FairFuse: Interactive Visual Support for Fair Consensus Ranking

Hilson Shrestha* Kathleen Cachel† Mallak Alkhathlan‡ Elke Rundensteiner§ Lane Harrison¶

Worcester Polytechnic Institute

Figure 1: FairFuse provides multiple views supporting fairness-oriented ranking workflows: A) Consensus Generation View, B) Rank Similarity View, C) Attribute / Protected Attribute Legends, D) Group Fairness View, E) Ranking Exploration View. (Right) Illustrating a fairness-oriented ranking workflow enabled by FairFuse.

Abstract

Fair consensus building combines the preferences of multiple rankers into a single consensus ranking, while ensuring any group defined by a protected attribute (such as race or gender) is not disadvantaged compared to other groups. Manually generating a fair consensus ranking is time-consuming and impractical— even for a fairly small number of candidates. While algorithmic approaches for auditing and generating fair consensus rankings have been developed, these have not been operationalized in interactive systems. To bridge this gap, we introduce FairFuse, a visualization system for generating, analyzing, and auditing fair consensus rankings. We construct a data model which includes base rankings entered by rankers, augmented with measures of group fairness, and algorithms for generating consensus rankings with varying degrees of fairness. We design novel visualizations that encode these measures in a parallel-coordinates style rank visualization, with interactions for generating and exploring fair consensus rankings. We describe use cases in which FairFuse supports a decision-maker in ranking scenarios in which fairness is important, and discuss emerging challenges for future efforts supporting fairness-oriented rank analysis. Code and demo videos available at `https://osf.io/hd639/`.

Index Terms: Human-centered computing—Visualization—Visualization systems and tools

*e-mail:hshrestha@wpi.edu
†e-mail:kcachel@wpi.edu
‡e-mail:malkhathlan@wpi.edu
§e-mail:rundenst@wpi.edu
¶e-mail:ltharrison@wpi.edu

1 Introduction

The ubiquitous task of combining preferences by multiple stakeholders into a consensus is challenging for decision-makers that steer this process. Decision-makers often grapple with diverging preferences provided by different stakeholders, and must reach a single decision that all stakeholders accept and agree with. A frequent approach to such decision-making is to employ rankings, where each stakeholder provides their ranking over the candidates. Candidates might include lists of people, organizations, or other entities. Decision-makers combine these base rankings from individual stakeholders into a single consensus ranking as part of the process.

However, when ranking candidates, stakeholders may provide biased or unfair rankings [13]. Bias can be implicit (unintended), for example, when favoring candidates from a particular university who happen to be overwhelmingly white. Bias can also be explicit, for example, weighing women candidates lower due to a perceived lack of ability for the target role. One way to mitigate such unfair outcomes is by promoting measures from the algorithmic fairness community, such as *group fairness* or *statistical parity* [38]. Statistical parity, for example, is a requirement that all groups receive an equal proportion of the positive outcome; in our case, favorable positions in the consensus ranking. Without intervention in the ranking process, there is substantial risk of perpetuating unfair practices, and thus harming marginalized groups.

Unfortunately, constructing a consensus ranking is challenging [6, 16] and ensuring that this consensus ranking is fair is even more difficult [11, 28]. Numerous visualization tools have explored the design space of rankings [18] and rank-based decision making [22, 34]. But existing approaches have not dealt with the complications of incorporating fairness into visual encodings, nor with interactive workflows related to consensus rank generation. Similarly, while research in fair algorithms has developed rank-focused auditing metrics and fair rank aggregation methods [11, 28], they have been confined to (non-visual) algorithmic solutions requiring substantial

2022 IEEE Visualization and Visual Analytics (VIS) | 978-1-6654-8812-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/VIS54862.2022.00022

2771-9553/22/$31.00 ©2022 IEEE
DOI 10.1109/VIS54862.2022.00022

technical expertise to use.

