Pregelix: Dataflow-Based Big Graph Analytics
Yingyi Bu
Department of Computer Science, UC Irvine

Dataflow Approach

Our philosophy
Stop building one-off systems like Pregel, GraphLab, and Giraph, instead, build them on a data-flow engine!

Big Graph Analytics!

Think Like a Vertex
- Receive messages
- Update states
- Send messages

public class PageRankVertex extends Vertex<VLongWritable, DoubleWritable, FloatWritable, DoubleWritable> {

    @Override
    public void compute(Iterator<DoubleWritable> msgIterator) {
        sum = 0;
        while (msgIterator.hasNext()) {
            sum += msgIterator.next().get();
            setVertexValue((0.15 / getNumVertices()) + 0.85 * sum);
            sendMsgToAllNeighbors(vertexValue / getEdges().size());
        }
    }
}

Conclusions
- Vertex-oriented programming model is simple
- Dataflow implementation is neat and efficient
- We target Pregelix to be an open-source production system.

Sponsors:

- Machines: Yahoo! Research cluster ~ 180 8 core/12GB memory/4 disk machines.
- Dataset: Yahoo! AltaVista web graph (1,413,511,393 vertex, adjacency list, ~70GB)

http://hyracks.org/projects/pregelix/