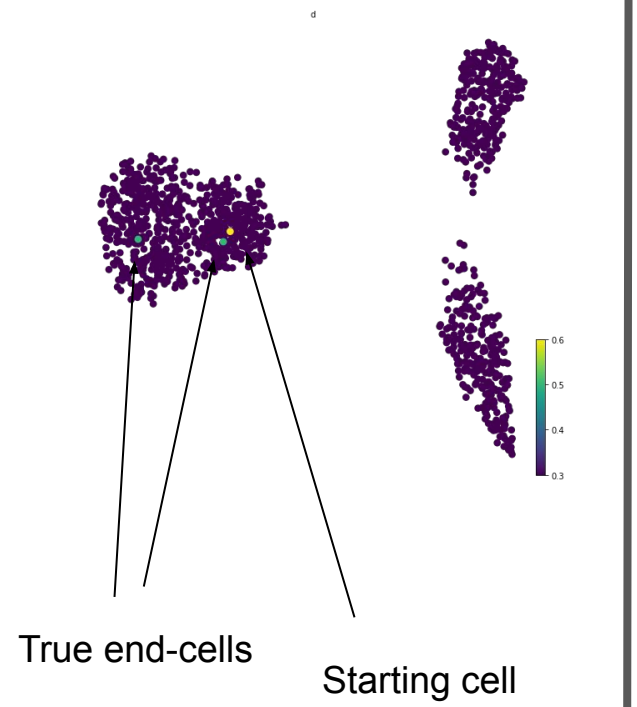
Update on dpt-distance

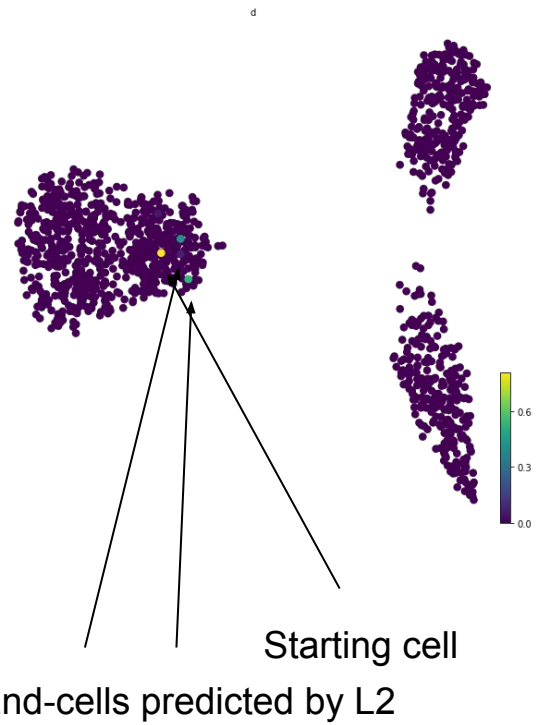
Overview of the pushs L2 / DPT give

Here you see the push of the L2-/dpt-map and the true descendants of the yellow cell. I checked several cells, and the different pushs are really similar to those depicted here.

