
Explainable Link Prediction for Privacy-Preserving Contact Tracing

Balaji Ganesan
IBM Research India
Bengaluru, India
bganesa1@in.ibm.com

Hima Patel
IBM Research India
Bengaluru, India
himapatel@in.ibm.com

Sameep Mehta
IBM Research India
Bengaluru, India
sameepmehta@in.ibm.com

Abstract

Contact Tracing has been used to identify people who were in close proximity to those infected with SARS-Cov2 coronavirus. A number of digital contact tracing applications have been introduced to facilitate or complement physical contact tracing. However, there are a number of privacy issues in the implementation of contact tracing applications, which make people reluctant to install or update their infection status on these applications. In this concept paper, we present ideas from Graph Neural Networks and explainability, that could improve trust in these applications, and encourage adoption by people.

1 Introduction

Contact Tracing is the task of searching people who might have come in physical contact with individuals of interest, like those who have tested positive for SARS-Cov2 coronavirus. Digital Contact Tracing is the version of contact tracing which relies on mobile phones and/or wearable devices for monitoring events when individuals were in close proximity.

As governments have tried to encourage the use of such digital contact tracing applications in response to COVID19, a number of privacy related issues were raised. Bengio et al. [2020] proposed a number of recommendations for contact tracing applications. Although a number of solutions have been proposed including those based on federated learning, the adoption of contact tracing apps by people has been lukewarm. As shown by MIT COVID Tracing Tracker [2020], the adoption is less than 15% of the population in several countries that have seen significant COVID19 outbreaks.

The low adoption of contact tracing and the related exposure notification apps have lead to concerns that these apps are not going to work if people are not motivated to use them Bengio [2020]. Even among the people who have downloaded such apps, the usage remains particularly low, making them ineffective in combating the spread of SARS-Cov2 virus. Abueg et al. [2020] argued that contact tracing apps could be useful even when the adoption is low. But increased adoption and information sharing benefits the society at large.

In this work, we propose ideas that can encourage the adoption of digital contact tracing and exposure notification applications, while adhering to the privacy considerations. First, we propose that not all instances of proximity between individuals need to generate exposure notifications. Instead only predictions of possible exposure to the coronavirus should lead to alerts. This can be accomplished using link prediction and node classification tasks in graphs.

Link Prediction is the task of finding missing links in a graph. Given a property graph where nodes are people, and their physical contacts are links, Graph Neural Network (GNN) models can be trained to predict additional *exposure links*. These links can happen even when there is no recorded physical proximity event (because of apps being switched off or people not carrying the phone on them during

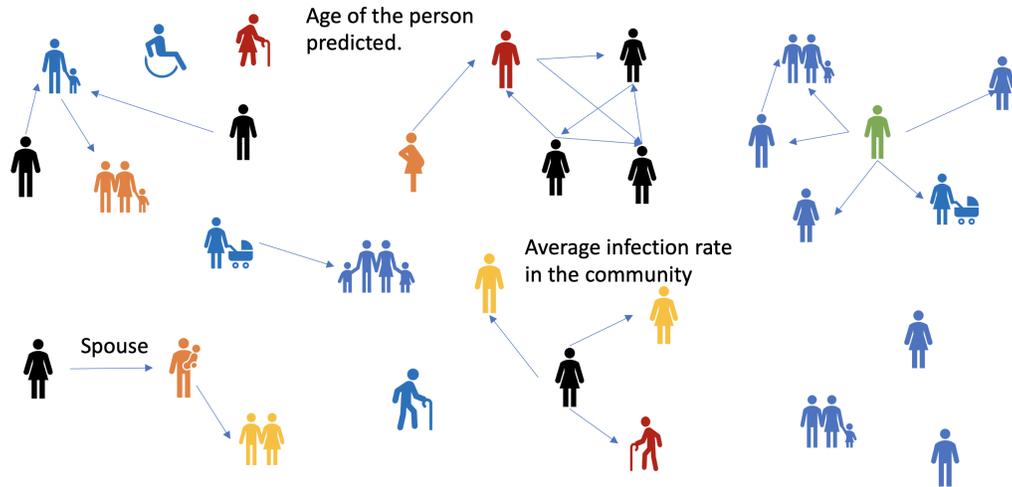


Figure 1: Contact tracing of COVID19 positive persons

a chat in office and other similar potential exposure events). If we are able to predict exposure links well, classifying whether a node is exposed, is a relatively easier problem. So in this work, we focus only on link prediction. The other related task in graphs, namely entity resolution has applications in physical contact tracing and the solutions proposed here are applicable for entity resolution as well.

