HAL open science

Trends in Combinatorial Analysis: Complex Data, Machine Learning, and High-Performance Computing

Ariful Azad, Bora Uçar, Alex Pothen

To cite this version:

Ariful Azad, Bora Uçar, Alex Pothen. Trends in Combinatorial Analysis: Complex Data, Machine Learning, and High-Performance Computing. 2019, pp.1-3. hal-02304457v2

HAL Id: hal-02304457

https://inria.hal.science/hal-02304457v2

Submitted on 17 Oct 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Trends in Combinatorial Analysis: Complex Data, Machine Learning, and High-Performance Computing

By Ariful Azad, Bora Uçar, and Alex Pothen
(For the original document see http://bit.ly/sinews-03-sep-2019)

Discrete algorithms and combinatorial analysis were well represented at the 2019 SIAM Conference on Computational Science and Engineering (CSE19), which took place earlier this year in Spokane, Wash. Over 15 talks—dispersed among five minisymposia—covered topics ranging from graph algorithms and machine learning to scientific computing. Like many other communities in applied mathematics and computational science, the SIAM Activity Group on Applied and Computational Discrete Algorithms (SIAG/ACDA)—founded this spring—is innovating techniques to learn from large-scale combinatorial data. Challenges that accompany learning from data include increasing data volumes and complexity, the need for novel algorithms, and efficient employment of high-performance computing resources. Researchers at CSE19 presented ongoing work that addressed these challenges and captured the following four trends.

Scientific Data and Combinatorial Representation

Several presenters spoke about combinatorial models they have developed to represent scientific datasets, and algorithms they have designed for efficient solution of the aforementioned problems. Possible application areas include astrophysics, biology, agriculture, and neuroscience. In some cases, researchers used complex networks to model the data. For example, Ariful Azad (Indiana University Bloomington) utilized correlated brain segments, as measured by functional magnetic resonance imaging (fMRI), to create a brain connectivity network. Ananth Kalyanaraman (Washington State University) analyzed agricultural phenomics data via persistent homology techniques. And Francesca Arrigo (University of Strathclyde) modeled complex datasets as multilayer networks, where different layers of each network capture various features or modalities. Overall, members of SIAG/ACDA are actively finding innovative ways to model increasingly complex scientific and business datasets.

Algorithmic Innovation

After representing data with a sound mathematical model, the next step often involves developing algorithms to solve problems of scientific interest. New scientific and business challenges lead to the creation of novel, efficient algorithms in the SIAG/ACDA community. For example, Arrigo also discussed an eigenvector-based centrality measure in multilayer networks capable of calculating centrality in a network with several different types of interactions among various entities. Syed M. Ferdous (Purdue University) described the design of parallel approximation algorithms for computing edge covers using the primal dual linear programming framework, with applications to semi-supervised classification.

Machine Learning

Many presenters at CSE19 focused on machine learning methods and algorithms. In fact, a minisymposium entitled “The Intersection of Graph Algorithms and Machine Learning” featured several talks on clustering and learning from discrete datasets and networks. For instance, Nesreen Ahmed (Intel Research Labs) introduced an algorithm that learns useful graph representations from diverse datasets. Azad spoke about his group’s search for a consensus clustering among segmented brain fMRI images (see Figure 1). Several other researchers addressed the application of deep neural networks to the solution of challenging problems in data analytics.

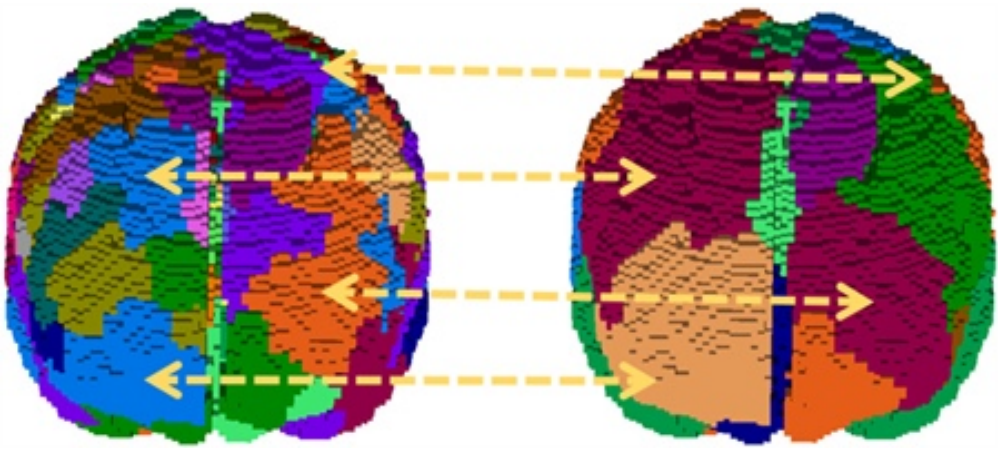

Figure 1. Parcellation of the brain using functional magnetic resonance imaging (fMRI) data and a clustering algorithm. The left side depicts large communities and the right side depicts small communities. Figure courtesy of Ariful Azad.

High-performance Computing

With ever-increasing amounts of data and the availability of parallel computers, parallel algorithms and high-performance computing are trending themes in high-end data analytics that involve the SIAG/ACDA community. A two-part minisymposium at CSE19 entitled “Graph and Combinatorial Algorithms for Enabling Exascale Applications” was dedicated to massively-parallel graph and combinatorial algorithms. Arif Khan (Pacific Northwest National Laboratory) exhibited a parallel approximation algorithm to solve a data privacy problem called adaptive anonymity for a terabyte-scale healthcare dataset on a computer at Lawrence Berkeley National Laboratory’s National Energy Research Scientific Computing Center. Mostofa Patwary (NVIDIA) demonstrated a scalable clustering algorithm called BD-CATS that can cluster 1.4 trillion cosmological particles using 100,000 cores on a Cray supercomputer. Speakers across various minisymposia introduced parallel combinatorial algorithms for multicore processors, graphics processing units, distributed supercomputers, and cloud infrastructures.

We are looking forward to more talks of this flavor at the SIAM Workshop on Combinatorial Scientific Computing, to be held in Seattle, Wash., from Feb 11-13, 2020, as well as the first SIAM Conference on Applied and Computational Discrete Algorithms in 2021. SIAG/

ACDA invites new and existing SIAM members with interest in designing “algorithms in the real world” that solve application-motivated problems to join us.

Ariful Azad is an assistant professor of intelligent systems engineering at Indiana University Bloomington. He received his Ph.D. from Purdue University and was a Research Staff Member at Lawrence Berkeley National Laboratory. Bora Uçar is a CNRS researcher at the Laboratoire de l’Informatique du Parallélism, École normale supérieure in Lyon, France. He received his Ph.D. from Bilkent University in Ankara, Turkey, and has worked at the Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique in Toulouse, France. Alex Pothen is a professor of computer science at Purdue University. He is chair of the SIAM Activity Group on Applied and Computational Discrete Algorithms.