9^{th} International Conference on Computational Social Science IC^2S^2
July 17-20, 2023, Copenhagen, Denmark

In vino veritas: Are the opinions of wine critics more informative than those of amateurs?

Keywords: social influence, crowd wisdom, expertise, recommender network, taste homophily

Introduction Whether it is the film currently running in cinemas, an experience in a restaurant that just opened, or the taste of vintage wine, people like to voice their judgments on matters of taste. For a few select individuals—the critics—expressing judgments on matters of taste has turned into a profession. Critics are employed in the daily and weekly press to assess fine dining restaurants, theater pieces, movies and wine labels, while some even run popular television shows. The role of critics as information producers is especially pronounced in websites such as Metacritic and Rotten Tomatoes, that have based their business model entirely on devising and displaying scores summarizing critics' opinions, or when they are hired as judges in prestigious competitions. Still, a number of questions about the ways in which critics can inform and influence the broader public in matters of taste remain unanswered. First, are the opinions of different critics more informative than those of amateurs in matters of taste? Second, how can the opinions of critics and amateurs be best combined and how would information flow between critics and amateurs in an informationally efficient network? And third, is it possible to identify influential critics and talented amateurs who might have the potential to become critics?

Method In this contribution, we put forward a novel approach for providing answers to the above questions combining methods from the recommender systems [3], machine learning [4] and network science communities [2]. We leverage the weighted k-nearest neighbors algorithm (k-nn), a classic recommender systems algorithm that encodes an array of strategies for choosing among experts as special cases [1]. For each individual and item to be predicted, our implementation of the algorithm draws advice from the k most similar other individuals in the rating database who have evaluated that item and weights their opinions according to a similarity sensitivity parameter ρ amplifying or dampening observed similarities. By assessing the out-of-sample performance of the algorithm when drawing advice only from critics, amateurs or both groups, we can compare the performance of different strategies involving each of these groups, and thus assess their relative informational value. In essence, the k-nn algorithm spans an advice network among different individuals, and it is thus possible to visualize and study the properties of such a "taste network" in any ratings database, even when the overlap in the ratings of different people is relatively small (i.e., sparse rater– item matrix). Individuals whose tastes appear relevant for many similar others, be they critics or amateurs, are sought often for their advice from the k-nn algorithm—they have a large recommender potential. This potential, however, can only materialize to recommender influence when the advisers have experienced the items that their advisees are considering. To assess the flow of influence between two categories of individuals—critics and amateurs—we adapt the notion of homophily from network science and show how it can be applied to domains of taste.

Dataset We apply this new framework to compare seasoned critics and amateurs in the domain of wine. We created a new dataset consisting of the ratings from both renowned wine critics' and regular wine consumers' (i.e., amateurs, non-professionals). We first obtained ratings by professional critics from Bordoverview and ratings by amateurs from Vivino by scraping the two websites. We then combined the two datasets and restricted our analyses to wine labels that were included in the Bordoverview list and had at least 5 reviews in Vivino and to Vivino users with more than 50 ratings in the resulting wine label list. This resulted to a dataset comprising of 1978 wine labels (322 different wines across 15 different vintages from 2004 to

2018 included), 14 professional critics (or wine magazines employing critics) and 120 Vivino amateurs. The dataset has 25907 ratings in total, and average density (i.e., mean proportion of rated wine labels relative to all wine labels) in the dataset is 4.7 % for amateurs and 53.3 % for professional critics.

