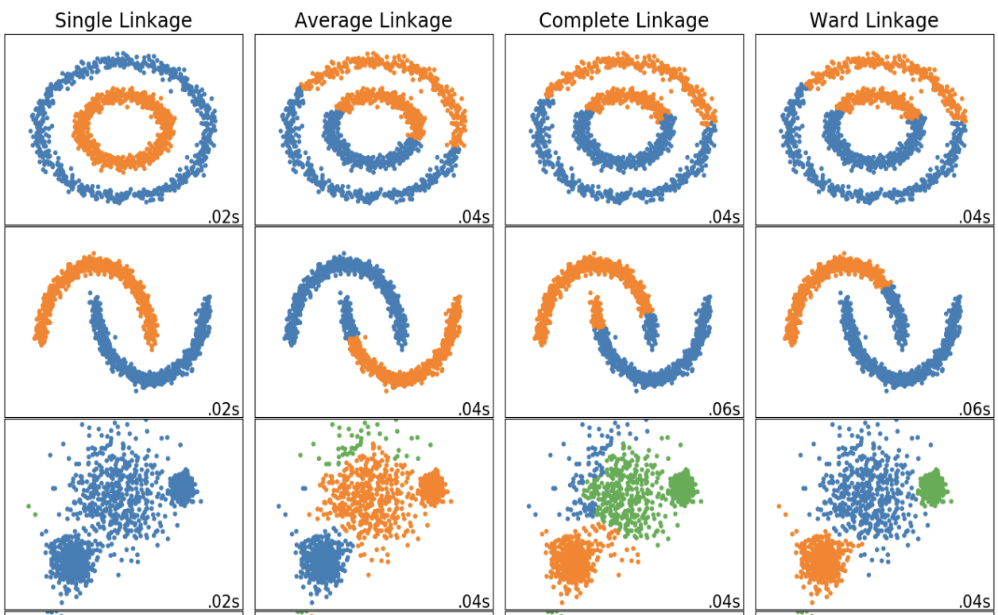
**MAA 507 SPRING 2022**

**SEMINAR 3 WRITTEN ASSIGNMENT**

1. Elif Cemre Durgut – Group 3 (Me, Felicia Modin, Nudzejma Pozder)
2. I was happy to work with my group. Both Felicia and Nudzejma were willing to work together. We searched online individually and then try to come up with new ideas together. Newman Girvan clustering was simpler to work with, but we struggled a bit about constructing a graph for mean linkage.
3. Our solutions for practical questions:
4. Our solution for complete linkage is to have a graph such that there are two main clusters, but they are very close to each other. So that single linkage would fail in this case.

For the single linkage, we thought of a graph that has a specific shape that consists of points that are very close to each other like an S shape. As can be seen in the image below, single linkage works better with specific shapes.

(Image by ScikitLearn)

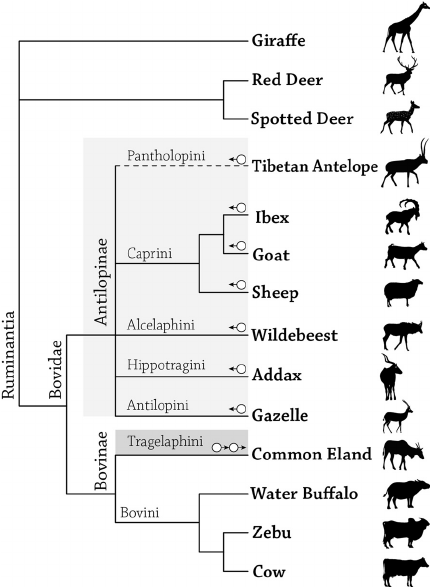
1. We could not find an example graph that is appropriate for mean linkage. Therefore, I will give the example that we discussed with the teacher. Two ellipse-like shapes that are close to each other:

Single linkage would not work here because the ellipses are too close to each other. And complete linkage would fail by clustering them as on top of each other.

1. We removed all the edges one by one starting with the one in the right middle. Then we get a square graph, a triangle graph, and another triangle graph with one extra node. Then we remove the edge between that triangle and the other node. Eventually, all edges are removed.
2. We found this information on the Internet, but we could not figure out a graph example ourselves.

Text

Description automatically generated(Slide by Kathy Macropol

 As we discussed during the seminar, parity is the problem here.

My reflection on discussion questions:

1. One example of such a dendrogram could be genes or species in biology where kinship/similarities between species are clearly visible.
2. Computational complexities of

NG 🡪 O(|V|3)

H 🡪 O(|V|2)

MC 🡪 O(|V|3)

1. Non-graph clustering problem could be any data points without connections. Our example was a dataset about age & income relations. The features(columns) are id, gender, age, monthly income. They are just points and they do not have connections between them to form an edge.

Graph clustering example could be any network data such as connections on LinkedIn, following relations on Instagram, friends on Facebook.

1. Extras:
2. If our data is sparse, this would decrease the computational complexity of the algorithms since most of the entries in the matrix will be zero. To speed up the algorithm, we can omit values that are smaller than a certain threshold as we do operations on the matrix.
3. We could not find time to discuss this as Group 3. We can try to find some connections between data points using a correlation matrix as we discussed at the end of the seminar.