Universal Information Graphs

Benefit: *Tracking evolving patterns and communities of interest in social networks is a central challenge in intelligence analysis. Our project is developing the algorithmic and computational infrastructure to support analysis of dynamic, and sometimes massive, networks.*

Mission

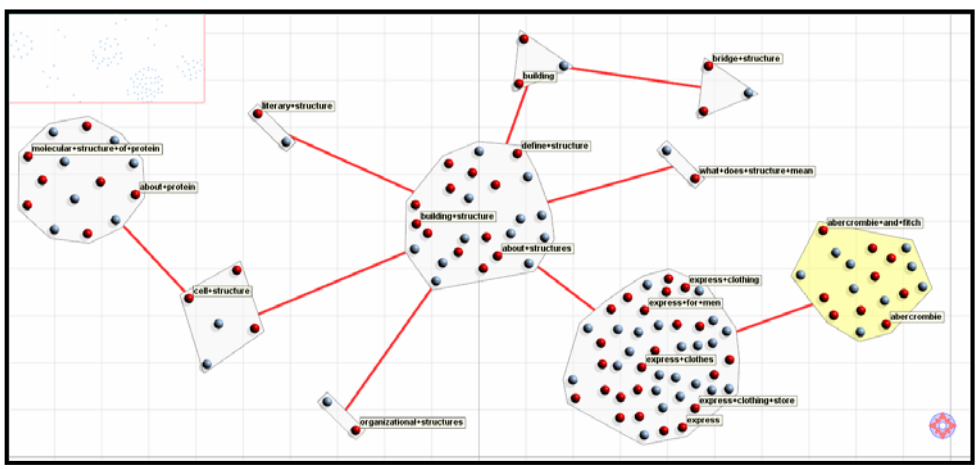
Many pattern detection applications require analysis and visualization of massive graphs that are obtained by fusing information from multiple sources. Universal Information Graphs (UIGs) are a tool for analysts that are the result of fusing a collection of graphs on separate vertex sets. As an example we can consider a person having unique characteristics in their separate email, telephone and web graphs. The universal graph must contain a fused view of these separate local neighborhoods.

Our project is building infrastructure and algorithms to support accessing, partitioning, navigating, manipulating, and visualizing massive weighted graphs obtained as projections of a universal information graph. As part of this work, we have built a computational infrastructure to support research on specialized algorithms to efficiently uncover and classify patterns in massive graphs that are too large to fit within a computer's main memory. Our newly built External Memory Algorithms (EMA) cluster enables researchers to develop and test fundamental algorithms for visualizing large graphs connecting entities of interest (such as people, organizations, places, events, documents, etc.) and rapidly identifying patterns in such graphs.

Our project marries notions from relational databases to notions from graph theory in order to develop new record similarity measures to enable better identification of groups. These database similarity measures are based on what we call a "stratified semantic dot product" which is a semantic generalization of the usual binary dot product that encodes the size of the intersection between two sets. We have implemented an external memory algorithm to compute stratified semantic dot products between the records of a relational database, and we have implemented a method that is able to process a semantic dot product graph with up to a billion edges in about an hour.

Collaborator(s):

• *Lawrence Livermore National Laboratory*



This picture shows a set of macro-categories and the connections between them that are derived from a network of search queries. The underlying data contains on the order of a billion edges and is pre-processed by our current hardware-software infrastructure to yield more manageable representations for query processing.

Funded by: *US Department of Homeland Security and Lawrence Livermore National Laboratory*

Early Development — Lab Prototype — Commercial Product

For more information, contact:
James Abello, abello@dimacs.rutgers.edu
(dydan.rutgers.edu)