# Fast and Scalable Combinatorial Algorithms for Data Analytics (FASCADA)

## PI: [Assefaw Gebremedhin](http://www.eecs.wsu.edu/~assefaw/)

School of Electrical Engineering and Computer Science

Washington State University

## Motivation

We are in an age when massive digital data continues to be collected at an extraordinarily rapid rate and with high and growing complexity (and concomitant uncertainty), when the data and the actors behind it are increasingly interconnected, and when architectures of computing platforms continue to rapidly change. These circumstances (data richness and complexity; interconnectedness; computing platform advances) collectively afford us with an unprecedented potential for discovery and innovation, but they also pose tremendous challenges for the design and implementation of algorithms that are up to the task of the required data analysis. Fast, robust and scalable algorithms are acutely needed for the purpose of analyzing massive datasets and extracting knowledge and insight from the data. For many problems, such algorithms do not exist, and when they exist, they are a poor match to the task and need fundamental re-examination.

## Research Goals

This project, named FASCADA, targets a segment in this grand challenge that spans a very wide spectrum. FASCADA’s overarching goal is to explore the interplay between *graph* and *matrix* algorithms in order to develop methods for data analytics that perform at scale on contemporary platforms, with a primary focus on data that are expressed in terms of networks. Its specific aims are organized under four intertwining areas:

1. *Enabling Scalable Data Analytics*.

Devise novel “problem-partitioning” methods that are useful for solving, at scale, optimization problems underlying a large class of machine learning algorithms.

1. *Network Analysis*.

Develop fast algorithms for discovering -- and analyzing -- dense sub-graphs in real-world networks arising from diverse domains.

1. *High Performance Computing.*

Develop effective paradigms for the parallelization of inherently sequential graph algorithms targeting many-core architectures.

1. *Algorithmic Differentiation (AD).*

Advance AD as a technology by designing better graph-based algorithms for Hessian computation, and use AD in emerging applications, including quantifying uncertainty.

## Integrated Education and Outreach Component

Two new innovative courses, an undergraduate course on Data Science and a graduate course on Network Science, will be developed and taught as part of the FASCADA project. Students at both the graduate and undergraduate level will be engaged in the research, and they will be trained in computational modeling, algorithm design, and software development. The project will also reach out to underrepresented minority groups in computing sciences and engineering, to increase their participation, via two existing mechanisms in the Pacific Northwest (the Louis Stokes Alliance for Minority Participation (LSAMP) initiative and a partnership with Heritage University).



An overview of the Research and Education plan in FASCADA.

## Expected Significance and Broader Impact

A common thread that runs through all four of the research Aims of FASCADA is a focus on graph problems and their solution. Graph algorithms have long played a crucial role not only in computer science, but also in scientific computing and the engineering disciplines. Indeed, the area of *combinatorial scientific computing* is founded on the core idea of effectively applying graph and other combinatorial algorithms in scientific computing, where the computations predominantly involve matrices. This perspective has led to much progress in the development of a) efficient methods that exploit structural properties of problems and b) tools that enable applications make better use of high performance computing. In the other direction, there is a range of examples in which matrix algorithms are the engines behind network (graph) computations. Examples include: Google's PageRank algorithm, random walks on graphs, graph partitioning, socio-metric analysis, and many more.

The work to be carried out in FASCADA seeks to traverse *both* of these directions (*matrix* to *graph* and *graph* to *matrix)* in the context of the targeted problems. It strives to advance fundamental knowledge at the intersection of a range of areas, including data science, computational science and engineering, computational mathematics, and high performance computing. Algorithmic progress made through the project will be realized in software implementations, will be integrated with existing software tools when applicable, and will be made available to the wider community as open-source software. Research results will be widely disseminated to the scientific community through publications.