Towards intersectional fairness in community detection

Georgios Panayiotou [1*], Matteo Magnani [1], Ece Calikus [1]

[1] InfoLab, Department of Information Technology, Uppsala University, Uppsala, Sweden
[*] georgios.panayiotou@it.uu.se

Abstract

Despite the recent interest in fairness-aware community detection, existing methods only focus on fairness along a single demographic, failing to account for multiple demographic attributes and their intersections. This work investigates intersectional fairness in social network community detection, highlighting the impact of demographic distribution and algorithm choice on fairness outcomes. Our findings emphasize the need for community detection methods considering intersectionality and demographic proportionality in order to mitigate biases in social network analysis.

Code — https://github.com/uuinfolab/paper.25_ICWSM_Intersectional-Fairness-CD

Introduction

With the increasing interest in large-scale complex network analysis, algorithmic fairness in social network analysis has emerged as a nascent area of research. This is a direct result of recent work highlighting usage of fairness-oblivious social network analysis methods can lead to bias and power inequality issues, typically stemming from structural biases in the network (Avin et al. 2017; Masrour et al. 2020). Thus, the overall goal of the field is the design of methods that consider network inequalities, where the outcome should not be biased towards any specific demographic (Saxena, Fletcher, and Pechenizkiy 2024).

An emerging area where algorithmic fairness can have significant societal impact is community detection, particularly in large-scale systems such as online social media. In platforms such as Facebook and X, community detection is used to group similar users and offer tailored recommendations (Meta 2023; Twitter 2023). However, most current algorithms rely solely on sub-graph density, which can unintentionally create echo chambers, such as grouping users with similar political views, reducing exposure to diverse perspectives essential for a healthy democracy (Pariser 2012). To address this, we need community detection methods that balance user relevance with fair demographic representation to avoid biased outcomes.

While there have been recent efforts to incorporate fairness constraints into community detection, the field still remains underexplored compared to its non-graph counterpart (Saxena, Fletcher, and Pechenizkiy 2024). Recent works have introduced group fairness constraints for community detection algorithms (Manolis and Pitoura 2023; de Vink and Saxena 2025), and proposed the first fairness-aware community detection methods (Panayiotou and Magnani 2025; Gkartzios, Pitoura, and Tsaparas 2025). However, these methods focus on fairness along a single axis of identity, typically gender or race, failing to capture biases along the intersections of these identities. As the impact of identity intersections across multiple demographics produces unique biases towards each subgroup (Crenshaw 1989), considering intersectionality in community detection is crucial to prevent amplifying existing inequalities within social networks. This need is further highlighted by recent studies highlighting algorithmic biases against intersectional subgroups (Gohar and Cheng 2023; Martin-Gutierez, Cartier van Dissel, and Karimi 2024).

In this short paper, we provide a preliminary exploration of intersectional fairness within social network community detection. Specifically, we examine the fairness of partitions obtained through various community detection methods, centering on the following questions:

1. How balanced, with respect to each demographic, are communities detected by various types of algorithms?
2. How well are combinations of demographics represented in the obtained communities?

Considering the combinations of demographic variables, albeit important, only provides a partial understanding towards implementing the intersectionality framework into quantitative methods (Walby, Armstrong, and Strid 2012). Nevertheless, this experimental comparison provides insights into designing intersectionally fair community detection methods.

Related work

With the growing reliance on automated decision-making systems, algorithmic fairness has emerged as a key study area (Pessach and Shmueli 2023; Caton and Haas 2024). Clustering tasks, in particular, have received substantial attention, with multiple works proposing algorithms for fair clustering of non-graph data (Bera et al. 2019; Chhabra, Masalkovaitė, and Mohapatra 2021; Chierichetti et al. 2017). Recently, fairness definitions considering the intersectionality framework have also been introduced for non-graph applications (Gohar and Cheng 2023; Maheshwari et al. 2023).

