Error and Sensitivity Analysis for Graphs – Jason Riedy, GT

Two kinds of errors out there...

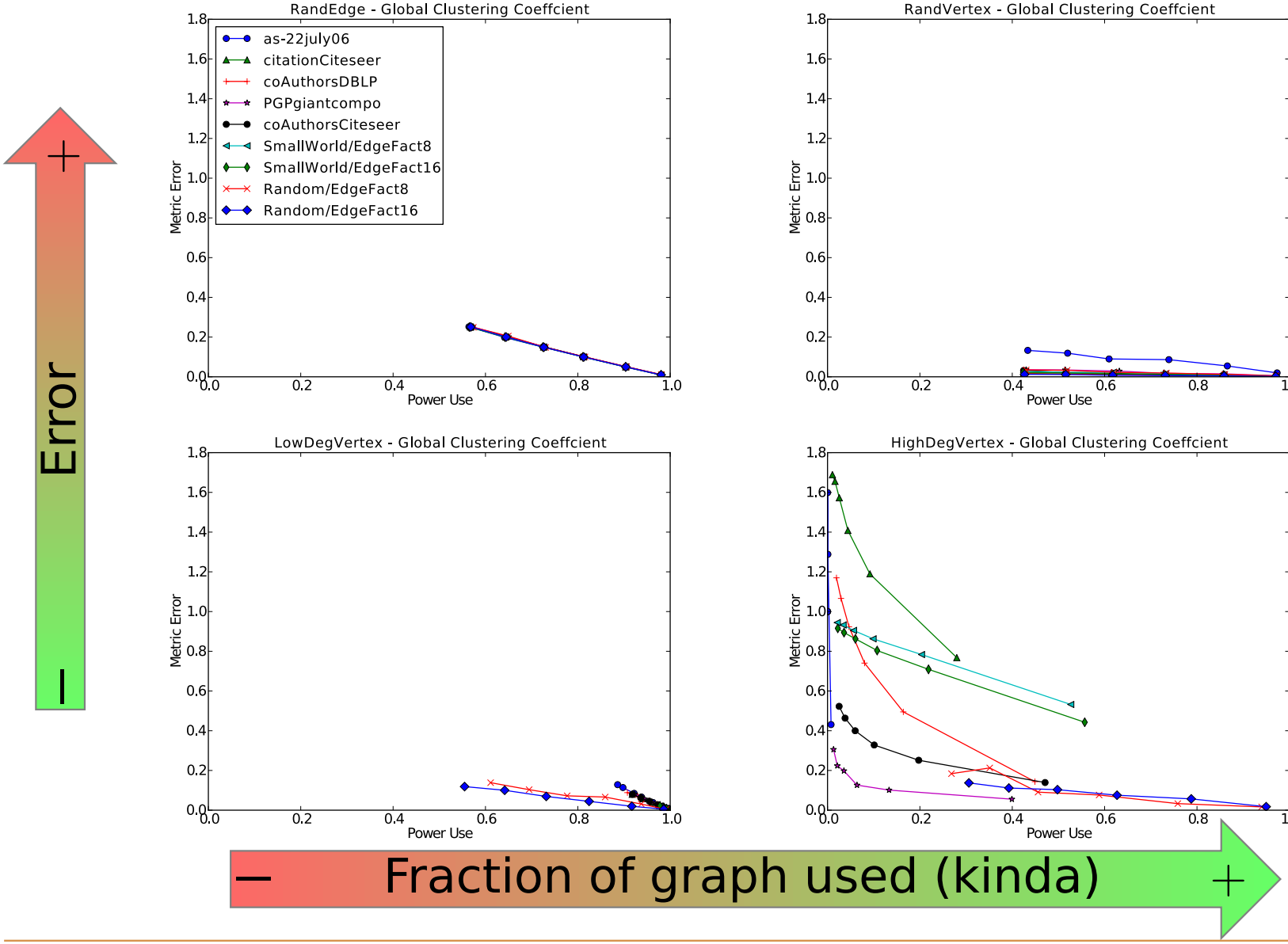
- Graphs imperfectly represent some real phenomenon.
  - Friendship: see LinkedIn
  - Health data: see privacy
- Computation imperfectly analyzes the graph.
  - Data may be “sampled” (aka dropped, lost) for energy...
  - Plain old computational error, bugs

Challenge: Quantify and Analyze Errors in Graphs

- Something that happens once in a billion times will pop up in large graphs...
- Except in limited cases, we don’t know what we’re doing.

