EXECUTIVE SUMMARY

- Popular parallel abstractions like MapReduce and Pregel do not efficiently express many machine learning algorithms:
  - MapReduce: does not express graph computation
  - Pregel: does not express asynchronous computation
- To fill this critical void we introduced GraphLab, a new abstraction targeted at advanced machine learning algorithms and capable of expressing adaptive asynchronous graph computation

THE GRAPHLAB ABSTRACTION

- Running Example: PageRank
  - Rank of a page is a weighted sum of the rank of neighbor pages:
    $$ R[i] = \alpha + (1 - \alpha) \sum_{(j,i) \in E} \frac{1}{L[j]} R[j] $$
  - $\alpha$ is the random reset probability
  - $L[j]$ is the number of links on page $j$
- Abstraction:
  - Data Graph
  - A graph with arbitrary data associated with each vertex and edge.
  - Update Function
    - A user defined program transforms the data in the scope of a vertex.
      ```
      struct pagerank : public update<vertex, pagerank> { double sum = 0;
      void operator_((context_type & context) { double old_rank = context.vertex_data(edge.source()).rank;
      context.num_out_edges();
      double residual = abs(old_rank - old_rank) / context.num_out_edges();
      if(abs(residual) > EPSILON) {
        context.reschedule_out_neighbors(pagerank());
      }
      }
      }
      ```
- Consistency Models
  - Ensures serializability:
    - Data Graph
    - Edge Consistency Model
    - Vertex Consistency Model

THE CHALLENGE OF NATURAL GRAPHS

- Natural Graphs:
  - Yahoo! Web Graph
  - Top 1% vertices is adjacent to 33% of the edges!

- The problem with high degree vertices
  - Touch a Large Amount of State
  - Requires Heavy Locking
  - Processed Sequentially

NEW FEATURE IN GRAPHLAB2

- Factorized Update Functors:
  ```
  struct pagerank : public update<graph, pagerank> { double delta = 0;
  void pagerank(const context_type & ctx) { delta += ctx.in_edges().size();
  ctx.schedule(edge.target(), delta);
  }
  ```

RESULTS

PageRank on 25.5M Vertex 355M Edge Web Graph

- Algorithm Comparison
  - Dynamic
  - Alpha
  - Async

- Abstraction Comparison
  - Delta
  - Factorized
  - GraphLab

Distributed PageRank

- GraphLab