Within the context of fair community detection on networks, however, literature remains scarce (Saxena, Fletcher, and Pechenizkiy 2024). Group fairness definitions, along with experimental analyses of community detection methods, have been recently introduced (Manolis and Pitoura 2023; de Vink and Saxena 2025). These works showcase the importance of considering both the community detection algorithm type and sensitive node attributes when incorporating fairness constraints into community detection, providing an important motivation for our work. Furthermore, fairness-aware community detection methods based on modularity have been recently proposed (Panayiotou and Magnani 2025; Gkartzios, Pitoura, and Tsaparas 2025). However, these works consider neither intersectional fairness, nor multiple sensitive attributes.

Demographic fairness

In the present work, we focus on the well-studied fairness definition of group balance, used in the seminal work of (Chierichetti et al. 2017) and later, as the fairness score in the Fair-mod method (Panayiotou and Magnani 2025). We extend this definition to consider multiple groups under each demographic (also referred to as a sensitive attribute) as follows.

Given a network $G = (V, E)$, a disjoint partition P of the nodes in V into communities, a set D with lists of $K_d \geq 2$ group classes under each demographic $d \in D$, and a set H containing $|D|$ sets, each denoting the membership of nodes in V in the groups under each demographic, the group balance of community $C_i \in P$ with respect to a demographic $d \in D$, is given by:

$$balance(C_i, d) = (K_d - 1) \min_{j \in d} \left(\frac{|H_{d,j} \cap V(C_i)|}{|H'_{d,j} \cap V(C_i)|} \right) \in [0, 1],$$

where $V(C_i)$ is the set of vertices in community C_i, $H_{d,j}$ denotes the vertices in group j of demographic d, and H'_j all other vertices not in group d, j. We scale the balance score by $K_d - 1$ to achieve a maximum score of 1 in a perfectly balanced community; note that when $K_d > 2$, the maximum balance score becomes $1/(K_d - 1)$ without scaling.

To further identify the impact of demographic intersections in community balance, we also want to consider the fairness between the combinations of groups from different demographics, a strategy also used in intersectionally fair machine learning applications (Gohar and Cheng 2023). We calculate the balance over a flattened membership set $H_{combined}$ as follows (for $|D| = 2$):

$$H_{combined} = \{\{H_{d_1,g_1} \cap H_{d_2,g_2}\} : d_1 \neq d_2 \in D\},$$

where g_1 and g_2 are groups under demographics d_1 and d_2 respectively.

Facebook, G/E	**A**	**B**	**C**
blue	608	1063	832
red	507	651	378

Twitch, M/L	**EN**	**FR**	**DE**	**Others**
blue	62783	2558	5778	7914
red	61628	4241	3650	19562

Table 1: Population under each demographic group for the Facebook (above) and Twitch Gamers (below) networks. Facebook demographics: gender (G, rows), education (E, columns). Twitch demographics: maturity (M, rows), language (L, columns). The group classes under the gender, education and maturity demographics are anonymized in the datasets. For Twitch, we report the three largest language groups; the other 18 languages in the dataset are summarized under Others.

Experimental setup

To illustrate what aspects should be taken into account when designing fair community detection methods, we want to highlight the impact of various methods of community detection on the fairness outcomes in attributed social networks. To achieve that, we evaluate to what extent the partitions obtained are balanced, with respect to each demographic attribute in the network, as well as the distinct groups from both demographics, using $H_{combined}$ as the membership set.

Settings

We evaluate the balance of partitions obtained using various types of methods for community detection. Namely, we focus on three algorithms; the modularity-based Louvain (Blondel et al. 2008), the random walk-based InfoMap (Rosvall and Bergstrom 2008) and SBM-DL, using the stochastic block model (Peixoto 2014). We use the CDLib (Rossetti, Milli, and Cazabet 2019) implementation of the algorithms under the default settings.

Datasets

As an illustrative example for the evaluation, we use real social networks with node attributes, obtained from the Stanford Large Network Repository [1]. We consider the following (a summary of their demographic distributions can be found in Table 1):

- *Facebook:* The dataset includes participants in a survey, and their Facebook friendships. The nodes contain demographic information about their gender and education.
- *Twitch Gamers:* The network includes Twitch streamers and their mutual following relationships. Node information includes their stream maturity and language.