To address this gap, we contribute the design and development of **FairFuse**, an interactive visualization system for generating, analyzing, and auditing fair consensus rankings. We develop a model capturing rankings and candidate attributes for identifying candidate groups, group-based fairness metrics, and algorithms to generate fair consensus rankings. We propose a parallel-coordinates style visualization design for rankings with a focus on the group membership of candidate attributes. We develop novel visual encodings for group-based fairness metrics. FairFuse enables an iterative ranking- and fairness-oriented workflow, allowing decision-makers to visually inspect and edit consensus rankings as part of their decision-making process. Our use cases demonstrate how a decision-maker can use FairFuse in fairness-oriented ranking scenarios. We conclude by discussing emerging challenges in supporting fairness in ranking-based decision-making through interactive visualization systems.

2 Related Work

Visualization systems have been designed to aid decision-makers in inspecting stakeholder preferences for decision-making tasks [4, 12, 15, 19, 20, 24, 32, 35, 39, 41, 43, 44]. Some consider settings, like ours, in which preferences from multiple stakeholders are modelled as rankings [19, 20, 22, 32]. Most recently, Hindalong *et al.* [23] developed visual abstractions for inspecting and comparing two or more preferences. However, these works neither apply algorithms to automatically construct one integrated fair ranking nor does their integration of multiple rankings address the critical real-world challenge that preferences tend to contain biases about socio-demographic groups (*e.g.* different gender or racial identities).

While interactive and visual systems [1–3, 5, 9, 25, 40, 45, 47] highlight and mitigate socio-demographic biases, they are not targeted towards rank-oriented workflows and fair consensus rankings. Instead, they target predictive machine learning tasks like classification [1, 2, 5, 9, 40, 45] or restrict themselves to single ranks [3, 47].

In the algorithmic fairness community, the predominant mitigation to bias and discrimination is the notion of "group fairness" [38]. Group fairness is conceptualized as treating groups similarly [38]. The state-of-the-art includes both *metrics* [7, 17, 29, 36, 42, 46] to quantify bias in rankings and *algorithms* [28] to generate such fair consensus rankings. Specifically, Kuhlman *et al.* [28] address fair-consensus ranking generation for two socio-demographic groups, while Cachel *et al.* [11] extend this scope to the multi-group setting. FairFuse takes initial steps towards leveraging these algorithmic solutions and their metrics to aid decision-makers in combining multiple stakeholder rankings into fair consensus rankings.

3 Basics of Fair Consensus Ranking

We characterize the data model and tasks for decision-makers analyzing multiple stakeholder preferences and ultimately combining them to generate a fair consensus ranking.

3.1 Abstraction of Data Model

We are given a set of **candidates**, described by **attributes**, to be ranked. One of the attributes, typically a categorical attribute referred to as the **protected attribute** (such as gender, race, or income-level), is associated with bias measurement and mitigation. We refer to candidates sharing the same value of the protected attribute as **groups**, such as Man, Woman, or Non-binary groups in the Gender attribute. Stakeholders in the committee (called **rankers**) each order (rank) the set of candidates to create a list of **base rankings** provided to the **decision-maker**. A decision-maker (head of the committee) using our system generates **consensus rankings** with the aim to order the candidates such that the base rankings, and thereby rankers, mostly agree with the consensus ranking.

The consensus ranking also must be fair. For auditing the fairness of rankings, we employ two metrics: a group-specific pairwise fairness metric **FPR** (Favored Pair Representation) [11] to measure the fair treatment of each group in the ranking, and an aggregate fairness metric **ARP** (Attribute Rank Parity) [11] to quantify if the overall ranking across all groups satisfies the statistical parity fairness criteria [38]. In generating a fair consensus ranking, the decision-maker sets the **fairness threshold** value which controls the level of ARP represented in the consensus ranking. The later is then generated by a function utilizing the **Fair-Copeland Algorithm** [11]. A function **Kendall Tau distance** [26] computes the similarity/agreement between any two rankings.

3.2 Task Analysis

We define nine abstract tasks to guide the development of FairFuse, following procedures from task abstraction methodologies such as Lam *et al.* [31], and recent work on group decision making from Hindalong *et al.* [22]. These tasks support comparing rankings, investigating bias in rankings, and iteratively generating fair consensus rankings.

