Pagerank Fairness in Networks

Evaggelia Pitoura[1,2]

[1] Computer Science and Engineering Department, University of Ioannina, Greece
[2] Archimedes Unit, ATHENA RC, Greece

Abstract
We live in a connected world where networks be it, social, communication, interaction, or collaboration networks, play an important role. In this paper, we focus on how power is distributed among the nodes of such networks. Specifically, we look at the structural importance of the nodes in a network as measured by the pagerank algorithm. Given two groups of nodes, we say that a network is fair, if the nodes of the two groups hold equally central positions in the network. First, we investigate whether real networks are fair and the conditions that may cause unfairness. Then, we present modifications of the pagerank algorithm such that the computed node importance is both fair and also as accurate as possible. Finally, we propose recommending connections whose additions in the network will improve fairness.

Keywords
Pagerank, node centrality, graph fairness, algorithmic fairness

1. Introduction

Networks offer a generic data structure for representing entities and the relationships and interactions between them. For example, in a social network, nodes correspond to people and edges to connections between them. In this paper, we look into network fairness. We adopt a group-based approach [1]: we assume that nodes belong to groups based on the value of one of their sensitive attributes, e.g., based on their gender, or race. We study fairness with respect to node centrality, i.e., with respect to whether nodes belonging to different groups hold equally central positions in the network.

One measure of node centrality is the degree of the node, i.e, the number of its neighbors. For example, in a social network degree centrality considers the number of followers. A more elaborate measure of centrality is provided by the celebrated pagerank algorithm (PR). In defining the centrality of a node, PR takes into account the PR centrality of its neighbors [2]. For example, in a social network, the PR centrality of a node considers not only how many followers the node has but also the PR centrality of these followers. Previous research has found that as networks evolve, biases arise in the degree centrality of nodes belonging to different groups [3, 4]. In this short article, we provide a high-level overview of our research on pagerank-based centrality fairness in networks [5, 6].

The pagerank algorithm assigns a weight $\mathrm{P}(u)$ to each node u that indicates the significance of u in the network, with the sum of the weights assigned to all nodes being equal to 1. We

EWAF'23: European Workshop on Algorithmic Fairness, June 07–09, 2023, Winterthur, Switzerland
pitoura@cs.uoi.gr (E. Pitoura)
0000-0002-3775-4995 (E. Pitoura)

assume two groups, the protected group, that we call the red group R and the privileged group, that we call the blue group B. We will use $\mathrm{P}(R)$ and $\mathrm{P}(B)$ to denote the total weight that PR assigns to the nodes in the red and blue group respectively. Clearly, $\mathrm{P}(B) = 1 - \mathrm{P}(R)$. For simplicity, we call $\mathrm{P}(R)$ red pagerank and $\mathrm{P}(B)$ blue pagerank.

We say that there is ϕ *(pagerank) fairness*, if the red pagerank is equal to ϕ, i.e., $\mathrm{P}(R) = \phi$, where ϕ is a parameter whose value depends on the fairness policy. For example, by setting $\phi = 0.5$, we ask that both groups are equally important. As another example, let r be he fraction of nodes that belong to R, that is, $|R| = r \times N$, where N is the total number of nodes. By setting $\phi = r$, we ask that the red nodes have a share in the weights proportional to their share in the population, a property also known as *demographic parity* [1].

Besides pagerank, we also consider personalized pagerank, where each node i assigns a weight $\mathrm{P}_i(u)$ to each node u in the network. $\mathrm{P}_i(u)$ indicates the significance that node u has for node i and can be seen as a measure of proximity between source node i and node u. Again, we use $\mathrm{P}_i(R)$ and $\mathrm{P}_i(B)$ to denote the weight that node i allocates to the red and blue groups respectively and ask that $\mathrm{P}_i(R)$ is equal to ϕ. Intuitively, fairness of P_i implies that node i weights the red and blue groups fairly. Overall, we can think of $\mathrm{P}_i(R)$ as a measure of how a specific node i weights the red group, while $\mathrm{P}(R)$ captures the weight that the network as a whole places on the red group.

In the following, we address briefly the following research questions:

RQ1 Are networks fair and what are the conditions that may cause unfairness?

RQ2 Can we modify the pagerank algorithm to produce weights that are both fair and as accurate as possible?

RQ3 Can we intervene in the evolution of a network through link recommendations so as to improve fairness?

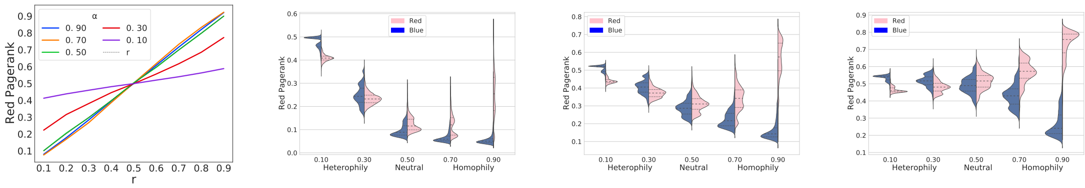
2. Towards fair networks

2.1. Are networks fair?

Previous research has shown that size imbalance between groups, and homophily (here the tendency to connect with nodes of the same color) may lead to degree unfairness [3, 4, 7]. Does unfairness arise also in the case of pagerank?