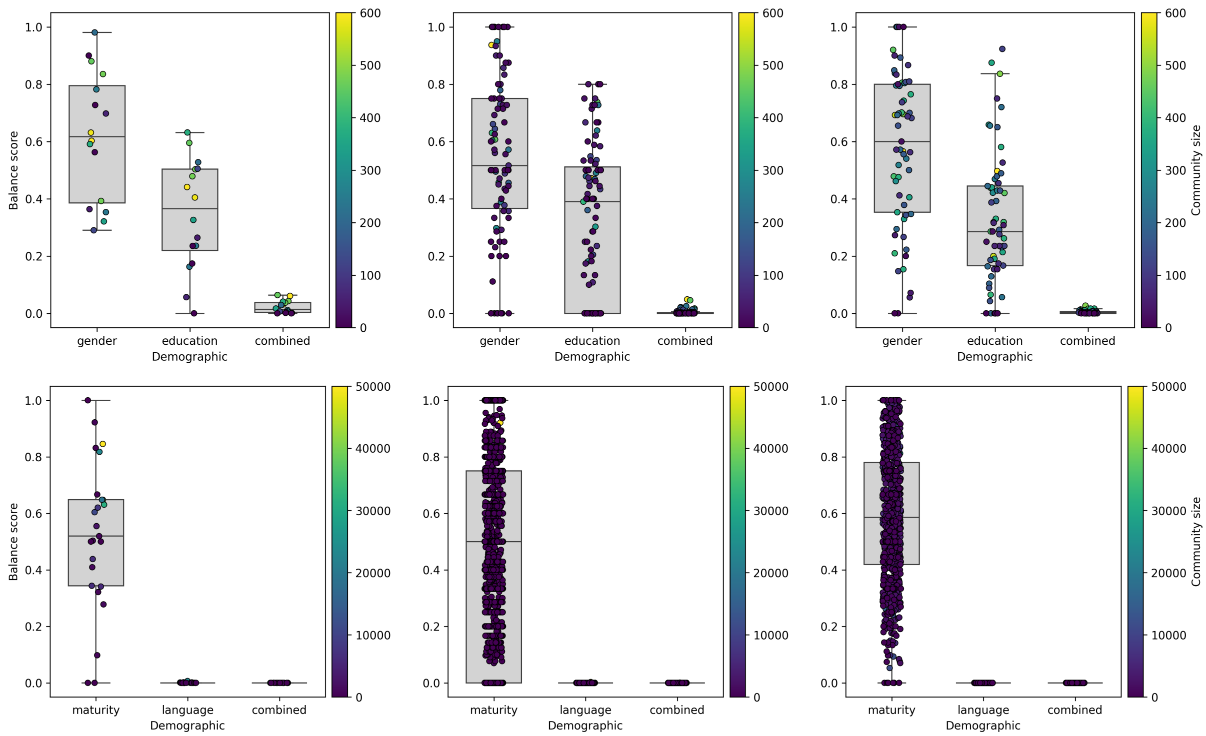

Figure 1: Distribution of community balance scores per demographic, and for all combinations of intersectional groups ($H_{combined}$), for communities detected on the Facebook (top) and Twitch Gamers (bottom) networks, using the Louvain (left), InfoMap (center) and SBM-DL (right) algorithms. Overlay dot colour represents community size.

Results

Looking at the communities detected by the three different algorithms on the Facebook social network, we observe some interesting trends. First, each algorithm discovers a different number of communities with varying sizes. While unsurprising, this has an effect on the distribution of the balance scores. In Fig. 1 above, SBM-DL and InfoMap discover a large number of smaller communities in the same network compared to Louvain, leading to a wider distribution of balance scores with respect to both the gender and education attributes. We note, however, that the largest communities identified by both SBM-DL and InfoMap also receive relatively high balance scores, with respect to both sensitive attributes.

Comparing the median demographic balance scores (Fig. 1, top) to their proportional balance in the Facebook network (calculated as the demographic's balance score on the entire network as a single community), all three algorithms are fairly close to the expectation for the gender attribute (0.613); the same can be observed for the gender balance scores of the largest communities (Fig. 2). In contrast, the median education balance score is much lower than the respective proportional balance score (0.763) for all methods. Overall, while gender balance is already reflected by the identified communities to an extent, the education attribute is not necessarily balanced proportionally in those same communities. This further highlights the need for a community detection method promoting fairness between multiple demographic groups, especially in networks where there is structural bias against one attribute.