T1: Identify candidate positions across base rankings to assess high and low performing candidates according to the rankers.

T2: Identify protected attribute values (*i.e.* group membership of) and additional attributes of ranked candidates.

T3: Analyze the (dis)similarity across rankings, both between base rankings and between base and consensus rankings.

T4: Explore the distribution of the placement of groups, in each ranking, to compare advantages across groups.

T5: Understand the fair treatment or lack thereof of each group per ranking, as captured by the FPR metric.

T6: Intuit fairness of each ranking as a whole with respect to the statistical parity fairness criteria for the specified protected attribute, as captured by the ARP metric.

T7: Compare group fair treatment and fairness across rankings, both base and consensus alike.

T8: Generate consensus rankings by initiating the generation algorithm, while controlling their level of fairness.

T9: Iterate on and adjust generated consensus rankings to satisfy the desired trade-off between base rankings and degree of fairness.

3.3 Data Sets for Use Case Scenarios

For the demonstration of this fair ranking problem, we use a dataset of 60 students with scores in 3 subjects: Math, Reading and Writing [27], and convert them into base rankings. We use the provided race attribute composed of 5 abstract groups (Group A, Group B, ...) as a protected attribute and map it to concrete race categories: White, Black, Asian, *etcetera*. Since names are not provided, we generate random names for each candidate. While this dataset is used to demonstrate the visual encodings and features, FairFuse supports other datasets— an example use case with employee bonus distributions data is included in the supplement.

4 FairFuse Overview

FairFuse is designed to support the process of both analyzing and combining preferences from multiple rankers into a *fair* consensus ranking. In designing FairFuse, we develop core views based on parallel coordinates, augmented with custom visual encodings for fairness metrics, and interactive components for generating fair consensus rankings.

4.1 Ranking Exploration View

The *Ranking Exploration View* (Figure 1E) contains all candidates ordered into two or more base rankings (**T1**). Columns on the left correspond to input base rankings; while fair consensus rankings generated by the decision maker are appended to the right upon their creation. Drawing on designs from Nobre *et al.* [37] and Maguire *et*

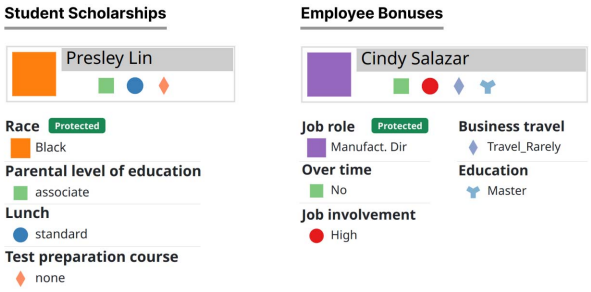

Figure 2: Candidate Card contains glyphs representing multi-variate attributes. The protected attribute is emphasized with a large (versus smaller) shape. Color represents the value of the attribute.

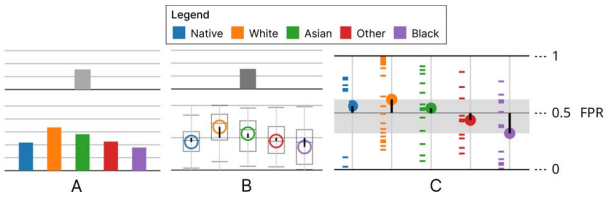

Figure 3: Group Fairness View design iterations. A) Colored barchart for FPR; gray bar for ARP. B) Dot plot for FPR, box-plot for distribution of groups in ranking. C) Final Design: Dot plot represents FPR, heatmap the distribution of groups, and shaded region the ARP.

al. [33], candidate attributes are displayed as glyphs in a *Candidate Card* (Figure 2) (**T2**).