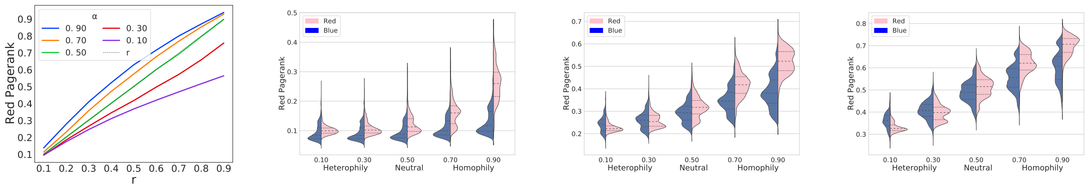
To address this question, we generate synthetic networks using the biased preferential attachment model [3] and study the conditions under which demographic parity holds. The ratio r of the red nodes controls the size imbalance between the two groups. Parameter α controls the level of homophily: $\alpha < 0.5$ corresponds to heterophily, $\alpha = 0.5$ to neutrality and $\alpha > 0.5$ to homophily. We consider: (a) a symmetric case, where α is the same for both groups and (b) an asymmetric case, where we set $\alpha = 0.5$ for the blue group, making it neutral, and vary the level of homophily only for the red group,

We plot in Figures 1(a) and 2(a) the red pagerank. *Note that demographic parity corresponds to the identity line, i.e., P*$(R) = r$. Values above the identity line indicate unfairness towards the blue group, while values below the line unfairness towards the red group. We also plot the red personalized PR in Figures 1(b)-(d) and 2(b)-(d). We plot two distributions, one for the $P_i(R)$ of the red nodes (i.e., $i \in R$) and one for $P_i(R)$ of the blue nodes ($i \in B$). Distributions are



(a) P(R) (b) r = 0.1 (c) r = 0.3 (d) r = 0.5

Figure 1: Symmetry: (a) Red PR, (b)-(d) Red personalized PR distribution for the red and blue nodes.



(a) P(R) (b) r = 0.1 (c) r = 0.3 (d) r = 0.5

Figure 2: Asymmetry: (a) Red PR, (b)-(d) Red personalized PR distribution for the red and blue nodes.

plotted in the form of violin plots. *Demographic parity corresponds to the case in which the two distributions overlap, with their mean on value r.*

Symmetric case. In case of homophily, (α = 0.7, 0.9), nodes tend to form two clusters, one with red nodes and one with blue nodes sparsely connected to each other. This leads to almost pagerank fair networks (with a very slight unfairness towards the smaller group), but the nodes are personalized pagerank unfair. In case of heterophily (α = 0.1, 0.3), there are no clusters, nodes tend to connect with nodes of the opposite color, and the larger group favors the smaller one. There is pagerank unfairness towards the larger group, and both blue and red nodes are personalized pagerank unfair towards the larger group. This is especially evident when there is large size imbalance (small r).

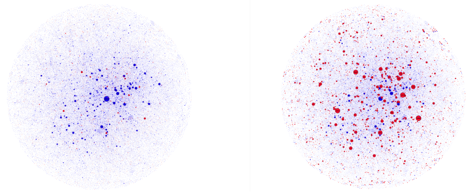
Asymmetric case. When the red nodes are homophilic (α = 0.7, 0.9), the red group keeps the pagerank to itself. As a result there is both pagerank and personalized pagerank unfairness towards the blue (that is, the neutral) group. When the red nodes are heterophilic (α = 0.1, 0.3), the red group favors the blue group, and as a result, there is both pagerank and personalized pagerank unfairness towards the red (that is, the heterophilic) group.

The only case when there is both pagerank and personalized pagerank fairness is when both groups are neutral (α = 0.5 for both groups) and of similar sizes (r = 0.5) (middle violin plots in Figures 1(d), and 2(d)).

2.2. Fairness-aware Pagerank

We would like to modify the pagerank algorithm so that the produced weights are fair independently of whether the network is fair or not. Furthermore, we would like these weights to be as accurate as possible, that is, as close as possible to the weights assigned by the original pagerank algorithm. It was shown that the only way to produce both global and personalized fair weights is by the family of *locally fair PR algorithms* [5]. The original PR is based on a vanilla random walk. When at node u, the walk moves to any of the neighbors of u with equal probability. In a locally fair PR, when at node u, the random walk moves with probability ϕ to a red node and with probability 1- ϕ to a blue node.