However, when each group combination is considered separately, the communities identified receive very low scores in general. In the bottom row of Fig. 2, we observe that community size also impacts the balance score for this strategy. While larger communities receive a slightly higher balance score, smaller communities are unlikely to contain a good balance between all group combinations, especially considering one group is underrepresented in the network (cf. Table 1). This trend suggests that considering balance between all demographic intersections as distinct groups may not be an appropriate strategy for fairness-aware community detection methods, at least not without considering their overall proportionality in the network.

The observed trends are further highlighted when we compare the results to the much larger Twitch Gamers network, where users streaming in English are heavily overrepresented. In Fig. 1 below, we note the trend of modularity-based methods to discover fewer (and larger) communities compared to other optimization methods, leading also to higher maturity balance scores for the largest communities. However, none of the algorithms is able to produce well-balanced communities with respect to the users' language,

[1] https://snap.stanford.edu/data/

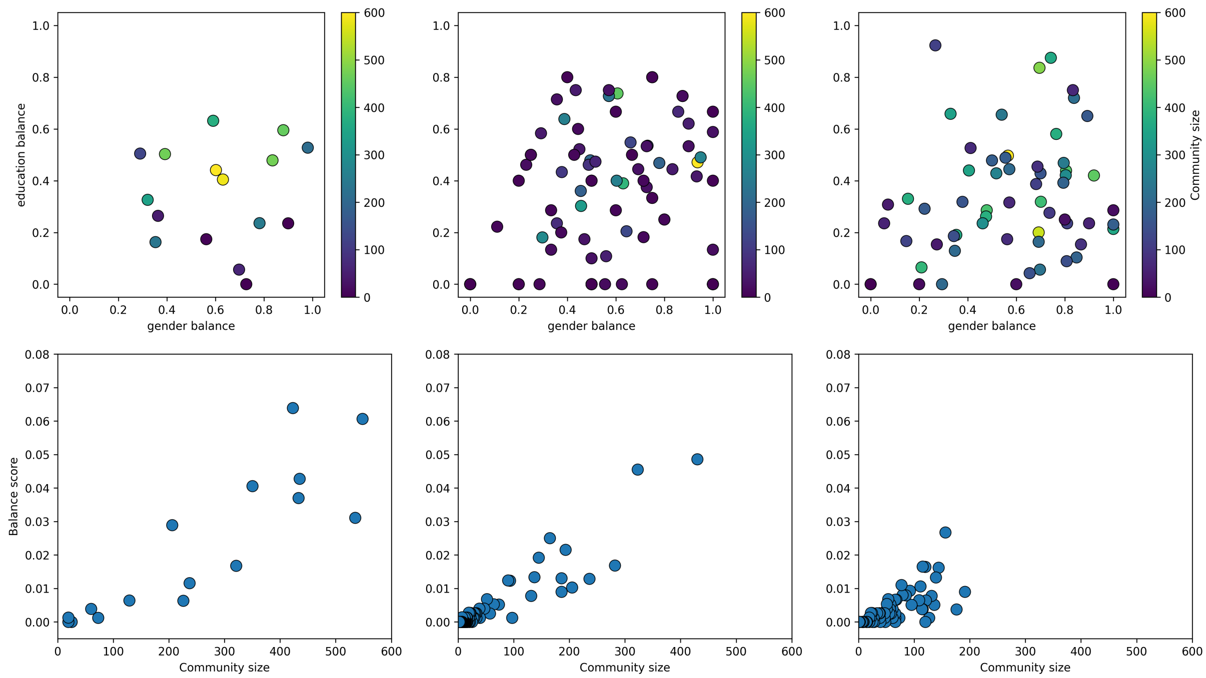

Figure 2: Comparison of balance scores for communities detected on the Facebook network using Louvain (left column), InfoMap (center column), SBM-DL (right column). Top row: comparison of balance scores for the two demographics. The dot colour represents community size. Bottom row: balance scores for $H_{combined}$.