This view uses parallel coordinates to compare candidates across different rankings (**T3**), drawing on features from similar rank oriented systems such as LineUp [18] and Hindalong *et al.* [22]. The order of candidates in a given column is based on candidate rank in the case of a base ranking columns, or the Fair-Copeland Algorithm in generated rankings. Each candidate appears across all rankings, with lines connecting them to illustrate change in position across rankings. Lines connecting the candidate across the rankings are colored based on the degree of change in the candidate's position between adjacent rankers. Candidates ranked higher in the subsequent ranking are colored in a gradient scale of blue, while those ranked lower are colored in a gradient scale of red.

To reduce parallel coordinates clutter (*e.g.* [21]) while maintaining task effectiveness, we hide lines for which both candidates on adjacent rankers are not visible within the screen. Clutter can also result from orderings of parallel coordinate columns [10]. Users can drag to re-arrange columns, and FairFuse can be readily extended with automatic ordering techniques. We also design a *Compressed Ranking View* mode (Figure 4) which represents a scaled-down version of the rankings. In this mode, the candidate cards (Figure 2) are initially hidden, but appear when hovering over a particular candidate. The protected attribute glyph is displayed with full saturation so that the decision-maker can explore how groups are distributed in each ranking (**T4**), while other attributes are desaturated so as to be visible while interfering less with the protected attribute color.

4.2 Group Fairness View

To support auditing rankings in terms of fairness, the *Group Fairness View* (Figure 1D) compactly captures fairness of a ranking at multiple levels of granularity: both at the level of individual groups, and holistically across groups for assessing fairness across rankings. The FPR metric [11] captures if a specific group is fairly treated throughout the ranking (**T5**). Specifically, FRP score $= 0.5$ denotes totally fair group treatment, while < 0.5 represents under-advantage and > 0.5 over-advantage. The ARP metric [11] captures if statistical parity fairness is satisfied by the ranking overall (**T6**), *i.e.*, *all* groups are comparably treated to each other. Here, ARP $= 0$ is absolute fairness, anything higher is further and further from total fairness. This novel fairness view is critical to capture the notion that in multiple-group settings one or more groups may be fairly treated, while others may be unfairly over- or under-advantaged.

In designing the Group Fairness View, we initially explored 2 alternate prototypes (Figure 3). Because FPR and APR are scalar values, we first represented the FPR and ARP fairness scores with bar encodings at the top of ranking columns (Figure 3A). However, after determining that this design made it difficult to identify over-advantaged and under-advantaged groups (**T5**), our second design placed an axis at FPR 0.5 and adopted a hybrid dot-plot and box-plot encoding (Figure 3B). This change supports more semantically meaningful visual queries. For example, dots above FPR 0.5 represent over-advantaged groups, informing the decision maker that they are unfairly receiving a larger share of favorable rank positions (**T5**).

As explained below, our third and final design variation of the Group Fairness View as depicted in Figure 3C offers additional advantages. Since ARP measures the difference between the maximum and minimum FPR scores, we can visualize the ARP score with the region between the scores of the respective group. This change enables visual queries within and across rankings to assess group fairness of each ranking(**T6**), as a smaller ARP value would create a smaller shaded region. To mitigate the limitations of box-plots for showing non-contiguous distributions, we adopt a marginal mark-based distribution plot. Finally, the view affords interactive features, such as displaying exact FPR and ARP values on hover, and highlighting groups in the parallel coordinates plot on click.

4.3 Consensus Generation and Similarity View

The *Consensus Generation and Similarity View* (Figure 1A) provides functionality for the decision maker to generate and compare fair consensus rankings (**T8**). The consensus generation component includes a fairness threshold slider used to adjust the level of fairness that should be reflected in the generated consensus ranking. Algorithmically, this is accomplished by passing the base rankings and the fairness target value to the recently innovated Fair-Copeland algorithm [11], which then computes and returns a new consensus ranking. Because a set of base rankings are unlikely to be completely unfair from the outset, the slider includes a gradient overlay to indicate that the fairness threshold will only produce fairer results if changed in a particular region. Similarly, on the other extreme, if the slider is set to 0, it will generate a consensus ranking soley based on the input base rankings.

