

Project: Due May 2, 8pm

In the densest subgraph problem you are given an undirected graph $G = (V, E)$ and are asked to find a subset of vertices $V' \subseteq V$ that maximizes

$$\rho(V') = \frac{|E[V']|}{|V'|},$$

where $E[V']$ is the set of edges both of whose endpoints lie in V' . (Formally, $E' = \{(u, v) \in E : u, v \in V'\}$.)

For the project you are asked to come up with a MapReduce algorithm that finds the densest subgraph on a large graph.

Logistics

- You may (but are not required) to work in groups of two or three people
- There are three data files (two from the first homework), and a third at <http://coms6998.s3.amazonaws.com/data3.txt>
- We will maintain a leaderboard of the best answer found so far on each of the three datasets (together with the team name) and refresh it every 24 hours. To be on the leaderboard, send to the instructor and the TAs an email formatted as follows:

Subject: [Team Name]: Density [\rho]
Attachment: file.gz

The file should be formatted as follows: Line 1: team name. Line 2: claimed density, ρ . Then a vertex id for vertices in V' , one per line. For example, if team *Densest* found a subset with $\rho = 0.22$, it would submit:

```
Team Densest
0.22
123765
435690
...
987653
```

- With the final result, you will submit a writeup describing your approach
- No late homeworks will be accepted.