Parallel Software Tools for the Construction and Analysis of Complex Networks

LINC Workshop 4 – Montevideo
24-26/March/2014
Outline

- Introduction
- Computational challenges
- Parallel Software tools
  - Network construction
  - Network analysis
- Software architecture
- Usage
Introduction – large-scale data analysis

Social Networks
Application: e.g., identifying communities, information spread modeling

Bioinformatics
Application: e.g., identifying drug target proteins

Climate
Application: e.g., identify patterns? analyze spatio-temporal interaction of climate variables

➡ Sources of data: simulations, experimental devices, the Internet, sensor networks
➡ Challenges: data size, heterogeneity, uncertainty, data quality, computational time
Introduction – from domain-specific to computation

**Application**
- Social Network Analysis
- WWW
- Computational Biology
- Scientific Computing
- Climate Research

**Problems**
- Community detection, central entities
- Metabolic pathways, gene regulation
- Graph partitioning, coloring, matching
- Community detection, teleconnections
- Marketing, social search

**Graph Algorithms**
- Traversals, shortest paths
- Centrality measures
- Connectivity
- Community detection

**Architecture**
- Single processing unit
- Parallel machines
  - GPUs
  - x86 multicore servers
  - Massively multithreaded clusters
  - Multicore clusters
  - Distributed memory clusters
  - Clouds
Introduction – from domain-specific to computation

Input data

Graph kernel

Which algorithm?

factors.....

Network

find ..

- paths
- clusters
- partitions
- matchings
- patterns
- orderings

• traversal
• shortest path algorithms
• flow algorithms
• spanning tree algorithms
• topological sort

• graph sparsity
• static/dynamic nature
• weighted/unweighted, weight distribution
• vertex degree distribution
• directed/undirected
• simple/multi/hyper graph
• problem size
• granularity of computation at nodes/edges
• domain-specific characteristics

Computing architecture
Introduction – computational challenges

๏ degree
๏ centrality
๏ traversals
๏ . . . .
Hard “drive”
**Introduction – computational challenges**

- **SLOW**
- **BIG**

- degree
- centrality
- traversals
- . . . .
Introduction – computational challenges

- parallel reading?
- data reduction/compression?
- parallel tools to analyze complex networks?
Parallel tools to analyze complex networks

Pigraph
Pigraph – parallel library for graph analysis

igraph

- GNU R
- Python

Graph operations (C library)
Graph algorithms

Pigraph

- GNU
- Python

Graph operations (C library)
Parallel Graph algorithms

➡ OpenMP
➡ Shared memory platforms
Pigraph – example: “shortest path”: Floyd-Warshall algorithm

\[
D(0) = \begin{bmatrix}
0 & 3 & 8 & \infty & -4 \\
\infty & 0 & \infty & 1 & 7 \\
\infty & 4 & 0 & \infty & \infty \\
2 & \infty & -5 & 0 & \infty \\
\end{bmatrix}
\]

\[
D = \begin{bmatrix}
0 & 3 & 8 & \infty & -4 \\
\infty & 0 & \infty & 1 & 7 \\
\infty & 4 & 0 & \infty & \infty \\
2 & 5 & -5 & 0 & -2 \\
\end{bmatrix}
\]

For \(k=1\) to \(n\) {
    Parallel
    For \(i=1\) to \(n\) {
        Parallel
        For \(j=1\) to \(n\)
            \(D[i,j] = \min(D[i,j], D[i,k] + D[k,j])\)
        
    }
}

(a) (b)
Data reduction / compression

CSR
Compressed sparse matrices
Compressed sparse matrices

Compressed Sparse Row (CSR)

$\begin{array}{c}
\begin{bmatrix}
0 & 0 & 1 & 7 & 0 & 0 \\
0 & 2 & 8 & 0 & 0 & 6 & 0 \\
5 & 0 & 3 & 9 & 0 & 6 & 0 \\
0 & 6 & 0 & 4 & 0 & 5 & 3 & 9 \\
1 & 7 & 2 & 8 & 5 & 3 & 9 & 6 & 4 \\
\end{bmatrix}
\end{array}$

$\begin{array}{cc}
\text{Matrix} & \# \text{entries} \\
\hline
\text{row offsets} \quad 0 & 2 & 4 & 7 & 9 \\
\text{column indices} \quad 0 & 1 & 1 & 2 & 0 & 2 & 3 & 1 & 3 \\
\text{values} \quad 1 & 7 & 2 & 8 & 5 & 3 & 9 & 6 & 4 \\
\end{array}$

$N^2$  
$2 \text{ nnz } + N + 1$
Parallel reading

NetCDF-4
(HDF5, MPI/IO)
Parallel I/O
Parallel software tools for the construction and analysis of complex networks

parallel I/O

parallel (MPI)

Pigraph (openMP)

CSR compressed

val 10 -2 3 9 3 7 8 7 3 ... 0 13 4 2 -1
col_ind 1 5 1 2 6 2 3 4 1 ... 5 6 2 5 0
row_ptr 1 3 6 9 13 17 20
Parallel software tools for the construction and analysis of complex networks

- User interface
  - config file
  - select data
  - threshold (in case)
  - graph metrics
  - # of computing elements

Parallel File System

- parallel I/O
- calculation of Correlations

- Parallel MPI
- R
- Python

- Graph operations (C library)
- Parallel Graph algorithms

- Pigraph
- interface

- CSR
- [Matrix with values: 1 0.81 0.41 0.922 0.01 0.41; 0.81 1 0.9 0.01 0.9 0.43; 0.41 0.92 1 0.37 0.92 0.322; 0.922 0.01 0.37 1 0.01 0.41]
Thanks!