Explaining the neural model predictions is naturally very important to build trust in the digital contact tracing apps. But making the explanations more human understandable is particularly important in applications aimed at the general population. We propose an improvement to the state of the art Anchors solution of Ribeiro et al. [2018], and also introduce a new path ranking based explainability solution. Like in any social network, the contact tracing apps are only as good as the number of people participating in the network, and especially willing to share information with the network. We propose Graphsheets, based on Factsheets by Arnold et al. [2019], to provide standardized information, to increase trust in the contact tracing applications and the underlying GNN model predictions.

Finally, we draw on the *nudge* idea introduced by Thaler and Sunstein [2009] in behavioural sciences, to encourage users to share relevant information based on explainability techniques and graphsheets.

2 Related Work

Tang [2020] presents a survey of currently available contact tracing applications and the technology choices. MIT Technology Review [2020] maintains a Contract Tracing tracker apps with data on their adoption in respective geographical areas. Mosoff et al. [2020] also maintain a similar list of apps.

Hamilton et al. [2017] introduced GraphSAGE (SAmple and AggreGatE) an inductive framework that leverages node feature information (ex: text attributes, node degrees) to efficiently generate node embeddings for previously unseen data or entirely new (sub)graphs. In this inductive framework, we learn a function that generates embeddings by sampling and aggregating features from a node’s local neighborhood. You et al. [2019] proposed a Position Aware Graph Neural Network that significantly improves performance on the Link Prediction task over the Graph Convolutional Networks.

Much of the recent work on explanations are based on post-hoc models that try to approximate the prediction of complex models using interpretable models. Vannur et al. [2020] present post-hoc explanations of the links predicted by a Graph Neural Network by treating it as a classification problem. They present explanations using LIME (Ribeiro et al. [2016]) and SHAP (Lundberg and Lee [2017]). Arya et al. [2020] introduced the AIX360 toolkit which has a number of explainability solutions that can be used for post-hoc explanation of graph models, if they can posed as approximated as interpretable models. Agarwal et al. [2020] introduced Neural Additive Models which learn a model for each feature to increase the interpretability.

3 Link Prediction for Contact Tracing

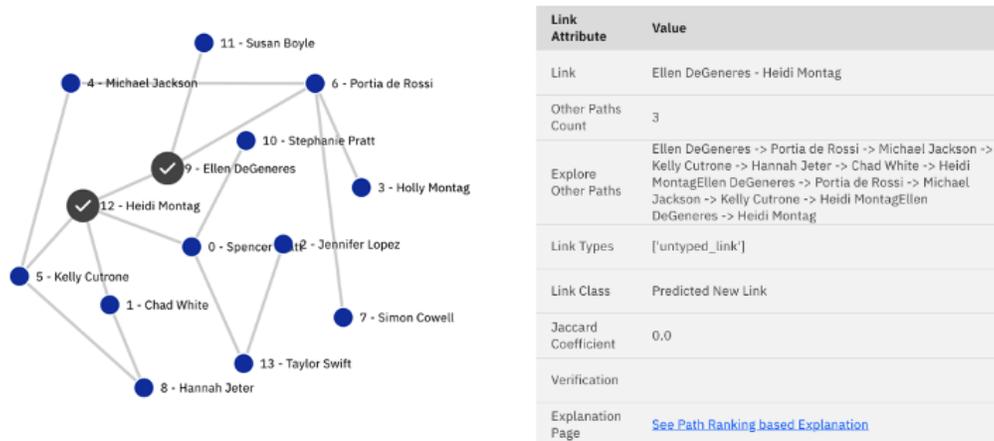


Figure 2: Example Graph which includes links predicted by a GNN model

We conducted our experiments on the UDBMS data Lu et al. [2016] as a proxy for people in the real world. Each node in our graph has attributes similar to the personal data collected by the various Contact Tracing apps. A sample connected component is as shown in the Figure 2. In this particular example of the dataset, Ellen DeGeneres and Heidi Montag, do not seem to have any common neighbors or any shared node attributes, though they have paths involving other nodes. The challenge here is to explain links like these to the users of contact tracing applications.