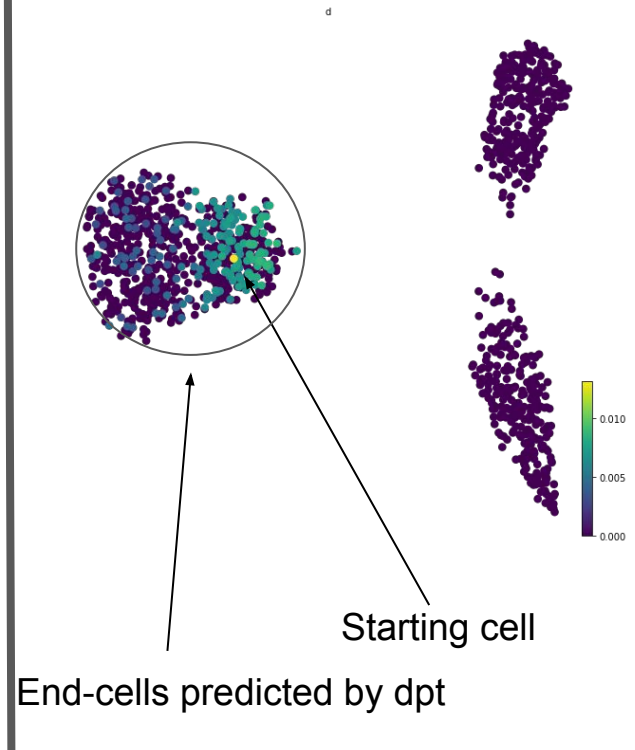
True map



L2 approach

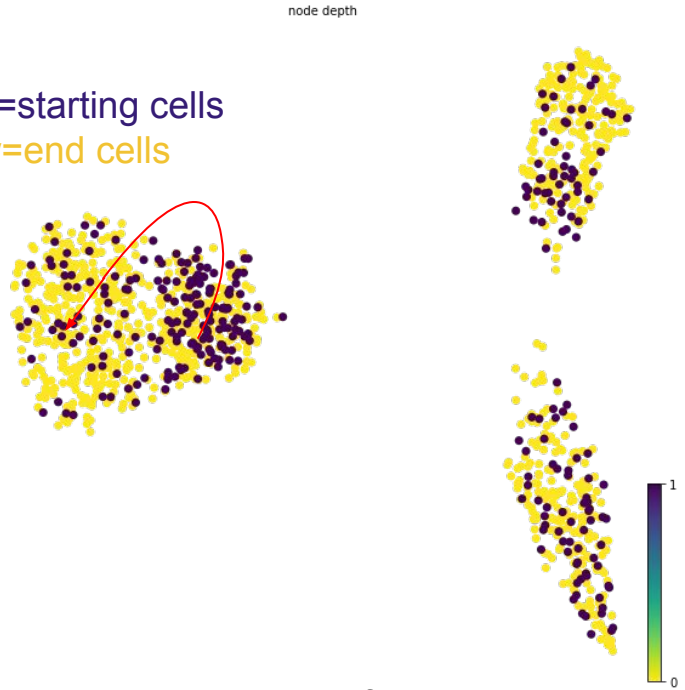


Dpt approach



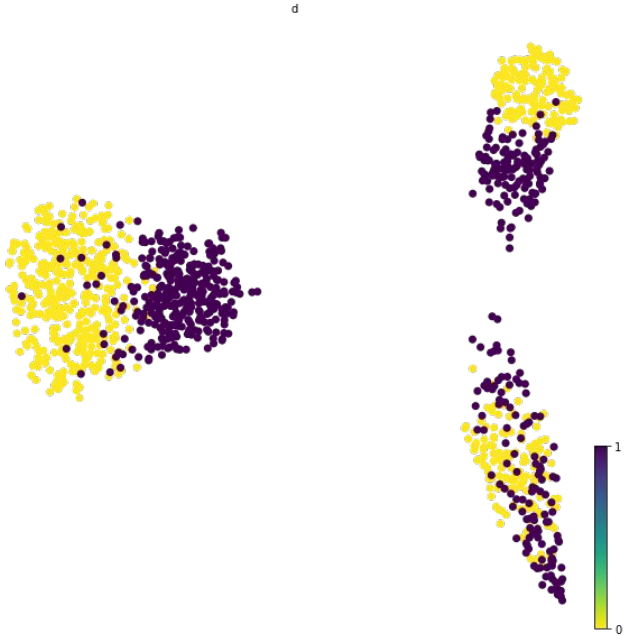
Right now, the time points I use for OT look like this (this is how moscot-Lineage does it):

violett=starting cells
yellow=end cells



The red arrow shows the path of one descendent. With the barcode information as well it is most likely possible to link them, but using only the expression profile it is not.
Cell from both time points are spreading more or less over the same area, which is quite untypical for a developmental process .