In the simplest form of the locally fair PR algorithm, called *neighbor-based* locally fair PR, when at node u, the walk moves with probability ϕ to a red neighbor of u, and with probability $1 - \phi$ to a blue neighbor of u [6, 8, 9]. In Figure 3, we see the weights of the original PR and of the neighbor-based locally fair PR that results in increasing the weights of the red group. The *residual-based locally fair* PR algorithm generalizes this idea. When at a node u that has fewer red neighbors than ϕ, the walk moves with equal probability to any of the neighbors of u, and with the remaining residual probability to a red node appropriately selected from R so as to optimize accuracy.



(a) (b)

Figure 3: Visualization of a subset of the DBLP co-authorship network: (a) original PR, (b) neighbor-based locally fair PR. Blue nodes correspond to male and red nodes to female authors; the size of a node is proportional to its weight; $\phi = 0.5$.

2.3. Link recommendations for fair networks

Instead of modifying the PR algorithm to produce fair weights, a more effective intervention would be to "correct" the network so that the original PR algorithm produces fair weights. The growth of many real-world networks is affected by link recommendations [10], for example recommendations of people to follow in a social network. The goal is to recommend links such that if accepted the fairness of the network will improve. To this end, we have derived analytical formulas for estimating the effect of edge additions on PR fairness and we have used them to design linear time link recommendation algorithms for maximizing fairness [6].

It was shown that the most important edges in terms of fairness are edges that connect nodes whose neighborhoods are of a "different color". It was also shown that in general the characteristics of the target node of the recommended link have a stronger correlation with fairness than the characteristics of the source node. Among these characteristics, the most relevant ones are the group (i.e., color) and the personalized PR of the target node.

3. Conclusions

In this short paper, we have summarized our previous work on PR centrality fairness. An important research question is to study how centrality unfairness is reflected on the fairness of downstream ML tasks. For example, is pagerank unfairness reflected in graph embeddings or in GNN models? Along this line, the effect of unfairness on network processes such as diffusion [11], and phenomena such as filter bubbles [12] is also of interest.

It is also important to view fairness in networks from an interdisciplinary perspective. Understanding the confounding social and psychological factors that lead to network unfairness would help in designing more relevant link recommendation algorithms. Furthermore, studying both the short term and long term effect that network unfairness may have in the lives of individuals would be instrumental [13]. Finally, questions of ethics arise in terms of how we design fairness interventions in networks.

Acknowledgments

The author was supported by the “ARCHIMEDES Unit: Research in Artificial Intelligence, Data Science and Algorithms” OPS code 5154714 within the framework of the National Recovery and Resilience Plan “Greece 2.0” with funding from the European Union – NextGenerationEU.

References

[1] C. Dwork, M. Hardt, T. Pitassi, O. Reingold, R. S. Zemel, Fairness through awareness, in: Innovations in Theoretical Computer Science, 2012, pp. 214–226.
[2] D. F. Gleich, Pagerank beyond the web, SIAM Review 57 (2015) 321–363.
[3] C. Avin, B. Keller, Z. Lotker, C. Mathieu, D. Peleg, Y. A. Pignolet, Homophily and the glass ceiling effect in social networks, in: ITCS, 2015, pp. 41–50.
[4] A. Stoica, C. J. Riederer, A. Chaintreau, Algorithmic glass ceiling in social networks: The effects of social recommendations on network diversity, in: WWW, ACM, 2018, pp. 923–932.
[5] S. Tsioutsiouliklis, E. Pitoura, P. Tsaparas, I. Kleftakis, N. Mamoulis, Fairness-aware pagerank, in: WWW, ACM, 2021, pp. 3815–3826.
[6] S. Tsioutsiouliklis, E. Pitoura, K. Semertzidis, P. Tsaparas, Link recommendations for pagerank fairness, in: WWW, ACM, 2022, pp. 3541–3551.
[7] F. Fabbri, F. Bonchi, L. Boratto, C. Castillo, The effect of homophily on disparate visibility of minorities in people recommender systems, in: ICWSM, 2020.
[8] T. A. Rahman, B. Surma, M. Backes, Y. Zhang, Fairwalk: Towards fair graph embedding, in: IJCAI, 2019, pp. 3289–3295.
[9] A. Khajehnejad, M. Khajehnejad, M. Babaei, K. P. Gummadi, A. Weller, B. Mirzasoleiman, Crosswalk: Fairness-enhanced node representation learning, in: AAAI, 2022, pp. 11963–11970.
[10] D. Liben-Nowell, J. M. Kleinberg, The link prediction problem for social networks, in: CIKM, ACM, 2003, pp. 556–559.
[11] J. Ali, M. Babaei, A. Chakraborty, B. Mirzasoleiman, K. P. Gummadi, A. Singla, On the fairness of time-critical influence maximization in social networks, in: ICDE, IEEE, 2022, pp. 1541–1542.
[12] S. Haddadan, C. Menghini, M. Riondato, E. Upfal, Repbublik: Reducing polarized bubble radius with link insertions, in: WSDM, ACM, 2021, pp. 139–147.
[13] E. Pitoura, Social-minded measures of data quality: Fairness, diversity, and lack of bias, ACM J. Data Inf. Qual. 12 (2020) 12:1–12:8.