As the decision maker initiates the generation of a consensus ranking, they can assess the overall agreement between the base versus this new consensus rankings (**T3**) via the similarity matrix view (Figure 1B). For example, if they generate rankings with a high level of fairness, the resulting ranking may deviate more from some base rankings than others. Similarity is calculated using a common measure for rank dissimilarity called Kendall-Tau distance [26], with darker squares representing more similarity between two rankings. This similarity component also aids the decision-maker in iterating over alternate consensus rankings (**T9**) to finalize the consensus decision. In designing this view, we considered alternatives such as arc diagrams embedded into the ranking exploration view (which were too cluttered), dissimilarity as the metric was traditionally defined (with inversion considered more interpretable), and variations of the matrix orientation.

4.4 Additional Interactions and Workflow

FairFuse provides additional interactions to support the decision-maker in a fairness-oriented rank analysis and generation workflow.

Figure 4: Compressed view with a group selected. When a group is selected in the Group Fairness View, all candidates of that group are highlighted, facilitating the decision-maker in focusing on both group and individual fairness considerations.

Ranking Exploration, Group Fairness, and Consensus Generation and Similarity Views include design elements that respond to user actions such as clicks and hovers. A user hovering in the Ranking Exploration view, for example, will highlight a Candidate Card for easier exploration across views (**T1, T2**). A click in this scenario "pins" a candidate for comparison against other candidates. Brushing is also enabled [21], allowing the user to drag select ranges of candidates within particular columns, which is particularly useful in the compressed views. Similar hover and click functions are available in other views, mainly oriented towards emphasizing or de-emphasizing visual components to enable the decision-maker to focus on particular tasks.

To support iteration and adjustment of consensus rankings (**T9**), FairFuse provides the decision-maker with editing features on consensus rankings. FairFuse supports manual editing of fair consensus rankings (**T7**), as the decision-maker may have additional context and information that they need to preserve in the resulting ranking. Decision-makers may adjust the fairness threshold of a consensus ranking to obtain another result, create or "pin" rankings, and manually adjust the position of candidates. Importantly, repositioning candidates immediately triggers the recalculation of fairness metrics, showing the decision-maker how fairness is lost or gained through their manual editing.

5 Use Case Scenarios Using the FairFuse System

A scholarship administrator, Jo, is responsible for determining the merit scholarship package of prospective students[1]. Jo needs to combine the recommendations of three rankers, teachers in Math, Reading, and Writing, and form a single ranking to allocate the merit scholarships. Cognizant that systemic and societal biases can affect how students of differing races perform in academic subject exams, which in turn can affect how students are perceived by subject-specific rankers, Jo seeks to detect and mitigate excess bias in the consensus ranking to ensure all groups are comparably treated.

Jo loads the data of base rankings given by the teachers along with candidate attribute information into the FairFuse. Jo uses the *Similarity View* to assess to what degree each base ranking agrees with others, along with visually inspecting the lines between adjacent rankings in the *Rank Exploration View*. At this point Jo uncovers that the Math teacher disagrees to some extent with the other teachers, which can make a consensus challenging, even before considering fairness.

Next, Jo switches to the compressed view to evaluate how candidate attributes are distributed across rankings. Auditing primarily for fairness, however, Jo pays particular attention to the protected attribute, race. For this task, Jo studies the *Group Fairness View* on top of each ranking (Figure 1D), which shows distributions of protected attributes throughout the ranking. Jo notices immediately that the FPR fairness metric indicates that white students have a stronger advantage over students from other races. On closer examination, Jo discovers that across all rankings, students from the white group are clustered at the top, while students from the black group are clustered more towards the bottom. This is then reflected in the ARP scores (gray area) of the rankings, indicating the base rankings in general are far from fair as defined by statistical parity.