What I want the time points to look like



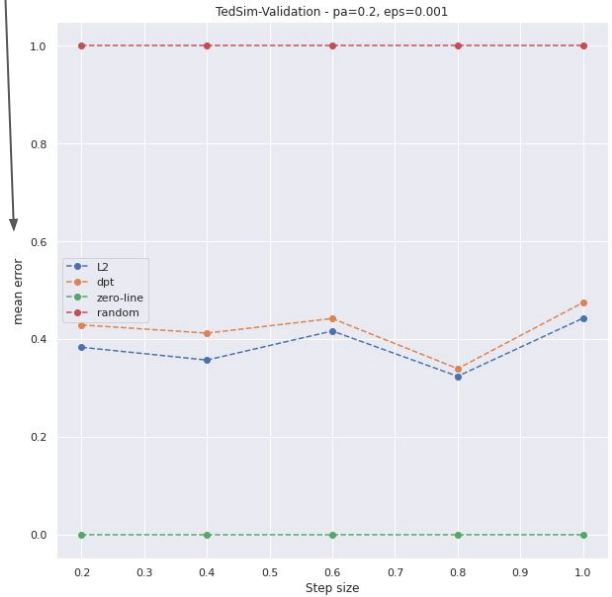
This seems like a more realistic depiction of cell development since the cells develop in a coordinated way.
I obtained this data by changing some parameters in TedSim. I'm doing the analysis on this data now.

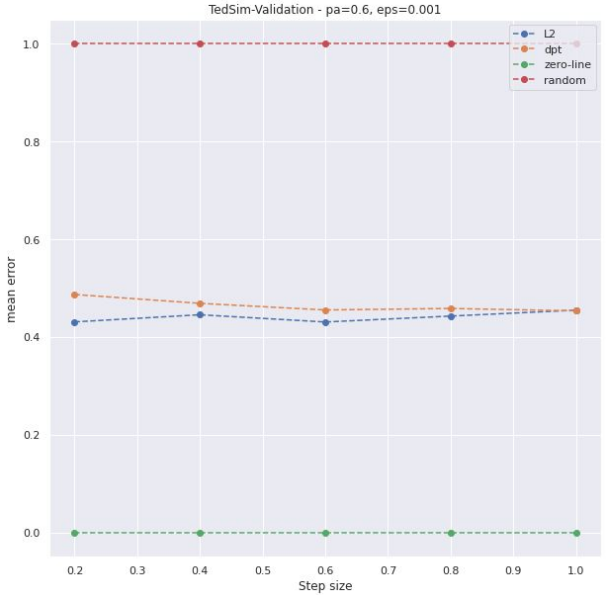
Here are the results for the approach as in moscot-Lineage

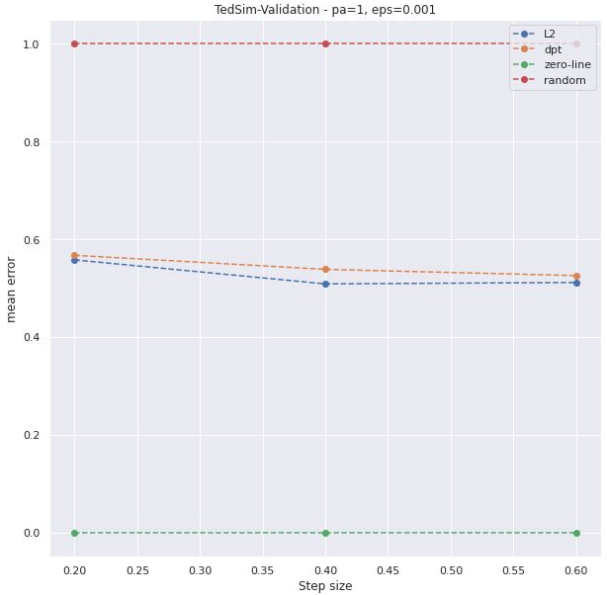
moscot-Lineage approach

Mean error measures the emd of predicted push to true push.

pa is the probability of asymmetric cell division (corresponds approximately to the probability of entering a new cluster)





