], 'minsamplessplit': [5, 10, 15, 20, 50],'minsamplesleaf':[1, 3, 5, 7, 10] and Random Forest - previous Decision Tree grid values along with 'maxfeatures':[1,3,5,7]. After assessing each combination and determining the best values for each classifier, we trained 3 different candidates using the best parameters. Choosing the best tuned candidate, we retrained it on the full training set to become the best tuned model. All in all, the chosen classifier, built using the Decision Tree machine learning algorithm and Cn Soundarajan Hopcroft modeling function, was validated using Stratified Kfold Cross validation method - F1 score.

1.5 Final results

The accuracy score of the final model on the test set was 0.82 with an AUC score of 0.73, proving that it is better than a random classifier. Concerning the solutionInput.csv, the classifier achieved an accuracy of 0.73.

1.6 Limitations

One limitation we encountered during our efforts to achieve the best predictions was our hesitation in creating more features beyond the provided attributes that were used as communities. It would have been good to do more feature engineering beforehand. For instance, we could have considered techniques like k-means clustering to introduce new features into the graph, and this approach would have made it possible to enhance the evaluation scores of the prediction more.

2 Ethical Implications

2.1 Introduction

Link prediction in online social networks is crucial due to the vast number of people interacting online. With over 4 billion people constantly connecting through various purposes, such as academia or dating, link prediction helps us understand these networks ([3], [10]). It serves two main purposes: finding missing links and predicting future connections. Social networks, which are dynamic and ever-changing, facilitate the spread of trends and behaviors on a large scale. We discuss the ethical aspects of our model in this essay, using online spaces as an illustrative example.

As shown in Section 1, our model relies on the number and similarity of the common neighbors of two nodes when predicting if there will be a link between them. Importantly, these neighbors are inspected in light of the communities they belong to, which are derived through a categorical variable that could represent a sensitive feature. These choices introduce some ethical issues in terms of model fairness, privacy and transparency, for which some examples are listed below.

- **Fairness.** As the model relies on the number of common neighbors of two nodes, as well as the number of common neighbors belonging to the same community, the model might suffer from biases towards less/more connected people, regardless of their other attributes or position in the network. As an example, in terms of information spread, this might lead to content spreading more forcefully in well-connected and populous communities, which might create an amplified effect, especially if the content plays on people's emotions or sensationalism for the sake of virality.

- **Privacy.** Privacy is a crucial issue in social applications, as predicting trends in networks often requires platforms to analyze large volumes of user data. Ethical issues arise if users are unaware of the extent of data collection and analysis for these purposes. In terms of our own model, a person's privacy might be infringed upon with respect to the control they have over the spread of their information. For instance, a user may find it difficult to reach people outside of their own circle of friends or community if an algorithm decides to decrease their visibility for people with whom they have no predicted links.
- **Transparency.** Users may not fully understand how prediction algorithms work, and companies may lack transparency in their operations, which can make it challenging to hold platforms accountable for the content they promote. In case of an algorithm such as ours, which relies on the common neighbors of nodes in a social network, media bubbles or echo chambers might be a harmful phenomenon. If users are unaware that the algorithm amplifies content aligned with their own views and downplays content going against it, they might feel a false sense of confirmation that their own opinion is the only viable one. This might be mitigated by sufficient transparency about the algorithm's workings as it could encourage users to actively seek out different viewpoints instead of blindly relying on the platform's recommendations.

In our essay, we focus on the first of these problems, namely the issue of bias arising from link prediction which relies on the quantity of common and similar neighbors of nodes. After elaborating on this limitation of the model, we aim to demonstrate some feasible mitigation strategies and discuss suggestions for future link prediction model designs.

2.2 Elaboration

In this part of the paper we focus on bias associated with link prediction. As mentioned in the introduction, the model's predictions rely on the shared connections between nodes and within the same community. Therefore, people who have more existing connections or who are part of a more populous community might be more likely to have new links predicted by the algorithm. This could result in biases in the application which favor individuals with more or fewer connections, regardless of their other characteristics or position in the network. For instance, it may cause information to spread more aggressively in large, interconnected communities, potentially amplifying the impact, particularly if the content exploits emotions or sensationalism to go viral. Furthermore, other potentially harmful impacts as well as advantages might reach members of the node in a similarly unfair manner.

For instance, research has shown that community structure can affect the speed of propagation and the activity of the users which shapes the way information is diffused [15], underlining the importance of fair link prediction. There are two different modes of information diffusion: broadcasting and contagion. Broadcasting is when just one user can send information to a large number of users in a “one-to-many” manner resulting in a network with one influential node [14]. Contagion is when information is spread person-to-person, just as in the communities investigated by Weng, Menczer and Ahn [15]. They demonstrated that community detection can be used to predict the virality of memes based on early patterns in terms of community structure.