We begin by explaining our experimental setup. We use the framework provided by You et al. [2019], which in turn uses pytorch Paszke et al. [2019] and more specifically pytorch geometric Fey and Lenssen [2019]. One of the pre-processing steps we do is finding the all pairs shortest paths calculation using appropriate approximations. This pre-processing step comes in handy to explain the links as well.

Following the procedure in You et al. [2019], we choose only connected components with atleast 10 nodes for our experiments. A positive sample is created by randomly choosing 10% of the links. For the negative sample, we use one of the nodes involved in the positive samples and pick a random unconnected node as the other node. The number of negative samples is same as that of the positive samples. We discuss hard negative samples in our future work section. Our batch size is typically 8 subgraphs and for PGNN, we use 64 anchor nodes.

Dataset	Model	ROC AUC	Std. Dev.
UDBMS	GCN	0.4689	0.0280
	P-GNN	0.6456	0.0185

Table 1: Comparison of Link Prediction performance with different GNNs

As shown on Table 1, the P-GNN model seems to perform better on this dataset. In the next section, we'll explore ways to explain link prediction for the example shown in Figure 2.

4 Explanations

One of the commonly used methods for explainability is LIME Ribeiro et al. [2016]. Anchors Ribeiro et al. [2018] is an improvement over LIME which works by focusing only on the important features. The Link Prediction problem can perhaps be posed as a simple classification problem of predicting whether a person is exposed. A typical explanation for Link Prediction is as shown in Figure 3. However, such interpretability solutions are rarely satisfactory to end users. Hence we propose an improvement to Anchors by incorporating ideas from GNN Explainer Ying et al. [2019].



Figure 3: Explanations for Contact Tracing

4.1 Graph Anchors for Link Explanation

We train a simple classifier to predict if a source and target nodes are linked or otherwise. This classifier is trained based on the output of the GNN model described in Section 3. Compared to the GNN model, this post hoc classification model is considered to be more interpretable.

In the above explanation, the classifier model is predicting a link between Ellen DeGeneres and Heidi Montag just like the original GNN model. Anchors model then explains that these two nodes will have a link 52.3% of the time, if the conditions shown in Figure 3 hold. This result suffers from loss of information between the GNN model and the post-hoc interpretable model. Our classifier above is rudimentary and with more feature engineering, we could make the classifier model closer to the GNN model. However, even with reduced accuracy, providing a summary of the conditions under which a model predicts a link is more intuitive than the more complex graph explainability solutions.

Hence we tried generating Anchors explanation on the output of the GNN Explainer model, rather than the post-hoc classification model. The model consumed by the Anchors now consists of a subset of original features, and the target is a binary variable based on the output of the original GNN model. This simple modification to the input of the Anchors explainability solution seems to work reasonably but is still unsatisfactory from the end user point of view, as shown in Figure 4. GNN-Explainer and Anchors both use the broad idea of removing unnecessary features and using only the important features for explaining the links.

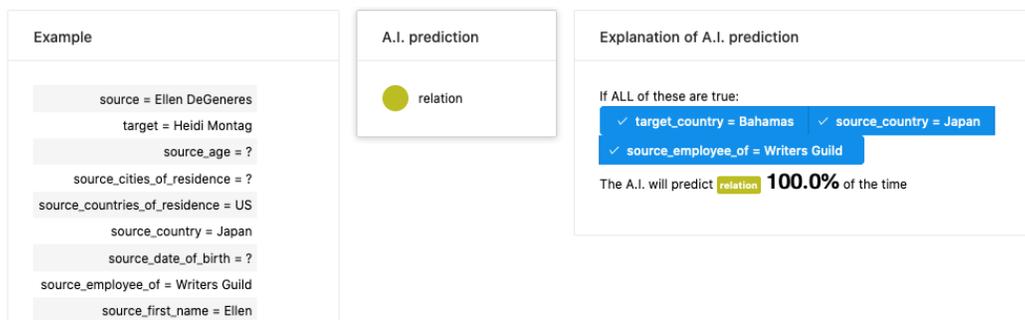


Figure 4: Anchors Explanations using GNN-Explainer output

4.2 Path Ranking based Link Explanation

Our next explainability solution is inspired by ideas in Error detection in Knowledge Graphs. In particular, we use the PaTyBRED approach described in Melo and Paulheim [2017] and Melo and Paulheim [2020].

