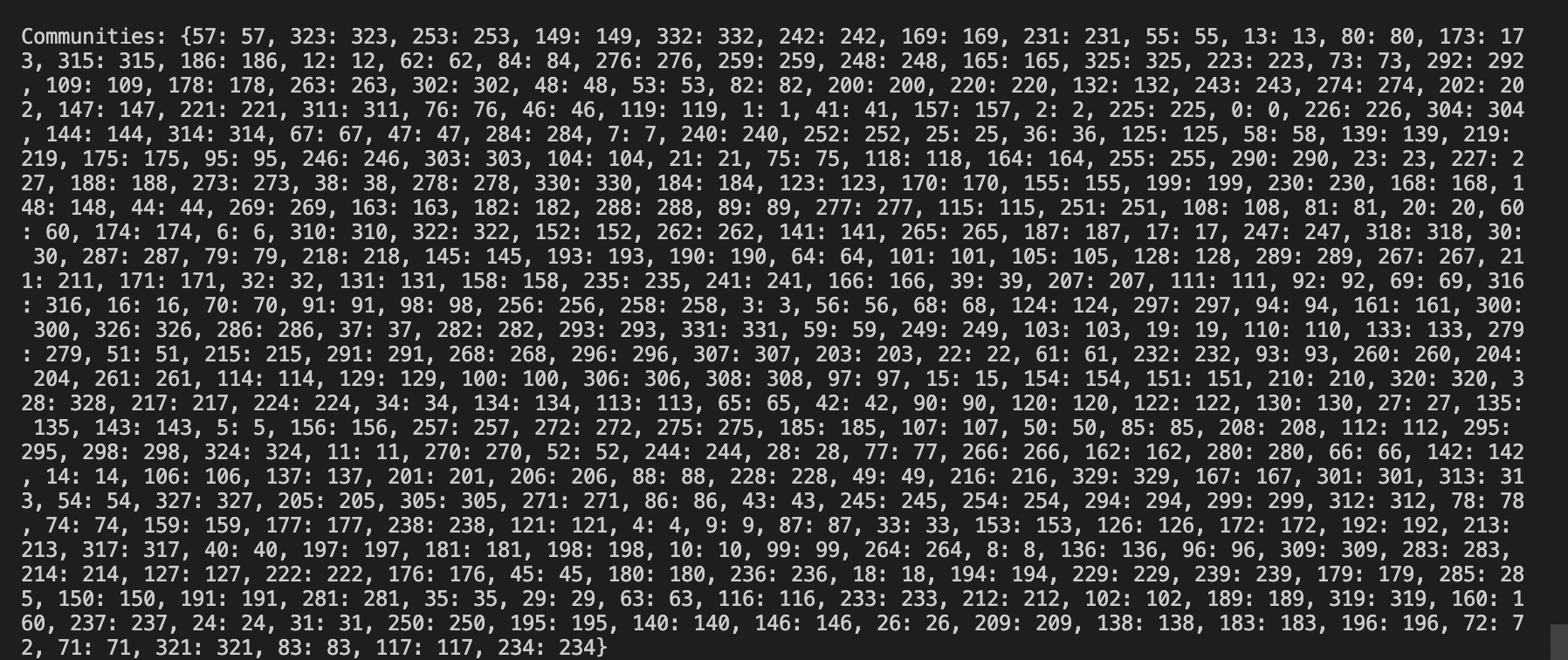
Aymen Tiguite

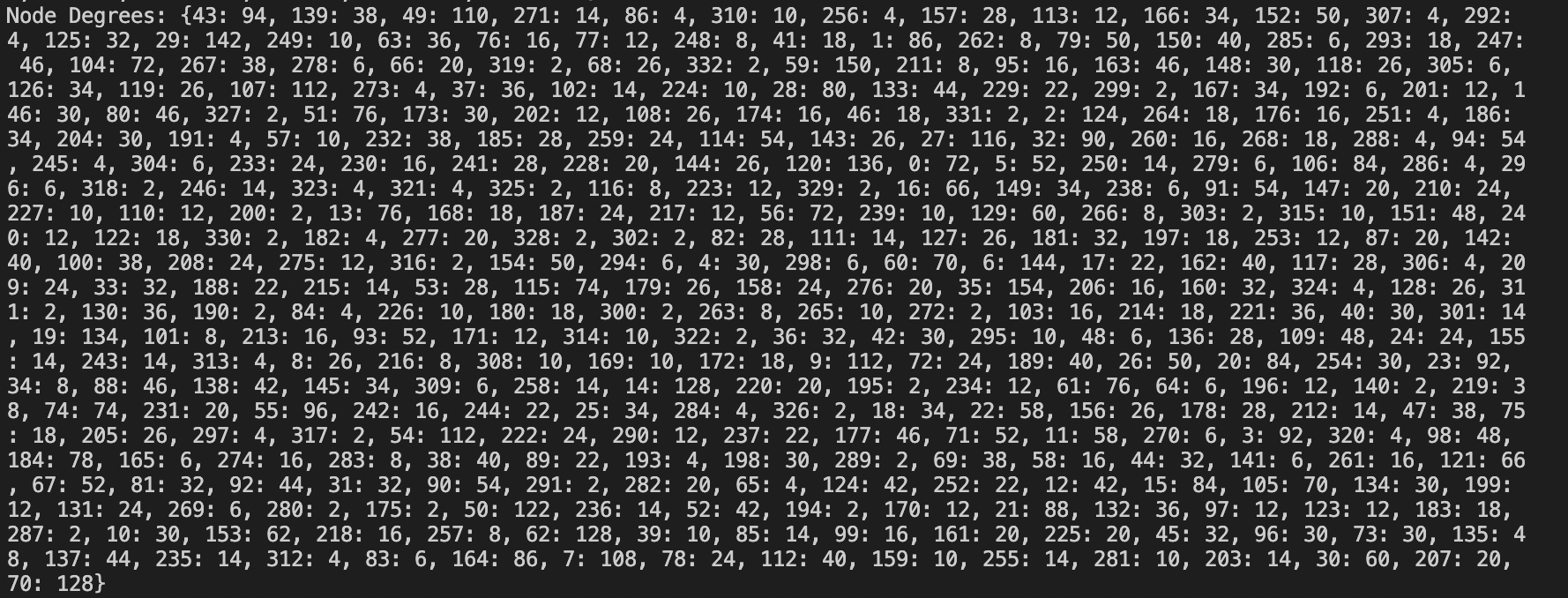
DS210

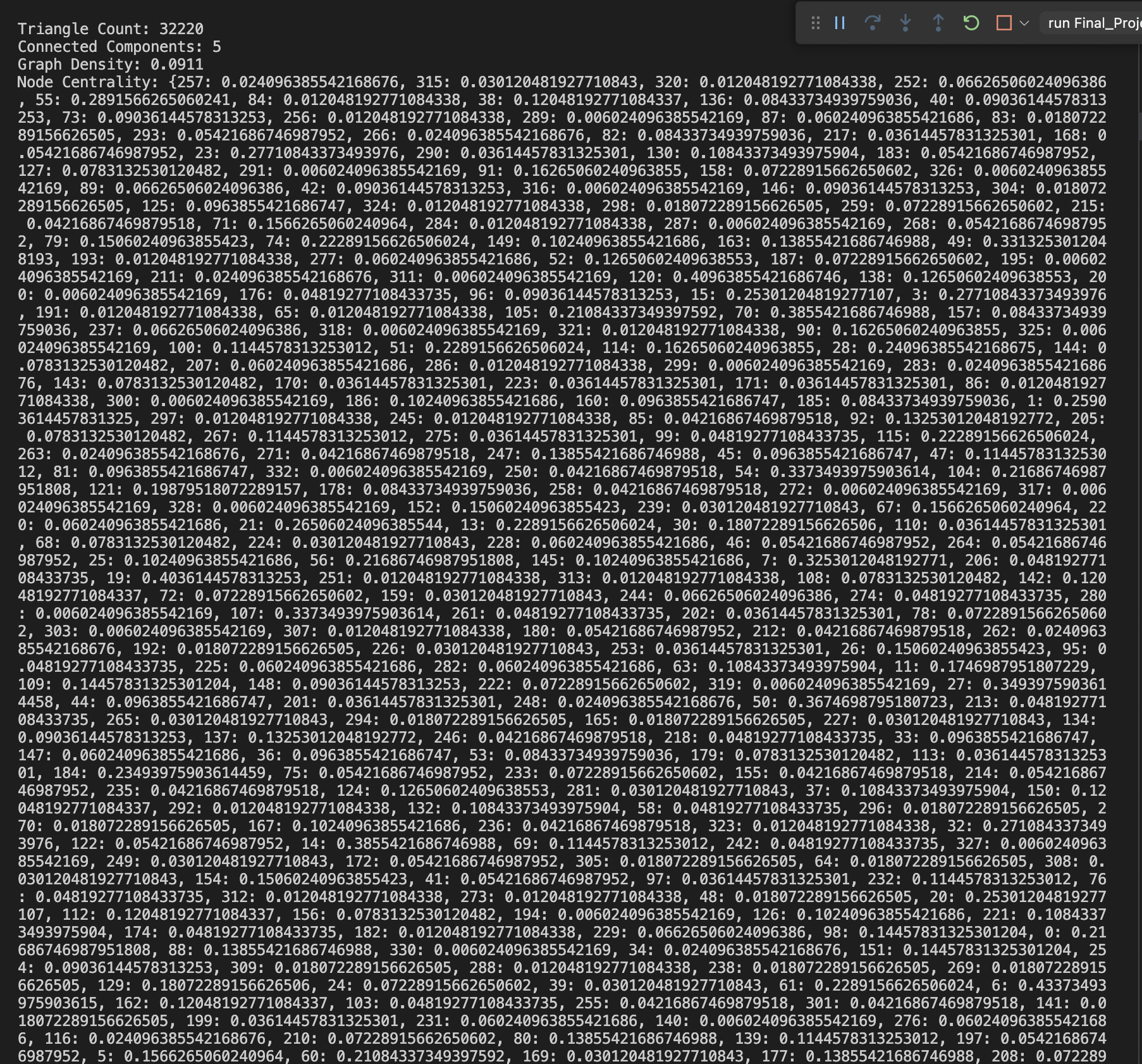
Professor Leonidas

December 15, 2023

The point of the project is to find the connections people make on facebook and analyze how these groups (circles) interact with one another. The program reads an edges file, facebook/0.edges, where each line is an edge linking two nodes (our social beings in the network). It's quite intriguing to see these connections come to life in the form of a graph, which we conveniently represent in DOT format for easy visualization. The “detect\_communities” function, currently in its nascent stage, assigns each node to its own community. It's a humble beginning, but it lays the groundwork for more sophisticated community detection algorithms that could unveil hidden clusters and social dynamics. We then dive into calculating the degree of each node. It's a simple yet effective way to spot the social butterflies of our network, or the nodes with the most connections.

