* Description of the package
* Presumably we’re going to use Franze’s data. What about Tiffany’s?
* We will want to show annotated gene UMAP(s). Use of hoverable UMAPs seems to dictate using the top 2000 most variable genes.
* Speaking of UMAPs of genes, can we say anything about why all the UMAPs I’ve been making have such a strong visual resemblance and why they look so different from the first gene UMAP that Anna made? Cat’s Cradle exploits the duality between rows and columns of a matrix, in our case, the duality between cells and genes. Is there an argument for working with non-scaled data? Scaling is a little bit asymmetric as regards genes vs. cells.
* What can we say about gene clusters vs. other gene sets. I’m thinking here of pre-fab gene sets like Hallmark or GO, but also of other gene clustering techniques such as nichnetr or …?
* Cats’s Cradle of gene clusters vs. cell clusters. Do we have any insight as to the relationship between the Cat’s Cradle as revealed by the 2000 most variable genes vs. using all the genes?
* We’ve talked about a Cat’s Cradle between gene clusters and cell clusters. But it seems to me this implies a relationship between cell clusters, where two cell types are related if they both highly express a common gene cluster. This is one of a number of different ways we can think about the relationship between clusters. We’ve discussed a relationship between cell clusters (and we can do the same with gene clusters) that comes out of asking the total weight of the edges in the nearest neighbour graph that need to be cut to separate the two clusters. Perhaps this total weight should be normalized by the product of the sizes of the two clusters. The result could either be displayed using force-directed layout of the resulting graph, or again making a Cat’s Cradle, but here having both sides being the cell clusters.
* Attributes in the nearest neighbour graph. Suppose we are looking at the genes in a particular gene cluster. Some of these may be genes we know things about, others not. We would like to know more about these unknown genes. We can leverage the fact that genes have known individual attributes in a way that individual cells typically do not. To do this, we can look at a neighborhood of a given gene and ask what are the attributes of its neighbors. This can be done either by drawing a circle of a given size around this gene and querying its immediate neighbors or by querying the entire graph in a way that accords less weight to genes which are further away in the graph. One way of validating this proceedure is to see whether it rediscovers the attributes of known genes.