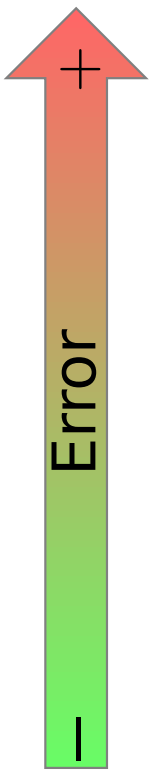
Quick Example: Global Clustering Coefficient

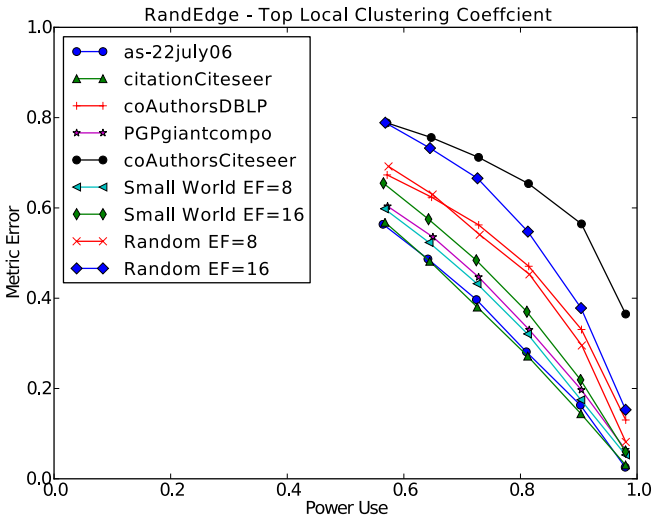
From Zakrzewska & Bader, "Measuring the Sensitivity of Graph Metrics to Missing Data," PPAM 2013

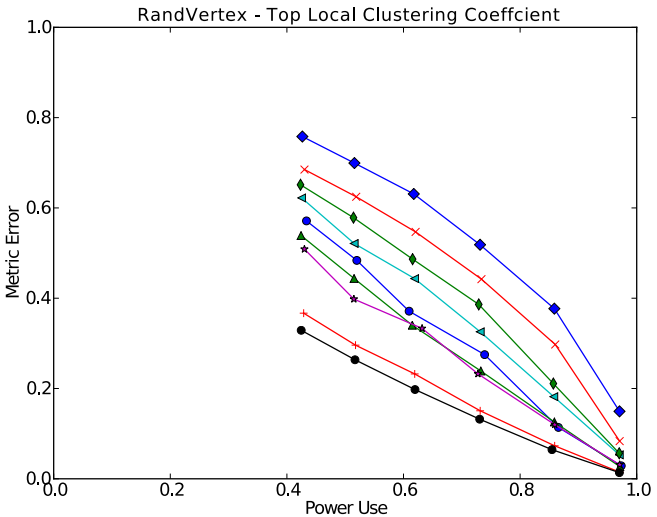


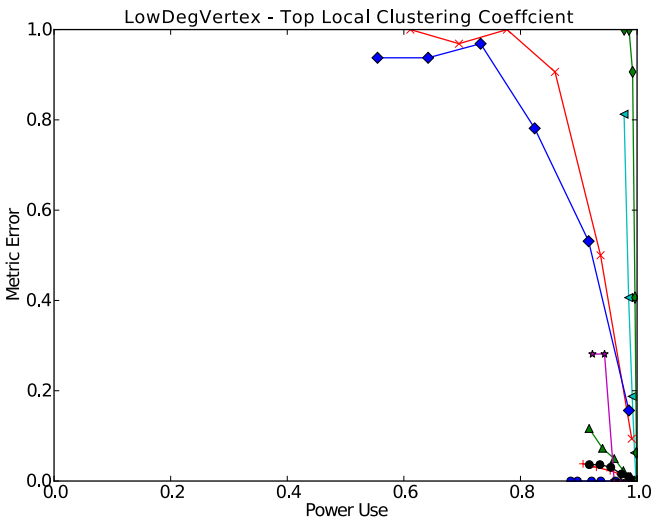
Quick Example: Local Clustering Coefficients

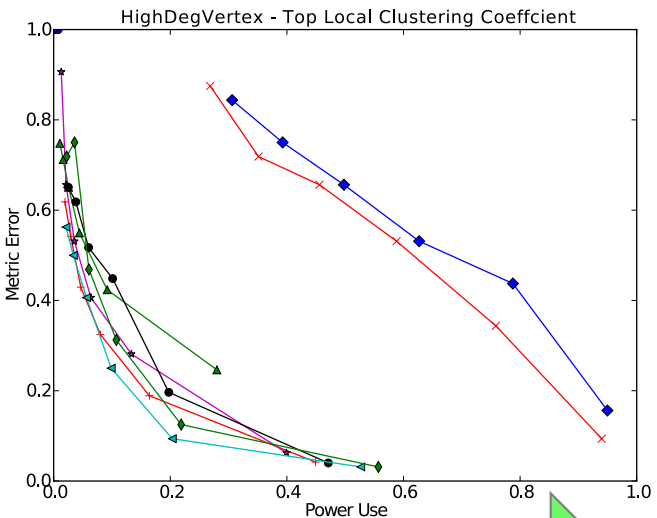
From Zakrzewska & Bader, "Measuring the Sensitivity of Graph Metrics to Missing Data," PPAM 2013