The output “Communities: {...}” shows each node initially assigned to its own community, indicating a starting point for community detection. The “Node Degrees: {...}” part of the output gives us a count of connections for each node. For example, node 43 with 94 connections. Nodes with high degrees, like node 18 with 330 connections, could be pivotal in the network, potentially acting as “information hubs” or influencers with a huge following and mutuals that come with that following.



I was able to identify 5 connected components within the network, hinting at distinct sub-networks or isolated groups within the larger social structure. This insight is valuable for understanding the network's compartmentalization and the potential for information flow (or lack thereof) between different groups. The triangle count in a graph is a measure of the number of triangles that exist within the graph. A triangle is a subgraph consisting of three nodes and three edges, forming a closed loop. In social network analysis and various other fields, triangles have significant importance and provide insights into the structure and connectivity of a network. There are 32220 triangles meaning there have been 32,220 instances where three nodes in the dataset form closed loops or triangles The calculated graph density is 0.09113933210318753. This relatively low density suggests that our network, while having areas of intense interconnectivity, is overall quite sparse. This characteristic is often observed in large social networks, where numerous acquaintances (edges) are spread across a vast number of individuals (nodes). The “Node Centrality: {...}” output provides a centrality score for each node, highlighting their relative importance. Higher scores indicate nodes that are more central in the network's structure, possibly playing crucial roles in maintaining network cohesion or spreading information. The people with lower scores are most likely to not be on facebook as much or just don’t find the need to make connections on social media. 

This project has been a very interesting experience when looking at social interactions/networks as numbers and not as things that people experience. It also showed me the fun that comes with working with large datasets, which I am excited to do in the future.