Unsupervised Discovery in Hypergraphs Using Tensor Decompositions

Hypergraphs as Tensors

Tensors = Multi-Dimensional Arrays
- Graphs represented with adjacency matrix
- Hypergraphs represented with adjacency tensor

Tensor Decomposition

Hypergraph Adjacency Tensor

ETTB++ Contributions (Patent Pending)
- ETTB++ builds upon Sandia’s C++ Tensor Toolbox
- New sparse tensor formats
- Runtime optimizations and load balancing scheme featuring mixed static/dynamic scheduling
- Compiler optimizations to
  - Increase parallelism
  - Improve data locality
  - Reduce global barrier synchronization and communication
- Data locality and low global communication keys for seamless mapping to distributed systems

ENSIGN Tensor Toolbox (ETTB++)

ENSIGN Tensor Format

Performance Results

Baskaran et al., HPEC 2014

Reservoir Network Traffic Case Study

The CANDECOMP/PARAFAC (CP) Decomposition
- Similar to a higher-order SVD
- Breaks a tensor into a weighted sum of “components,” each representing a discovered feature of the data
- Each component includes one vector for the indices of each dimension that indicates how involved the indices are in the feature
- For the sample tensor above, each component will have a sender vector, receiver vector, and a key term vector

Baskaran et al., HPEC 2012

ENSIGN Tensor Format Performance Results

- Time: 9 am to 5 pm every day
  - Senders: Intern’s computers
  - Receivers: DNS servers
  - Requests: Google, Stack Overflow, various yale.edu websites

- Time: Constant and regular
  - Senders: Two business computers
  - Receiver: Broadcast address
  - Request: Faulty printer’s address (Request denied so repeated constantly)

- Time: Between 2 am and 5 am in the morning
  - Senders: Blacklisted Chinese IP addresses
  - Receiver: Reservoir code repository
  - Access denied

M. Baskaran, J. Ezick, E. Papenhausen, D. Bruns-Smith, B. Meister, R. Lethin