Graph Analytics

Graph analytics is the study and analysis of data that can be transformed into a graph representation consisting of nodes and links. The problem of graph analytics is among the most important research and development (R&D) areas within the visual analytics community. Researchers with diverse backgrounds from information visualization, human-computer interaction, computer graphics, graph drawing, and data mining have pursued graph analytics R&D from scientific, technical, and social approaches. The transdisciplinary nature of the area separates itself from the other graph studies such as 1) graph drawing, which studies optimal topological layout, and 2) graph mining, which champions knowledge discovery though algorithmic computation.

Scientists at the Pacific Northwest National Laboratory (PNNL) have been actively involved in graph analytics R&D in social network, cyber communication, electric power grid, critical infrastructure, bio-informatics, and earth sciences applications. The mission of graph analytics research at PNNL is not merely about research per se; it has the essential and enduring purpose of producing pragmatic working solutions that meet real-life challenges. This article gives a glimpse of some of the ongoing work in different application domains.

Foundational Graph Analytics Technologies

We have developed a series of cutting-edge graph analytics technologies to explore and analyze graphs with different sizes and complexities. For the exploration of small world graphs such as a social network, we developed the concept of a graph signature [1] that extracts the local features of a graph node or set of nodes and used it to supplement the exploration of a complicated graph filled with hidden features. For larger graphs with about one million nodes, we further developed the concept of a multi-resolution, middle-out, cross-zooming technique [2][3] that allows users to interactively explore their graphs on a common desktop computer. Currently, we are developing the concept of an extreme-scale graph analytics pipeline designed to handle graphs with hundreds of millions of nodes and tens of billions of links. Much of our work developed at PNNL has been loosely integrated into a graph analytics library know as Have Green [4].

GreenGrid, a power grid analytics tool, explores issues and areas of vulnerability in the western states power grid.
Domain Applications

We have used the Have Green library to develop a number of domain applications. For social network analytics, we explore the problems of co-citation networks, cyber and telecommunication networks, and criminal social networks. The graph data of these domain applications usually exhibit the so-called small-world properties with large clustering combined with small node separation (short characteristic path).

For critical infrastructure and utility graphs such as a power grid, we have investigated a number of problems from network vulnerability detection, contingency analysis, to situation awareness. The graph data of these domain applications usually are near planar. Some of them, such as the western North American Power Grid, also exhibit small-world graph properties.

More recently, we have been working on graph data transformed from scientific data modeling such as climate simulations. Cutting-edge scientific modeling usually requires extreme high-resolution simulation, which can potentially generate extreme-scale graphs with sizes in the range of terabytes to petabytes. Our extreme-scale graph analytics work is being conducted on high-performance computing platforms from Cray XMT to Cray XT4. Visual scalability is as important as computation scalability in these domain applications.

Unique Strengths and Capabilities

Our primary goal in graph analytics R&D is to develop pragmatic technology for real users and real applications. All our work has been co-developed with domain experts, analysts, and operators. All of our software is developed at PNNL, and no third-party commercial software license is required to use the software. Because we have full control of the software development cycle, we can customize our tools using different languages (such as C/C++, Java, C#, and python) and graphic libraries (such as OpenGL and Direct X) running on different platforms (including Windows, Windows CE, and Unix/Linux). With the exception of a few ongoing developments, all of our systems and technologies have been reported in peer-reviewed journals and conference proceedings.

For more information about the research described here, see http://nvac.pnl.gov

Team Members

Pacific Northwest National Laboratory: George Chin, Harlan Foote, Steve Libby, Patrick Mackey, Jim Thomas, and Pak Chung Wong. (Others include Daniel Chavarria, Yousu Chen, John Feo, David Haglin, Zhenyu Huang, Shuangshuang Jin, John Johnson, Ruby Leung, and Bryan Olsen.)

Point of Contact

Pak Chung Wong, Chief Scientist, PNNL, pak.wong@pnl.gov

References