We propose using the ranking function in the PaTyBRED algorithm to explain the link predicted by GNNs. The algorithm works by ranking already existing paths between two nodes. The idea here is

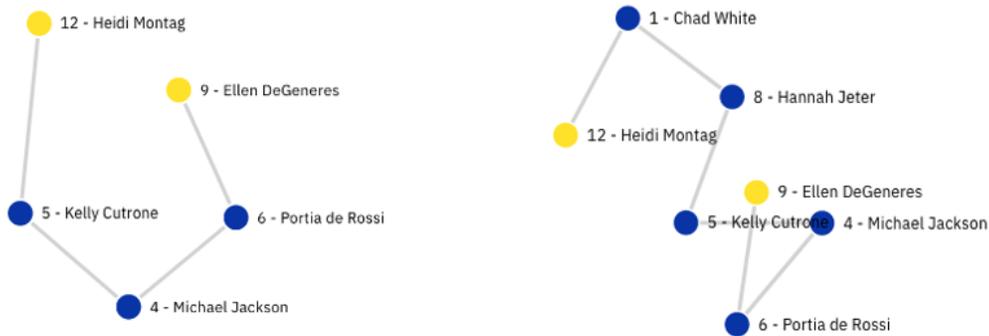


Figure 5: Path Ranking Algorithm based explanation

that the information contained in that path is more useful than other paths to understand the predicted link. This idea is somewhat similar to the GNN Explainer idea of exploring the subgraph around the predicted link, except that we prefer to use an independent algorithm to rank all the paths rather than subgraph around the two nodes. As shown on Figure 5, showing paths between two nodes that are predicted to be linked, can be a substitute for what would otherwise be unsatisfactory explanations using features and common neighbors.

5 Graphsheets

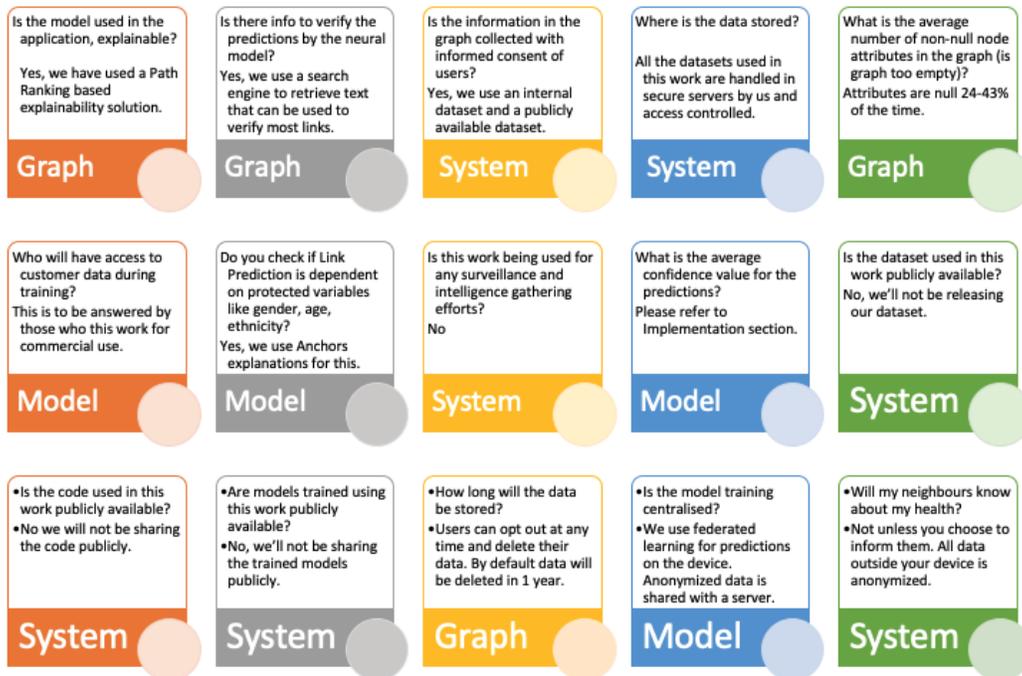


Figure 6: Graphsheet for Contact Tracing

After having trained models for contact tracing use-case, it is important to ensure that downstream applications and model builders follow similar practices for fairness, privacy, data protection and ethical reasons. Towards this goal, we introduce Graph Sheets. Based on Factsheets Arnold et al.