with very few communities receiving a score over 0. The effect is also repeated when all demographic combinations are considered as distinct classes, with only the largest communities detected (for Louvain and InfoMap) yielding a balance score slightly over 0. This is largely due to the English users representing roughly 74% of the network, in great contrast to the other 20 languages in the dataset. These findings confirm the above observations: first, group balance may not be an appropriate fairness function for imbalanced networks, particularly when users are characterized by multiple demographics. Second, there is a need for fairness-aware community detection methods taking into consideration multiple, and disproportionally represented, demographics.

Discussion

Our experiments prompt several interesting questions concerning fair community detection. First, what is a suitable function to quantify fairness in community detection? While group balance has been well-established as a fairness definition, it might not be suitable for all types of networks; especially when considering network data where one group is heavily underrepresented. Alternative fairness definitions can consider how the sensitive attributes are embedded into the network structure, or the demographics' proportionality in the data.

Second, how should multiple demographics be integrated into the fairness definition? Considering all identity intersections as distinct groups can be used as a strategy to extend previously proposed fairness-aware methods for a single demographic. However, this approach is not necessarily suitable for underrepresented subpopulations, nor scalable, as the complexity of such a method would increase alongside the number of groups in the data.

Finally, how does the choice of community detection algorithm impact the fairness of the detected partition? Our experiments, along with previous comparative work on fairness in community detection (de Vink and Saxena 2025), suggest that the type of optimization method impacts the size of detected communities, and as a result, the balance between different demographic groups. This emphasizes the need for an appropriate fairness function taking into account the types of communities each algorithm yields, or alternatively, a method to adjust the importance of network structure over the partition's fairness.

Due to space limitations, we report results on two social networks and a particular definition of fairness, to illustrate the impact of both the fairness constraint and underrepresented demographics in the network in fairness outcomes. Nevertheless, we already note several interesting aspects to be taken into consideration when designing fairness-aware community detection methods.

Our findings highlight the need for more extensive benchmarking of community detection methods. This includes experimentation with more attributed social networks and fairness definitions, and comparisons with fairness-aware community detection methods, to highlight how different algorithms behave under various network structures. Finally, future studies should also consider more complex network structures (e.g. multilayer networks), as fairness-aware community detection methods have not yet been extended to multiple layers (Panayiotou, Magnani, and Pinaud 2024).

Acknowledgments

This work has been partly funded by eSSENCE, an e-Science collaboration funded as a strategic research area of Sweden. The computations were enabled by resources provided by the National Academic Infrastructure for Supercomputing in Sweden (NAISS), partially funded by the Swedish Research Council through grant agreement no. 2022-06725.

References

Avin, C.; Lotker, Z.; Nahum, Y.; and Peleg, D. 2017. Modeling and Analysis of Glass Ceiling and Power Inequality in Bi-populated Societies. In Shmueli, E.; Barzel, B.; and Puzis, R., eds., *3rd International Winter School and Conference on Network Science*, 61–73. Cham: Springer International Publishing. ISBN 978-3-319-55471-6.

Bera, S.; Chakrabarty, D.; Flores, N.; and Negahbani, M. 2019. Fair Algorithms for Clustering. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc.

Blondel, V. D.; Guillaume, J.-L.; Lambiotte, R.; and Lefebvre, E. 2008. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10): P10008.

Caton, S.; and Haas, C. 2024. Fairness in Machine Learning: A Survey. *ACM Comput. Surv.*, 56(7): 166:1–166:38.

Chhabra, A.; Masalkovaitė, K.; and Mohapatra, P. 2021. An Overview of Fairness in Clustering. *IEEE Access*, 9: 130698–130720. Conference Name: IEEE Access.

Chierichetti, F.; Kumar, R.; Lattanzi, S.; and Vassilvitskii, S. 2017. Fair Clustering Through Fairlets. *Proceedings of the 31st Conference on Neural Information Processing Systems*, 5029–5037.