After exploring and comparing the similarities and fairness of the base rankings (Figure 1B,D), Jo initiates the auto-generation of a consensus ranking, using the Consensus Generation View. Immediately, Jo notices that the consensus ranking reflects the biases found in base rankings. Jo then progressively adjusts the Fairness Threshold (Figure 1A) to generate a fairer consensus ranking. Throughout this process, Jo references the Similarity View matrix and base rankings themselves to evaluate the extent to which base rankings are represented in the fair consensus. Honing in on a consensus ranking that balances the desired trade-off between the fairness and preference representation, Jo makes manual swaps between candidates to refine the target consensus ranking. With each edit, Jo's changes are audited visually by changes in the Group Fairness View (Figure 1D), helping ensure this manual manipulation does not drastically change the desired fairness measure. The resulting consensus ranking is both fair with respect to mitigating the over-advantage of white students and their disproportionately large merit awards, while ensuring the teacher recommendations expressed by base rankings are adequately combined and represented.

6 Discussion & Future Work

In designing FairFuse, we learn about the challenges and opportunities for integrating fairness-oriented algorithms into visualization-driven workflows. State-of-the-art algorithms only consider a single protected attribute per candidate, but future work should engage with the algorithmic and visualization challenges surrounding multiple protected attributes and intersectionality [14]. The current similarity view shows the agreement between any two rankings, but matrix views can be confusing [37]. Future designs might explore how particular aspects of the workflow can be used to inform new encodings that better represent agreement between base rankings and generated consensus rankings. For the current implementation, we use the Fair-Copeland algorithm. Future work could also explore how to support multiple algorithms in a single tool, and comparisons between them. While our use cases illustrate the utility of a system [30], task-based user studies involving fairness will also be important future work. User studies might explore ethical issues surrounding algorithmic fairness, such as the potential to "fair wash" results by deceptively using metrics to promote unfair outcomes [8].

7 Conclusion

Fusing preferences of multiple stakeholders into a fair consensus decision expressed by a result ranking is ubiquitous yet an incredibly challenging process. To support fair consensus-building workflows, we introduce FairFuse, a visualization system for auditing, analyzing and generating consensus rankings. FairFuse interactively aids the decision-maker in generating and refining fair consensus rankings given a set of base rankings. With custom visualizations encoding fairness metrics from fair-algorithms research, FairFuse enables decision-makers to visually and interactively explore and audit base- and generated- rankings. We demonstrate how FairFuse supports fairness-oriented ranking workflows through use cases, yielding a foundation for future studies at the intersection of fairness, ranking, and visualization.

Acknowledgments

This work was supported in part by NSF IIS #2007932.

[1] An additional usage scenario is presented in supplemental material.

REFERENCES

[1] Ai 360. https://ai-fairness-360.org/.

[2] What if tool. https://pair-code.github.io/what-if-tool/.

[3] Y. Ahn and Y.-R. Lin. Fairsight: Visual analytics for fairness in decision making. *IEEE trans. vis. and comput. graph.*, 26(1):1086–1095, 2019.

[4] S. Bajracharya, G. Carenini, B. Chamberlain, K. Chen, D. Klein, D. Poole, H. Taheri, and G. Öberg. Interactive visualization for group decision analysis. *International Journal of Information Technology & Decision Making*, 17(06):1839–1864, 2018.

[5] N. Bantilan. Themis-ml: A fairness-aware machine learning interface for end-to-end discrimination discovery and mitigation. *Journal of Technology in Human Services*, 36(1):15–30, 2018.

[6] J. Bartholdi, C. A. Tovey, and M. A. Trick. Voting schemes for which it can be difficult to tell who won the election. *Social Choice and welfare*, 6(2):157–165, 1989.

[7] A. Beutel, J. Chen, T. Doshi, H. Qian, L. Wei, Y. Wu, L. Heldt, Z. Zhao, L. Hong, E. H. Chi, et al. Fairness in recommendation ranking through pairwise comparisons. In *Proc. 25th ACM SIGKDD Int. Conf. on Knowledge Discovery & Data Mining*, pp. 2212–2220, 2019.

[8] E. Bietti. From ethics washing to ethics bashing: a view on tech ethics from within moral philosophy. In *Proceedings of the 2020 conference on fairness, accountability, and transparency*, pp. 210–219, 2020.