[2019], Graph Sheets include facts about the graph being studied and a number of FAQ types questions and answers that model developers have to consider before training Link Prediction models on their own datasets.

As shown in Figure 6, we provide some sample information regarding the model, graph and the systems used in training Graph Neural Models on the dataset. Other relevant information can be added based on the application.

6 Nudging

Having created a Graph Sheet and being able to generate human understandable explanations for the predicted links, we can try to nudge users to both use contact tracing applications and to share their personal data with the community. Thaler and Sunstein [2009] proposed *nudging* to alter the choice architectures to make people behave in ways that are assumed to be good for them (without making other choices difficult or costly). In privacy preserving contact tracing, there is an inherent tension between the societal good for people to share personal information and the privacy risk involved with sharing such information. Prainsack [2020] discusses the broader implications of nudging on public healthcare data. For the limited purpose of contact tracing where each participant is likely to benefit from sharing personal information, nudging them to share such information could be beneficial.

As shown in Figure 7b, insights drawn from both explainability and graph sheets can be used for nudging. Email Id, is not a useful feature for contact tracing and hence is better not collected, or made clear to users that it's optional and used only for administrative purposes, as the case may be.

Age on the other hand, is a significant factor in assessing risk for COVID19 or other health factors. Models could be trained with age as a feature or alternatively, the age could be used to tailor alerts while being stored only on the user's device. For example, if the model predicts an exposure link with less confidence, the app could make a decision to alert the user or otherwise based on the risk profile of the user.

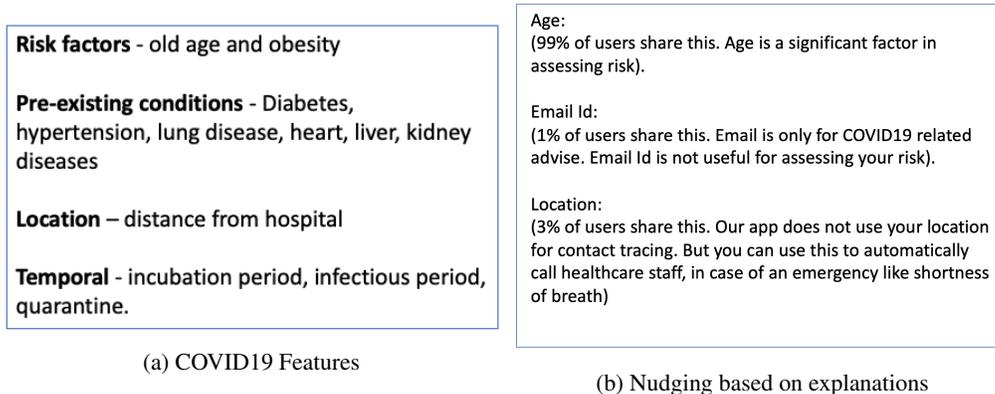


Figure 7: Users of contact tracing apps can be encouraged to share more information by explaining why the information is relevant

Conclusion

We described a Graph Neural Network model to predict exposure links in a contact tracing application to reduce false positives (unnecessary alerts) and also false negatives (lack of updates by participants). We proposed three methods to encourage people to both install and share information with the community, and optionally a centralized authority, by increasing trust in the applications. The three methods are namely, graph sheets to increase trust, human understandable explanations for the exposure notifications, and nudging based on important features found by explainability techniques.