Results First, when considering the similarities within groups we find that taste similarity among professional critics is much higher than among Vivino amateurs (see Figure 1), suggesting that critics are more consistent judges. What's more, the average similarity between critics and amateurs is higher than the average similarity among amateurs. This is also expressed in a recommender system based on the k-nn algorithm, as we find that a system based on critics performs better than a system based on amateurs. The difference is substantial, and larger than 3 % for the average amateur user (see Figure 2). A recommender system using both critics and amateurs can perform even better for k values equal or larger than 5, suggesting that there are some complementarities in using ratings from both groups. Yet for k values lower than 5 using ratings from both amateurs and critics does worse than using ratings only from critics (see Figure 2). This less-is-more effect is to a large extend produced because critics are more prolific raters and there are more data to be used when estimating their similarities with amateurs. Therefore, the apparent taste similarities between critics and amateurs can generalize better on unseen items as compared to the apparent similarities between amateurs and other amateurs. When drawing advice from the entire population, the tastes of amateurs are characterized by slight inbreeding homophily for low values k, and slight heterophily for intermediate values k, independently of the parameter ρ used. The wine critics, by contrast, are drawing most of their advice from other critics their tastes and are characterized by inbreeding homophily regardless of the parameters k and ρ used (Figure 4). Bettanne and Desseauve and Jeff Leve are the critics with the highest recommender potential (see Figure 3), yet the journal Decanter would have the most influence in a Bordeaux wine recommender system built from these data, as it has reported ratings on the largest number of items. Last, there is a sizable group of the amateurs users that appear to have a large recommender potential (larger than that all critics, see Figure 3 upper right panel), and some of them have very high correlations to both critics and amateurs (Figure 1, right), indicating that they might have the potential to become critics.

Conclusion We developed a new approach for assessing the informational value of critics (or other individuals) for different audiences, and identifying individuals with influential and informative tastes on any rating database. In sum, our results shed light to the function of wine critics as information producers, but also shows that there is some promise in building recommender systems that combine the ratings from both critics and amateurs.

References

[1] Pantelis P Analytis, Daniel Barkoczi, and Stefan M Herzog. Social learning strategies for matters of taste. *Nature Human Behaviour*, 2(6):415–424, 2018.

[2] Pantelis P Analytis, Daniel Barkoczi, Lorenz-Spreen Philipp, and Stefan M Herzog. The structure of social influence in recommender networks. In *Proceedings of the sixth ACM conference on the Web (WWW)*, pages 415–424. ACM, 2020.

[3] John S Breese, David Heckerman, and Carl Kadie. Empirical analysis of predictive algorithms for collaborative filtering. In *Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence*, pages 43–52. Morgan Kaufmann Publishers Inc., 1998.

[4] Özgür Şimşek. Linear decision rule as aspiration for simple decision heuristics. In *Advances in Neural Information Processing Systems*, pages 2904–2912, 2013.