What they found is that a high level of intra-community links, meaning people interacting more with members of the same community rather than between communities, suggests that communities can strongly trap communication. Weng, Menczer and Ahn [15] implicitly demonstrated how “echo chambers”, also known as “filter bubbles” and “information cocoons” arise in the online communities. The idea behind it is that by promoting content that aligns with a user’s existing beliefs and preferences, you create polarized communities with a wide knowledge gap between them. Even though some argue that only a small part of the population find themselves trapped in an echo chamber, the potential risks associated with such divides are quite high as they can limit limit exposure to diverse perspectives and lead to polarization of opinions [4].

Most social platforms are business that create a feed algorithm tailored to the consumer, mediating the content to fit the customers worldview which only further reiterates the creation of homophilic clusters [1], which might amplify the effect of bias if majority and minority groups are further separated. [1] suggests that the structure of the network and the clustering does differ across platforms, so that more research is needed to explore these differences. Furthermore, Guerini, Strapparava & Ozbal argued that “content virality, in social networks, hinges on the nature of the viral content itself, rather than on the simple structure of the social network” [8]. It can also be argued that the role of these elements can vary depending on the topic that is being discussed, as, for example, Wang et al. found that both the community and the sentiment of the content can play a vital role in what makes information go viral [14].

The effects of viral information spread are doubly problematic in terms of fairness as they might affect specific communities or individuals in highly different ways. One example of such an effect can be seen in a number of lawsuits against YouTube filed by LGBTQ+ content creators in 2019 [11]. The influencers believed that the videos they posted online were demonetized and censored based on the algorithm restricting LGBTQ+ content. Despite YouTube’s ML fairness initiative, the creators were not treated fairly by their algorithm. In this way, predictive algorithms can lead to social insensitivity. They may not adequately account for social, or in other cases, cultural differences, leading to a lack of promotion of content, offending to specific social groups. If an online community is a minority, the volume of nodes is already comparatively lower, which means that less links might be predicted. Therefore, besides demonatization and age-restriction of content, decreased link prediction might prevent minorities from joining the information flow or even growing through new connections.

To summarize, link prediction can detect communities and influential users and impact the spread of information. This section outlined how predicting based on the number of shared neighbors might influence the diffusion of information in the online world, how trend detection algorithms can lead to the creation of echo chambers as well as algorithmic bias and social insensitivity. The ethical concerns raised here are all related to the fairness of algorithms.

2.3 Mitigation

As we have identified that our link prediction model could introduce issues with regards to fairness, this section is dedicated to exploring how these concerns could be mitigated. Therefore, we will discuss options to counter the bias and unfairness evoked by predicting links based on the mutual (and similar) neighbors of nodes (i.e., people) in a social network.

Firstly, successful mitigation requires us to know where the problem originates: whether it is a result of faulty data collection (possibly due to human prejudices) or whether it comes from the flaws of the model. Our mitigation strategies will address these two possibilities separately below.

As discussed in the previous section, we believe that social links might be difficult to interpret and could be perceived wrongly in communities which are different from the data collector's own, especially if these are minorities. For this reason, domain knowledge and expertise in handling sensitive data from different minority groups is crucial. Particularly when recording social links (i.e., edges in the network), we believe that the researcher must be highly aware of the risk of bias and prejudice when surveying groups to which they themselves do not belong. Quantifying bias through different metrics during data collection (e.g., the researcher's probability to record a link within different communities), comparing to existing literature, as well as validation of the methods by the minority group (e.g., confirming whether the understanding of the social link at hand, for instance "friendship", is the same across groups) might be effective approaches to create an unbiased dataset.

At the same time, automatic data collection from online social platforms might be problematic for different reasons, and such data might not be representative of the actual population. One potential approach to resolve this issue is proposed in [2]. Namely, the weight of samples could be adjusted to reflect proportions in the actual population (in [2], this is done by comparing to census data). At the same time, as González-Bailón et al. explain, data collection from online social networks is also problematic because of the limited transparency of the APIs through which the data is downloaded [7]. Therefore, in such cases, unbiased data collection should start with an understanding of the limitations of the API, especially with respect to users and content on the periphery of the network [12] (such as exactly the minority communities or less well-connected people discussed previously). Additionally, there exist methods which are designed to collect more representative data about social networks by traversing online social graphs, such as the ones presented by Gjoka et al. [6]. Finally, as Hargittai suggests, it is important that even if biases remain, research on social networks is always designed while being mindful of them [9] - we believe that this holds true for any application, and that any application design should consider the potential biases in the underlying data to ensure the best possible level of fairness.

On the other hand, different mitigation strategies are required if the bias stems from the prediction model rather than the data. In our case, the probability of a link between two nodes is judged by a score which is a simple sum of the number of their shared neighbors plus one for each shared neighbor which belongs to the same community as the both of them do. It follows that, as discussed, nodes with a higher number of connections (or communities with more members) are more likely to achieve a high score.

Therefore, in order to mitigate this problem, a normalization step could be introduced, which would transform this simple sum to a ratio which compares the number of shared neighbors to the absolute number of connections of a node. This way, people towards the periphery of the network could have a more balanced likelihood to have a new link predicted.

