

Graph Mining Algorithms on Arabesque

Project Description:

If you look around, you will find that graphs are all around us. Social networks, road networks, the Web, user reviews, and even protein interactions in our bodies, can all be represented as graphs. One of the most common activities on graphs is discovering “interesting” patterns within the graphs. Which people are more likely to become friend in the future? Which road itineraries are more frequent? What are the items that you might be more likely to buy? Algorithms that give us deeper insights from graphs are called Graph Mining algorithms. They are one of the underlying data science technologies behind the magic we see everyday, for example, in social networks when we get friend suggestions or in advertising. With the current data deluge, graphs grow larger and larger, and we need scalable Big Data technologies to crunch them. If you find implementing scalable graph mining algorithms tempting, come and join our project!

Duties/Activities:

1. Understanding the programming API and the high-level architecture of the Arabesque framework, a distributed Big Data system
2. Understanding one or two graph mining algorithms from the literature
3. Implementing these algorithms on top of Arabesque

Required Skills:

- Mandatory skills: Java, Bash scripts, understanding of basic graph concepts
- Preferable skills: Scala, Jupyter notebook awareness is a plus

Expected Team Size:

1 or 2 *junior or senior students*

Mentors

Name: Abdurrahman Ghanem

email: abghanem@qf.org.qa

Name: Marco Serafini

email: mserafini@qf.org.qa