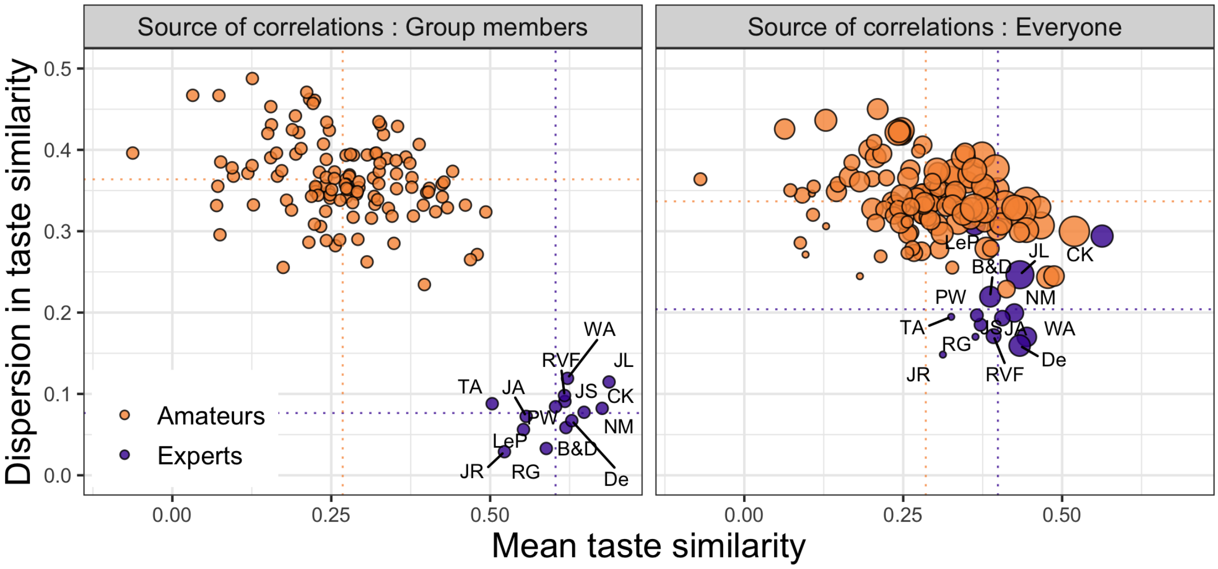

Figure 1: The position of 14 professional critics and 120 amateurs on a 2-dimensional plane defined by mean taste similarity (i.e., mean correlation) and dispersion in taste similarity (i.e., SD of correlations) with members of the same group Right: The position of the same 14 professional critics and 120 amateurs on the same plane, but this time with taste similarity calculated across all individuals (professional critics and amateurs). The color in both panels indicates whether the user is a professional critic or an amateur and the point size in the right panel indicates user recommender potential. Only correlations were two individuals had an overlap higher than 5 ratings were considered in this graph. **Initials of professional critics**: WA — Lisa Perotti Brown, NM — Neal Martin, JR — Jancis Robinson, TA — Tim Atkin, B & T — Bettanne and Desseauve, JS — James Suckling, JL — Jeff Leve, De — Steven Spurrier, James Lawther, Beverley Blanning and Jane Anson, RVF — Poels, Durange and Maurange, JA — Jane Anson, LeP — Dupont, PW — DeGroot, RG — Rene Gabriel, CK — Chris Kissack.

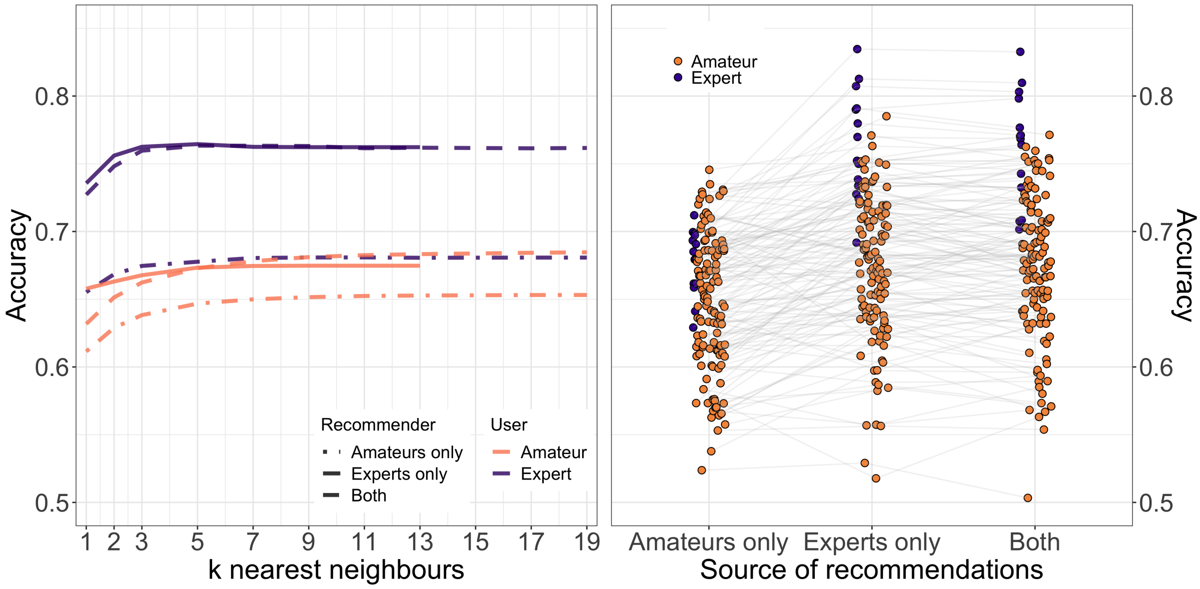

Figure 2: Performance of the recommender system for different groups (left) and individuals (right). Left: The average performance of the k-nn algorithm based only on amateurs, only on experts or both amateurs and experts for different values k for the amateur and experts groups Right: The individual level performance of the k-nn algorithm based only on amateurs,only on experts or both amateurs and experts for k = 5.