[9] S. Bird, M. Dudík, R. Edgar, B. Horn, R. Lutz, V. Milan, M. Sameki, H. Wallach, and K. Walker. Fairlearn: A toolkit for assessing and improving fairness in ai. *Microsoft, Tech. Rep. MSR-TR-2020-32*, 2020.

[10] M. Blumenschein, X. Zhang, D. Pomerenke, D. A. Keim, and J. Fuchs. Evaluating reordering strategies for cluster identification in parallel coordinates. In *Computer Graphics Forum*, vol. 39, pp. 537–549. Wiley Online Library, 2020.

[11] K. Cachel, E. Rundensteiner, and L. Harrison. Mani-rank: Multiple attribute and intersectional group fairness for consensus ranking. In *2022 IEEE 38th Intl. Conf. on Data Engineering (ICDE)*. IEEE, 2022.

[12] G. Carenini and J. Loyd. Valuecharts: analyzing linear models expressing preferences and evaluations. In *Proceedings of the working conference on Advanced visual interfaces*, pp. 150–157, 2004.

[13] A. Chouldechova and A. Roth. A snapshot of the frontiers of fairness in machine learning. *Communications of the ACM*, 63(5):82–89, 2020.

[14] K. Crenshaw. Mapping the margins: Intersectionality, identity politics, and violence against women of color. *Stan. L. Rev.*, 43:1241, 1990.

[15] E. Dimara, P. Valdivia, and C. Kinkeldey. Dcpairs: A pairs plot based decision support system. In *EuroVis-19th EG/VGTC Conf. Vis.*, 2017.

[16] C. Dwork, R. Kumar, M. Naor, and D. Sivakumar. Rank aggregation methods for the web. In *Proceedings of the 10th international conference on World Wide Web*, pp. 613–622, 2001.

[17] S. C. Geyik, S. Ambler, and K. Kenthapadi. Fairness-aware ranking in search & recommendation systems with application to linkedin talent search. In *Proceedings of the 25th acm sigkdd international conference on knowledge discovery & data mining*, pp. 2221–2231, 2019.

[18] S. Gratzl, A. Lex, N. Gehlenborg, H. Pfister, and M. Streit. Lineup: Visual analysis of multi-attribute rankings. *IEEE transactions on visualization and computer graphics*, 19(12):2277–2286, 2013.

[19] P. Hansen and F. Ombler. A new method for scoring additive multi-attribute value models using pairwise rankings of alternatives. *Journal of Multi-Criteria Decision Analysis*, 15(3-4):87–107, 2008.

[20] Q. Hayez, Y. De Smet, and J. Bonney. D-sight: a new decision making software to address multi-criteria problems. *International Journal of Decision Support System Technology (IJDSST)*, 4(4):1–23, 2012.

[21] J. Heinrich and D. Weiskopf. State of the art of parallel coordinates. In *Eurographics (State of the Art Reports)*, pp. 95–116, 2013.

[22] E. Hindalong, J. Johnson, G. Carenini, and T. Munzner. Towards rigorously designed preference visualizations for group decision making. In *2020 IEEE Pacific Vis. Symp. (PacificVis)*, pp. 181–190. IEEE, 2020.

[23] E. Hindalong, J. Johnson, G. Carenini, and T. Munzner. Abstractions for visualizing preferences in group decisions. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW1):1–44, 2022.

[24] S. Hong, M. Suh, N. Henry Riche, J. Lee, J. Kim, and M. Zachry. Collaborative dynamic queries: Supporting distributed small group decision-making. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2018.

[25] W. Jin, D. Gromala, C. Neustaedter, and X. Tong. A collaborative visualization tool to support doctors' shared decision-making on antibiotic prescription. In *Companion of the 2017 ACM Conf. on Computer Supported Cooperative Work and Social Comput.*, pp. 211–214, 2017.

[26] M. G. Kendall. A new measure of rank correlation. *Biometrika*, 30(1/2):81–93, 1938.

[27] R. Kimmons. Exam scores. http://roycekimmons.com/tools/generated_data/exams.