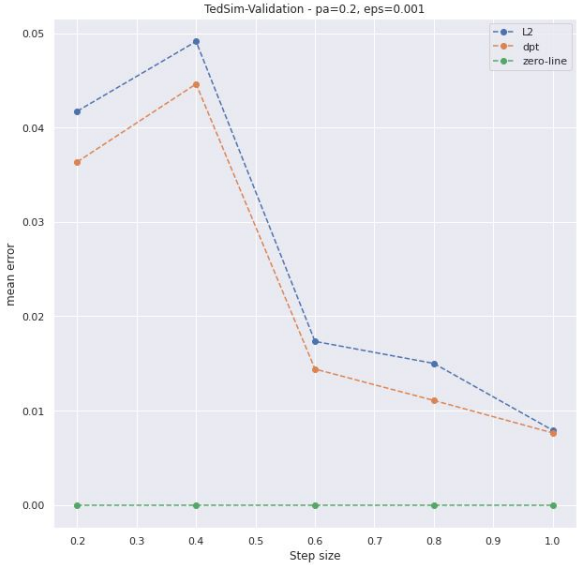


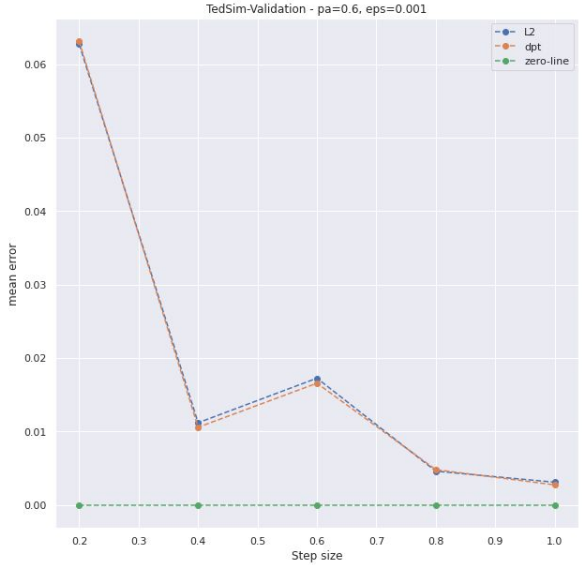
Step size is how much distance there is between cell clusters

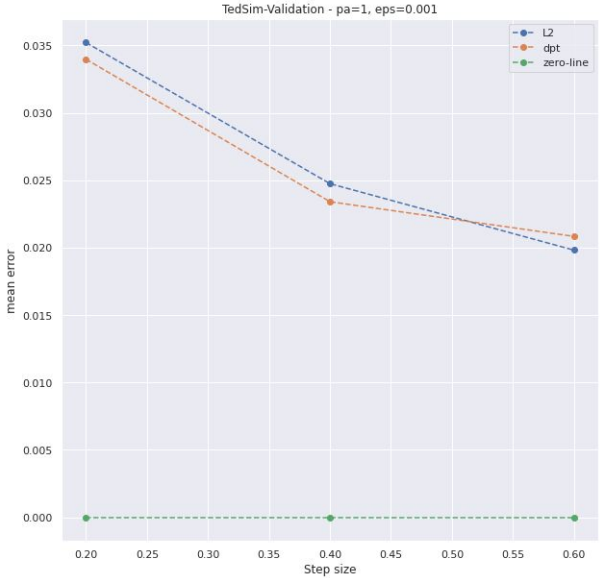
Here L2 seems to outperform dpt a little.

moscot-Lineage approach

For these last graphs, the mean error was calculated using the emd, and the emd used the L2-distance as transportation cost. With a look at slide 2, the push of dpt is much more spread out, so the emd from the dpt-push to the true descendents will be higher only by that fact alone (L2 squares distance, and dpt only cares about the graph, not about the distance in L2). Therefore, to make it fair, I did the validation where the emd is with respect to dpt distance. For this approach, dpt outperforms L2 a little.







Outlook

So as of now, there is not a big difference in performance between L2 and dpt.

I think how the data / time points as they are right now are not ideal to validate L2/dpt-maps based on gene expression alone. I've simulated some data for which it is easier / a more realistic depiction of cell development (like depicted on the right of slide 3. Usually cells from different time points are forming different clusters, no? Would appreciate some feedback on whether this makes sense). I've got some preliminary results on this new data where dpt outperforms L2 quite a bit. I will have the full results probably by Thursday.