Clustering over user features and latent behavioral functions with dual-view mixture models
-Reviews-

Dear Reviewers,

...

List of detailed responses:
(Q: reviewer comments. A: authors replies.)

Editor-in-Chief

Q1: Let me emphasize what the AE already mentioned: "Maybe modify the figures to not only differentiate groups by color, but also with different plot symbols so that they can also be distinguished when printed in gray scale."
Our print issue by default gets published in grayscale (unless you want to pay for color figures). So many of the differently colored points in your figures will become almost indistinguishable for readers of our print issue.
A1: ...

Q2: Please minimize the use of footnotes and add the text back to the main text, e.g., in parentheses.
A2: ...

Q3: In footnotes 3 and 6, you refer to the coda and mclust R packages. Formally cite R and R packages you use in the main text and add to the References. Look at a recently published article from our journal for reference.
A3: ...

Q4: Place table captions above the table.
A4: ...

Q5: Please check the completeness of your citations, in particular with respect to missing page numbers and missing editor names for conference proceedings.
A5: ...

Associate Editor

Q1: Please consider changing behaviors view to behavior view or behavioral view.
A1: ...

Q2: Maybe try to emphasize more already at the early beginning of the manuscript that the feature and the behavioral cluster structure are the same and maybe explain why and when they could be assumed to be aligned; otherwise the improvement in the cold-start problem is unclear.
A2: ...

Q3: The variables used by Cheng et al. (2014) are unclear.
A3: ...

Q4: Bonilla et al (2010) and Abbbasnejad et al (2013) seem to not use dual-view clustering. This should be more explicitly stated and it should be made clearer why they are mentioned, i.e., their relationship to social network clustering.
A4: ...

Q5: One could state that the generalization to infinite number of clusters leads to a non-parametric model.
A5: ...

Q6: Below Equation (1) please correct p_k to p_K.
A6: ...

Q7: The statement "In this case, every component is associated to a cluster." should be toned done as there also exist applications of mixtures where several components are used to form a cluster.
A7: ...

Q8: Use consistent bold or non-bold face for a_u, α and index u.
A8: ...

Q9: I am not sure if Equation (7) is correct. It gives the impression that this holds for all k where $n_k = 0$, i.e., an infinite amount of k. But rather I would assume that the combined weight of all empty components should be proportional to α. This comment also holds for Equation 24. Also Equations 25 and 26 could be modified to avoid the restriction of k within the probabilities.
A9: ...

Q10: It is unclear why the set $U = \{u_1, \ldots, u_U\}$ is introduced, i.e., why referring to $1, \ldots, U$ is not sufficient. For $\mathbf{a_u} = (\mathbf{a_1}, \ldots, \mathbf{a_D})$ the question arises if a subindex u should also be used for a_i, i.e., a_{ui}.
A10: ...

Q11: Below Equation 17 a different notation is used for the precision.

Q12: It would be good to provide some justification, e.g., a reference, for the use of the ML estimators for the priors. One might also explain why the use of these estimates is sufficiently flexible if used as priors for the mixture components.

Q13: One might explain why choosing m = 3 was assumed to be sufficient.

Q14: Add a reference for the iris dataset. Note also that Anderson (1935)'s iris dataset would seem to consist of more observations than used in the example and more variables.

Q15: Maybe explain why different initialization schemes are used for the two experiments and the iris dataset.

Q16: Maybe modify the figures to not only differentiate groups by color, but also with different plot symbols so that they can also be distinguished when printed in gray scale.

Q17: Regarding the additional analysis using the iris dataset it would be good to better motivate it. Please specify which observations are included, which variables are included and more clearly explain why it was complemented with artificial data and not a real dataset was used which already contains all required features for applying the proposed model. When evaluating the results it would be good to point out that the class labels are used as ground truth for evaluation while, however, it is unclear in cluster analysis if recovering true class labels is necessarily the aim.

Q18: Maybe add round parentheses around vectors, e.g., pi and alpha at the beginning of Appendix A.