[28] C. Kuhlman and E. Rundensteiner. Rank aggregation algorithms for fair consensus. *Proceedings of the VLDB Endowment*, 13(12), 2020.

[29] C. Kuhlman, M. VanValkenburg, and E. Rundensteiner. Fare: Diagnostics for fair ranking using pairwise error metrics. In *The World Wide Web Conference*, pp. 2936–2942, 2019.

[30] H. Lam, E. Bertini, P. Isenberg, C. Plaisant, and S. Carpendale. Empirical studies in information visualization: Seven scenarios. *IEEE trans. on visualization and computer graphics*, 18(9):1520–1536, 2011.

[31] H. Lam, M. Tory, and T. Munzner. Bridging from goals to tasks with design study analysis reports. *IEEE trans. on visualization and computer graphics*, 24(1):435–445, 2017.

[32] W. Liu, S. Xiao, J. T. Browne, M. Yang, and S. P. Dow. Consensus: Supporting multi-criteria group decisions by visualizing points of disagreement, 2018.

[33] E. Maguire, P. Rocca-Serra, S.-A. Sansone, J. Davies, and M. Chen. Taxonomy-based glyph design—with a case study on visualizing workflows of biological experiments. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2603–2612, 2012.

[34] N. Mahyar, W. Liu, S. Xiao, J. Browne, M. Yang, and S. P. Dow. Consesnsus: Visualizing points of disagreement for multi-criteria collaborative decision making. In *Companion 2017 ACM Conf. Computer Supported Cooperative Work and Social Comput.*, pp. 17–20, 2017.

[35] J. Mustajoki and R. P. Hämäläinen. Web-hipre: Global decision support by value tree and ahp analysis. *INFOR: Information Systems and Operational Research*, 38(3):208–220, 2000.

[36] H. Narasimhan, A. Cotter, M. Gupta, and S. Wang. Pairwise fairness for ranking and regression. In *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, pp. 5248–5255, 2020.

[37] C. Nobre, D. Wootton, L. Harrison, and A. Lex. Evaluating multivariate network visualization techniques using a validated design and crowdsourcing approach. In *Proceedings of the 2020 CHI conference on human factors in computing systems*, pp. 1–12, 2020.

[38] D. Pedreshi, S. Ruggieri, and F. Turini. Discrimination-aware data mining. In *Proc. 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 560–568, 2008.

[39] P. G. Pham and M. L. Huang. Qstack: Multi-tag visual rankings. *J. Softw.*, 11(7):695–703, 2016.

[40] P. Saleiro, B. Kuester, L. Hinkson, J. London, A. Stevens, A. Anisfeld, K. T. Rodolfa, and R. Ghani. Aequitas: A bias and fairness audit toolkit. *arXiv preprint arXiv:1811.05577*, 2018.

[41] C. Shah. Collaborative information seeking. *Journal of the Association for Information Science and Technology*, 65(2):215–236, 2014.

[42] A. Singh and T. Joachims. Fairness of exposure in rankings. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 2219–2228, 2018.

[43] D. Weng, R. Chen, Z. Deng, F. Wu, J. Chen, and Y. Wu. Srvis: Towards better spatial integration in ranking visualization. *IEEE transactions on visualization and computer graphics*, 25(1):459–469, 2018.

[44] D. Weng, H. Zhu, J. Bao, Y. Zheng, and Y. Wu. Homefinder revisited: Finding ideal homes with reachability-centric multi-criteria decision making. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2018.

[45] T. Xie, Y. Ma, J. Kang, H. Tong, and R. Maciejewski. Fairrankvis: A visual analytics framework for exploring algorithmic fairness in graph mining models. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):368–377, 2021.

[46] K. Yang and J. Stoyanovich. Measuring fairness in ranked outputs. In *Proc. 29th intl. conf. scientific and statist. db. manage.*, pp. 1–6, 2017.

[47] K. Yang, J. Stoyanovich, A. Asudeh, B. Howe, H. Jagadish, and G. Miklau. A nutritional label for rankings. In *Proceedings of the 2018 international conference on management of data*, pp. 1773–1776, 2018.