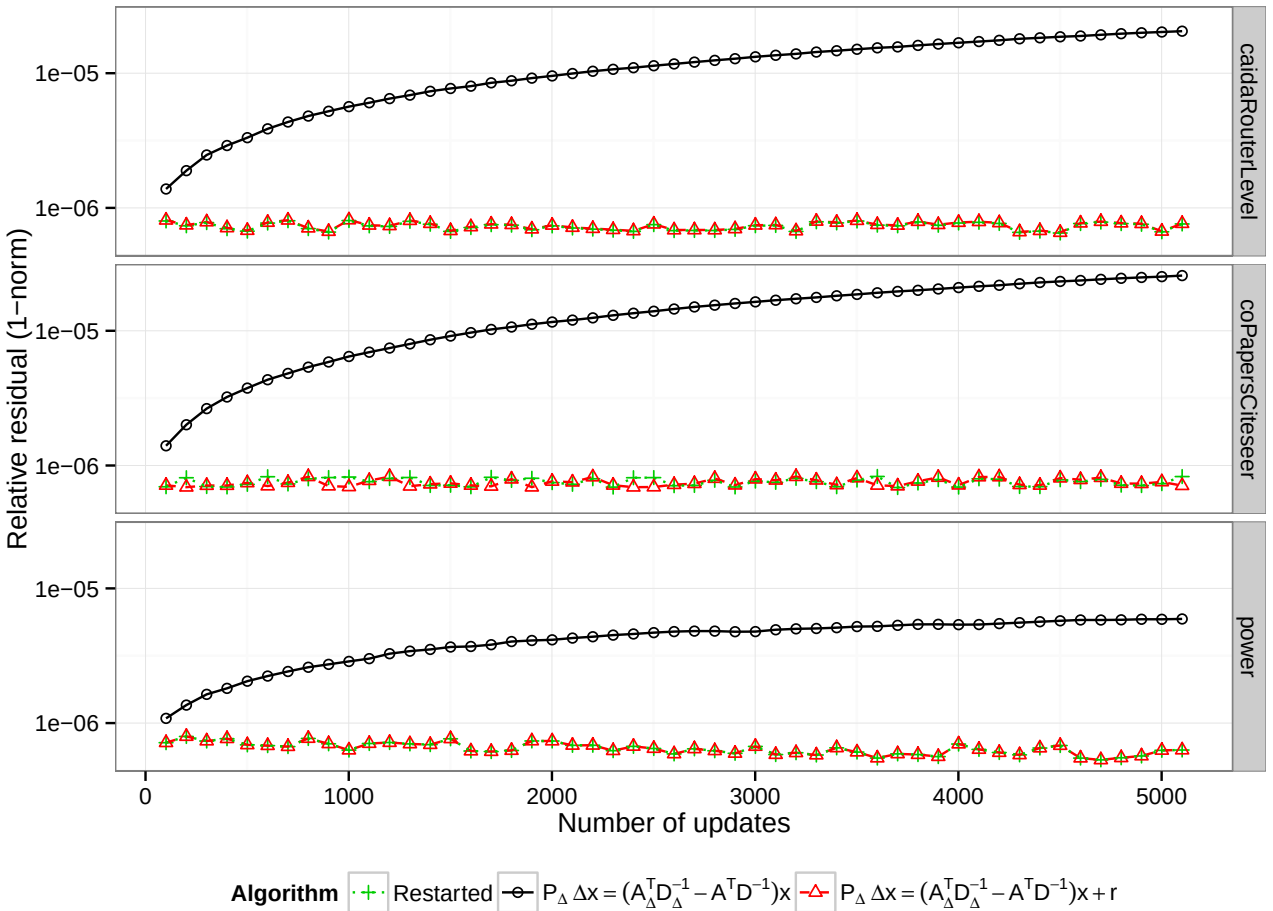






Quick Example: Streaming Magnifies Errors

Updating PageRank via simple linear algebra:



Ranking looks just fine! Until everything falls apart...

Paying attention to the initial error works.

Challenge: Build Error & Sensitivity Analysis for Graphs

Possible starting points

How do you measure or model error in...

- connected components?
  - Is the graph a window into the “real” network?
  - Can you leverage link prediction between components?
  - Measure precision and recall against... what?
- linear-algebra-ish metrics like PageRank?
  - Is this easier?
  - Mapping backward error analysis to a discrete matrix...

What is success?

Building mental and formal methods for addressing error and sensitivity that can be condensed to *rules of thumb*.