Data Intensive Computing - Review Questions 6

Daniele Montesi, Francesco Staccone

1. According to the greedy vertex-cut algorithm in PowerGraph, an edge (u, v) should be assigned to one of the machines from the vertex (either u or v) with the most unassigned edges, when A(u) intersection A(v) = empty. A(v) is the set of machines that vertex v spans. Explain why?

   The goal of partitioning in PowerGraph (i.e. using Vertex-cut partitioning) is to distribute the load over multiple machines in order to provide:

   - Balanced number of edges per machine
   - Avoid replication of vertices

   Greedy-algorithm works well in this case since it distributes the vertex load evenly whenever is possible. In the aforementioned scenario, the reason why the edge (u,v) is assigned to the machine whose vertex has the least edges means that we are balancing the partition loads. Differently, we would increase the chance of possible connection between that assigned vertex and another node, which is probably turning out into unbalanced partitions.

2. Assume we have two types of resources in the system, i.e., CPU and Memory. In total we have 28 CPU and 56GB RAM (e.g., 1 CPU = 2 GB). There are two users in the systems. User 1 needs 1CPU,2GB per task, and user 2 needs 1CPU,4GB per task. How do you share the resources fairly among these two users, considering the asset fairness and DRF.

   - Asset Fairness case:

     ```
     max(x,y)
     x+y<=28
     2x+4y<=56
     4x=6y
     ```

     Results:
     User 1: **x=12, <12 CPU, 24 GB> ; <43% CPU, 43% GB>**
     User 2: **y=8, <8 CPU, 32 GB> ; <28% CPU, 57% GB>**

   - DRF case:

     User 1 dominant resource: Equal share for CPU and RAM (1/28 = 2/56)
     User 2 dominant resource: RAM (1/28 < 4/56)

     ```
     max(x,y)
     x+y<=28
     2x+4y<=56
     1/28x=4/56y
     ```

     Results:
     User 1: **x=14, <14 CPU, 28 GB> ; <50% CPU, 50% GB>**
     User 2: **y=7, <7 CPU, 28 GB> ; <25% CPU, 50% GB>**