Crenshaw, K. 1989. Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics. *University of Chicago Legal Forum*, 1989(1).

de Vink, E.; and Saxena, A. 2025. Group Fairness Metrics for Community Detection Methods in Social Networks. In Cherifi, H.; Donduran, M.; Rocha, L. M.; Cherifi, C.; and Varol, O., eds., *Complex Networks & Their Applications XIII*, 43–56. Cham: Springer Nature Switzerland. ISBN 978-3-031-82435-7.

Gkartzios, C.; Pitoura, E.; and Tsaparas, P. 2025. Fair Network Communities through Group Modularity. In *Proceedings of the ACM on Web Conference 2025*, 506–517. New York, NY, USA: Association for Computing Machinery. ISBN 979-8-4007-1274-6.

Gohar, U.; and Cheng, L. 2023. A survey on intersectional fairness in machine learning: notions, mitigation, and challenges. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, IJCAI '23, 6619–6627. Macao, P.R.China. ISBN 978-1-956792-03-4.

Maheshwari, G.; Bellet, A.; Denis, P.; and Keller, M. 2023. Fair Without Leveling Down: A New Intersectional Fairness Definition. In Bouamor, H.; Pino, J.; and Bali, K., eds., *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, 9018–9032. Singapore: Association for Computational Linguistics.

Manolis, K.; and Pitoura, E. 2023. Modularity-Based Fairness in Community Detection. In *Proceedings of the International Conference on Advances in Social Networks Analysis and Mining*, 126–130. Kusadasi Turkiye: ACM. ISBN 9798400704093.

Martin-Gutierez, S.; Cartier van Dissel, M. N.; and Karimi, F. 2024. Intersectional inequalities in social networks. ArXiv:2410.21189 [physics].

Masrour, F.; Wilson, T.; Yan, H.; Tan, P.-N.; and Esfahanian, A. 2020. Bursting the Filter Bubble: Fairness-Aware Network Link Prediction. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(01): 841–848. Number: 01.

Meta. 2023. The AI behind unconnected content recommendations on Facebook and Instagram.

Panayiotou, G.; and Magnani, M. 2025. Fair-mod: Fair Modular Community Detection. In Cherifi, H.; Donduran, M.; Rocha, L. M.; Cherifi, C.; and Varol, O., eds., *Complex Networks & Their Applications XIII*, 91–102. Cham: Springer Nature Switzerland. ISBN 978-3-031-82435-7.

Panayiotou, G.; Magnani, M.; and Pinaud, B. 2024. Current challenges in multilayer network engineering. *Applied Network Science*, 9(75): 1–23. Publisher: SpringerOpen.

Pariser, E. 2012. *The Filter Bubble: What the Internet is Hiding from You*. Penguin Books. ISBN 978-0-241-95452-2.

Peixoto, T. P. 2014. Efficient Monte Carlo and greedy heuristic for the inference of stochastic block models. *Physical Review E*, 89(1): 012804. Publisher: American Physical Society.

Pessach, D.; and Shmueli, E. 2023. Algorithmic Fairness. In Rokach, L.; Maimon, O.; and Shmueli, E., eds., *Machine Learning for Data Science Handbook: Data Mining and Knowledge Discovery Handbook*, 867–886. Cham: Springer International Publishing. ISBN 978-3-031-24628-9.

Rossetti, G.; Milli, L.; and Cazabet, R. 2019. CDLIB: a python library to extract, compare and evaluate communities from complex networks. *Applied Network Science*, 4(1): 1–26. Number: 1 Publisher: SpringerOpen.

Rosvall, M.; and Bergstrom, C. T. 2008. Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences*, 105(4): 1118–1123. Publisher: Proceedings of the National Academy of Sciences.

Saxena, A.; Fletcher, G.; and Pechenizkiy, M. 2024. FairSNA: Algorithmic Fairness in Social Network Analysis. *ACM Comput. Surv.*, 56(8): 213:1–213:45.

Twitter. 2023. Twitter's Recommendation Algorithm.

Walby, S.; Armstrong, J.; and Strid, S. 2012. Intersectionality: Multiple Inequalities in Social Theory. *Sociology*, 46(2): 224–240. Publisher: SAGE Publications Ltd.