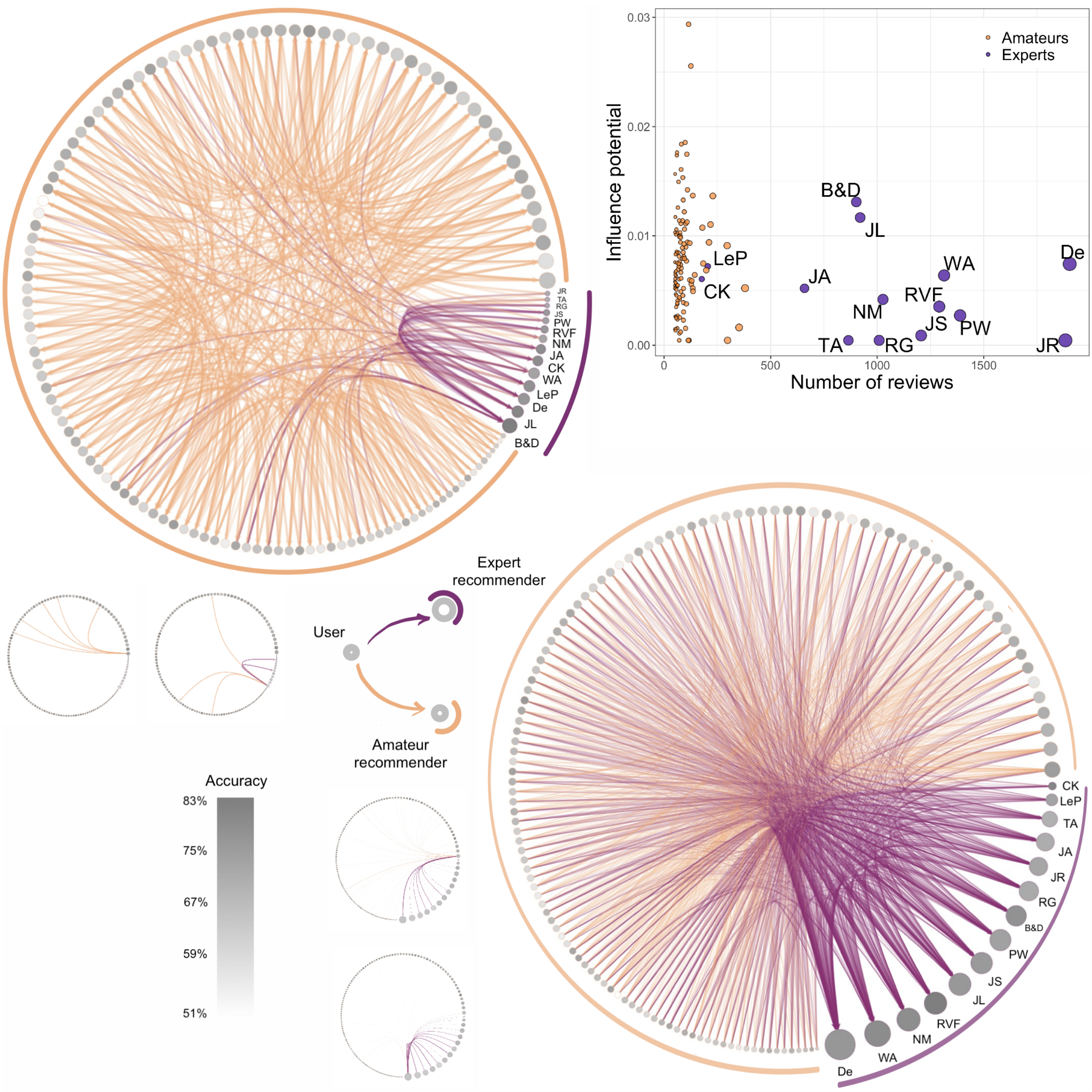

Figure 3: The recommender potential and recommender influence of different individuals. Nodes represent individuals, the size of the nodes represents the potential (upper left circle) or recommender influence (bottom right circle), respectively, of different people in the recommender network spanned by the k-nn algorithm. Orange edges indicate that advice is sought (or provided) from an amateur and purple edges indicate that advice is sought (or provided) from a professional critic (see bottom Left). The colour of the nodes indicates the accuracy of the algorithm for different individuals in the dataset (see bottom Left). Edges with weights smaller than 0.05 do not appear in the visualization to prevent overcrowding the graph. Upper Left: The influence graph produced by the initial call of k-nn, disregarding missing values. The edges (arrows) are pointing to the individuals from whom k-nn first seeks advice for the target individual. Lower Right: The influence graph eventually