Abstract: Nowadays, emerging distributed technologies enable the scientific community to perform large-scale simulations at a rate never seen before. The pressure those systems put on the scientists is twofold. First, they need to analyze the massive amount of data generated as a consequence of those computations. Second, scientists need to make sure they achieve meaningful scientific conclusions with the available resources, oftentimes by changing the course of an experiment at run-time. The first challenge implies the need of new and more efficient clustering and classification techniques that require at most linear time with respect to the amount of data generated. While the second challenge needs algorithms able to build knowledge from the data and make decisions on the fly, in a time-sensitive scenario.

In this talk I will present scalable algorithms that address both challenges; the first one in the context of a high-throughput protein-ligand docking application, and the second in the context of a Volunteer Computing system.

Following the seminar, please join us for refreshments in the Computer Science Conference Room (FLN 4.01.20) at 11:30am.