Abstract Evaluation

Name of Editor: Hayden Foote

Identify the below sections in the abstract – if you identify them, copy and paste the text/summarize as instructed. In all cases, add comments if: something is missing, the text could be made clearer and/or the arguments stronger.

* Started with one or two facts that relate to the problem statement (copy them here)

Next-generation surveys (including Rubin, Euclid, CSST, 4MOST, and DESI) will furnish cutting-edge data on an extensive sample of galaxies.

* Explained why these facts are important (copy line here)

These state-of-the-art photometric and spectroscopic datasets will offer an opportunity to further our understanding of galaxy evolution and dark matter (DM).

* Introduced the problem (rewrite the problem in your own words)

Copied problem: However, the density distribution of DM at galaxy scales is heavily influenced by astrophysical feedback processes. These complicated processes are difficult to fully incorporate into classical techniques to constrain DM masses at these scales.

My interpretation: Baryonic feedback is expected to heavily influence the distribution of DM within galaxies, but feedback is difficult to account for in “classical” models.

Notes: The problem itself (absence of feedback modelling) is very clear, and the proposal is very well-motivated in general. However, I’m not quite sure what is meant by “classical” models in this context. I assume this means anything not involving machine learning, but this could be made more clear.

* Stated the goal (copy it here)

We propose to expand beyond classical mass models with a supervised learning algorithm to infer the DM content of galaxies from luminous matter using data from the TNG100 simulation public catalog.

Notes: Again, I’m not sure what you mean by “classical mass models.” Knowing exactly what model you’re trying to improve upon would strengthen the argument as to why the new approach is needed. To improve the overall clarity of what you’re proposing, it would help to explicitly state that you’re training a new pipeline with simulation data for future use on survey data. You mention this further down, but it took me longer than it should have to grasp what the goal is. The way this is currently written, it makes me think the goal of the proposal is to simply analyze the TNG100 simulations, which is disconnected from the observational motivation above.

* What is the key component? (your words)

To train a new survey data analysis pipeline by analyzing the TNG100 catalog.

* What is the target? (your words)

To determine the DM content of galaxies based on their luminous matter.

* Explained the strategy. (copy here)

We will adopt several observable data metrics, including brightness in different bands, baryonic masses, stellar half-mass radii, and velocity dispersions.

Notes: This may just be my unfamiliarity with machine learning, but it’s still unclear to me how doing this is going to solve the problem. One more sentence about how this technique accounts for feedback would strengthen the argument. I’m not too familiar with ML so this may not be right, but it would help a lot to explicitly say that the sims do all the feedback, so your model - when correctly trained on the simulations - should also account for it, and therefore improve on existing pipelines. You do imply all of this throughout the abstract, but I had to read it way more times than a reviewer would to understand the overall argument.

* Stated the importance of the solution *to the subfield*  (copy here)

Using these parameters, we will recover DM masses and radii with verifiable accuracies from the TNG100 simulation. Future applications of this pipeline to next-generation surveys will allow for critical predictions about DM from the real galaxy data observed.

Notes: As I said in the previous comment, I’m still missing an explicit connection between the simulation analysis and the stated problem. The way the first sentence is worded, it doesn’t seem like your proposed pipeline is doing anything new, since the DM properties are already in the simulation data. Your last sentence here is great, but I needed to understand this much further up in the abstract. It may also help to replace “from the real galaxy data observed” (since this is implied by “surveys”) with something like “while accounting for baryonic processes.”

* Explained the broader implications of results to *other subfields*  (copy here)

Notes: I don’t think this was mentioned but take that with a grain of salt since that depends upon one’s definition of the subfield (I interpret it to be the galaxy-halo connection in general) in this case. If you define the subfield as ML pipeline development, then the application to real survey data would fit fine here. I guess that would depend on who you’re sending this proposal to.

Overall, great start! It’s very well motivated, and the pieces are definitely all there, just needs a bit of reworking to make the logic flow more smoothly.