References

- Matthew Abueg, Robert Hinch, Neo Wu, Luyang Liu, William JM Probert, Austin Wu, Paul Eastham, Yusef Shafi, Matt Rosencrantz, Michael Dikovsky, et al. Modeling the combined effect of digital exposure notification and non-pharmaceutical interventions on the covid-19 epidemic in washington state. *medRxiv*, 2020.
- Rishabh Agarwal, Nicholas Frosst, Xuezhou Zhang, Rich Caruana, and Geoffrey E Hinton. Neural additive models: Interpretable machine learning with neural nets. *arXiv preprint arXiv:2004.13912*, 2020.
- Matthew Arnold, Rachel KE Bellamy, Michael Hind, Stephanie Houde, Sameep Mehta, A Mojsilović, Ravi Nair, K Natesan Ramamurthy, Alexandra Olteanu, David Piorkowski, et al. Factsheets: Increasing trust in ai services through supplier’s declarations of conformity. *IBM Journal of Research and Development*, 63(4/5):6–1, 2019.
- Vijay Arya, Rachel KE Bellamy, Pin-Yu Chen, Amit Dhurandhar, Michael Hind, Samuel C Hoffman, Stephanie Houde, Q Vera Liao, Ronny Luss, Aleksandra Mojsilović, et al. Ai explainability 360: hands-on tutorial. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pages 696–696, 2020.
- Yoshua Bengio. A covid-19 exposure notification app won’t work if people aren’t motivated to use it: Ai expert. <https://montrealgazette.com/health/a-contact-tracing-app-wont-work-if-people-arent-motivated-to-use-it-ai-expert>, 2020. [Online; accessed 11-Oct-2020].
- Yoshua Bengio, Richard Janda, Yun William Yu, Daphne Ippolito, Max Jarvie, Dan Pilat, Brooke Struck, Sekoul Krastev, and Abhinav Sharma. The need for privacy with public digital contact tracing during the covid-19 pandemic. *The Lancet Digital Health*, 2020.
- Matthias Fey and Jan Eric Lenssen. Fast graph representation learning with pytorch geometric. *arXiv preprint arXiv:1903.02428*, 2019.
- Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. In *Advances in neural information processing systems*, pages 1024–1034, 2017.
- Jiaheng Lu, Jun Chen, and Chao Zhang. Helsinki Multi-Model Data Repository. <http://udbms.cs.helsinki.fi/?dataset>, 2016.
- Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. In *Advances in neural information processing systems*, pages 4765–4774, 2017.
- André Melo and Heiko Paulheim. Detection of relation assertion errors in knowledge graphs. In *Proceedings of the Knowledge Capture Conference*, pages 1–8, 2017.
- Andre Melo and Heiko Paulheim. Automatic detection of relation assertion errors and induction of relation constraints. *Semantic Web*, 11(5):801–830, 2020.
- MIT COVID Tracing Tracker [2020]. MIT Technology Review COVID Tracing Tracker Data. https://docs.google.com/spreadsheets/d/1ATa1AS08KtZMx_zJREo0vFh0nmB-sAqJ1-CjVRSC0w/edit#gid=0, 2020. [Online, accessed 12-Oct-2020].
- MIT Technology Review [2020]. MIT Technology Review COVID Tracing Tracker. <https://www.technologyreview.com/2020/05/07/1000961/launching-mittr-covid-tracing-tracker/>, 2020. [Online, accessed 14-May-2020].
- Ryan Mosoff, Tommy Friedlich, Teresa Scassa, Kelly Bronson, and Jason Millar. Global pandemic app watch (gpaw): Covid-19 exposure notification and contact tracing apps. <https://craiedl.ca/gpaw/>, 2020. [Online; accessed 11 October 2020].
- Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. In *Advances in neural information processing systems*, pages 8026–8037, 2019.

- Barbara Prainsack. The value of healthcare data: to nudge, or not? *Policy Studies*, pages 1–16, 2020.
- Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. " why should i trust you?" explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1135–1144, 2016.
- Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. Anchors: High-precision model-agnostic explanations. In *AAAI*, volume 18, pages 1527–1535, 2018.
- Qiang Tang. Privacy-preserving contact tracing: current solutions and open questions. *arXiv preprint arXiv:2004.06818*, 2020.
- Richard H Thaler and Cass R Sunstein. *Nudge: Improving decisions about health, wealth, and happiness*. Penguin, 2009.
- Lingraj S Vannur, Balaji Ganesan, Lokesh Nagalapatti, Hima Patel, and MN Thippeswamy. Data augmentation for personal knowledge base population, 2020.
- Rex Ying, Dylan Bourgeois, Jiaxuan You, Marinka Zitnik, and Jure Leskovec. Gnn explainer: A tool for post-hoc explanation of graph neural networks. *arXiv preprint arXiv:1903.03894*, 2019.
- Jiaxuan You, Rex Ying, and Jure Leskovec. Position-aware graph neural networks. *arXiv preprint arXiv:1906.04817*, 2019.