Similarly, as the current score includes the number of shared neighbors from the same community, smaller communities might have a lower likelihood of predicted links. This is especially problematic if the social links recorded in these communities (possibly due to data collection bias, as discussed before) are sparser in the first place. In that case, the likelihood of a link predicted for nodes within minorities could be severely decreased and the predictions skewed. Therefore, a normalization term which balances the score for the size of the communities which the nodes belong to could counteract this effect.

The problem of more well-connected nodes receiving more new predicted links is also referred to as the *rich get richer* effect and has been recognized in literature. For instance, Ferrara et al. suggest that vector-based link predictions (where nodes are represented by embedding vectors) could be effective in countering this effect as they rely more on the proximity of two nodes in the embedding space than on the number of connections [5]. It is possible then that such an approach in our model could also help in creating fairer predictions.

2.4 Conclusion

In this essay, we examined the ethical implications of link prediction in a social network based on the shared number of neighbors of two nodes. We established that such an approach could be more likely to predict new links for already well-connected people as well as in more populous communities, possibly amplifying harmful effects in minority communities and for less connected individuals. We proposed several strategies to create more balanced predictions. Firstly, the quality of manually collected data can be improved (e.g., through validation by minority communities, reflecting on bias metrics, etc.) as well as that of automatically collected data from social networks (e.g., through the re-weighting of instances using ground truth population data or applying graph traversing methods to create a representative sample). Secondly, our model could be improved by adjusting the scoring system to normalize for the number of connections of the nodes as well as the size of their communities to balance the likelihood of a predicted node in the centre and the periphery of the network. Furthermore, link prediction based on embedding vector has also been suggested to mitigate the *rich get richer* effect [5].

At the same time, it is important to note the limitations of our paper. Importantly, the argumentation could be expanded with more examples and implications in contexts outside of virtual spaces since our reasoning mostly focused on online social networks. Furthermore, more specific mitigation strategies could be developed with particular applications in mind, as ”fair” predictions might mean different things in different usecases of the algorithm. Bearing these in mind, we hope that our essay provided general insight into the ethical implications of link prediction in social networks using the number of common neighbors of nodes as input, as well as outlined some strategies for mitigation and emphasized the importance of mindful model design in social applications.

References

1. Cinelli, M., De Francisci Morales, G., Galeazzi, A., Quattrociocchi, W., Starnini, M.: The echo chamber effect on social media. Proceedings of the National Academy of Sciences **118**(9), e2023301118 (2021)
2. Culotta, A.: Reducing sampling bias in social media data for county health inference. In: Joint Statistical Meetings Proceedings. pp. 1–12. Citeseer (2014)
3. Daud, N.N., Ab Hamid, S.H., Saadoon, M., Sahran, F., Anuar, N.B.: Applications of link prediction in social networks: A review. Journal of Network and Computer Applications **166**, 102716 (2020)
4. Dubois, E., Blank, G.: The echo chamber is overstated: the moderating effect of political interest and diverse media. Information, communication & society **21**(5), 729–745 (2018)
5. Ferrara, A., Espín-Noboa, L., Karimi, F., Wagner, C.: Link recommendations: Their impact on network structure and minorities. In: Proceedings of the 14th ACM Web Science Conference 2022. pp. 228–238 (2022)
6. Gjoka, M., Kurant, M., Butts, C.T., Markopoulou, A.: A walk in facebook: Uniform sampling of users in online social networks. arXiv preprint arXiv:0906.0060 (2009)
7. González-Bailón, S., Wang, N., Rivero, A., Borge-Holthoefer, J., Moreno, Y.: Assessing the bias in samples of large online networks. Social Networks **38**, 16–27 (2014). https://doi.org/https://doi.org/10.1016/j.socnet.2014.01.004, https://www.sciencedirect.com/science/article/pii/S0378873314000057
8. Guerini, M., Strapparava, C., Ozbal, G.: Exploring text virality in social networks. In: Proceedings of the International AAAI Conference on Web and Social Media. vol. 5, pp. 506–509 (2011)
9. Hargittai, E.: Is bigger always better? potential biases of big data derived from social network sites. The ANNALS of the American Academy of Political and Social Science **659**(1), 63–76 (2015)
10. Kemp, S.: Digital in 2018: World’s internet users pass the 4 billion mark, https://wearesocial.com/uk/blog/2018/01/global-digital-report-2018/
11. Kleeman, J.: ’we can’t reach the women who need us’: the lgbt youtubers suing the tech giant for discrimination. The Guardian (2019)
12. Morstatter, F., Liu, H.: Discovering, assessing, and mitigating data bias in social media. Online Social Networks and Media **1**, 1–13 (2017). https://doi.org/https://doi.org/10.1016/j.osnem.2017.01.001, https://www.sciencedirect.com/science/article/pii/S2468696416300040

13. NetworkX: Link prediction, https://networkx.org/documentation/stable/reference/
14. Wang, X., Chen, L., Shi, J., Peng, T.Q.: What makes cancer information viral on social media? Computers in Human Behavior **93**, 149–156 (2019)
15. Weng, L., Menczer, F., Ahn, Y.Y.: Virality prediction and community structure in social networks. Scientific reports **3**(1), 1–6 (2013)