produced by k-nn. When a user called by k-nn has not rated a wine label, the next user in the correlation rank is consulted. This process continues until k advisers have been found or until the pool of potential advisers is exhausted. Upper Right: Amateurs and professional critics placed on a 2-dimensional plane defined by the number of items they have evaluated (x-axis) and their recommender potential (y-axis). Critics are depicted with purple color and amateurs with yellow. The size of the nodes indicates the total influence of different individuals.

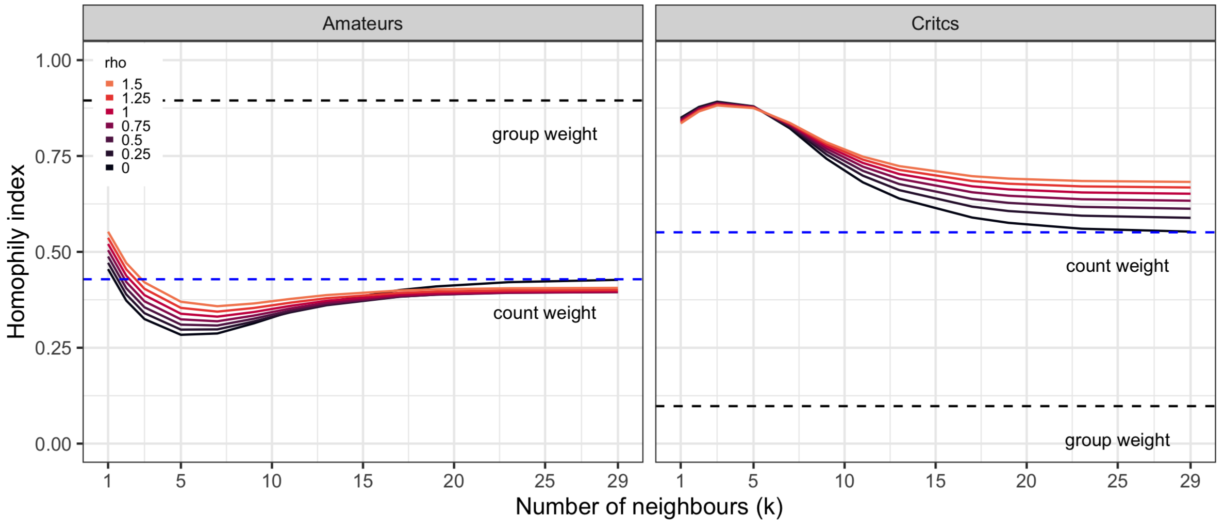

Figure 4: The homophily index of amateurs (left) and critics (right) as a function of the value k in the k-nearest neighbors algorithm. Different values are represented with lines of different colour. The vertical lines represent homophily baselines corresponding to the proportion of group members in the population (group weight) and the proportion of the ratings contributed by the members of each group (count weight).