

---

# EgoNet Sketching of Graph Streams

Bortik Bandyopadhyay, David Fuhry, Aniket  
Chakrabarty, Srinivasan Parthasarathy  
*Department of Computer Science and Engineering*  
The Ohio State University



# The Data Deluge: Streaming Data Everywhere



facebook

twitter

LinkedIn

YouTube

Instagram

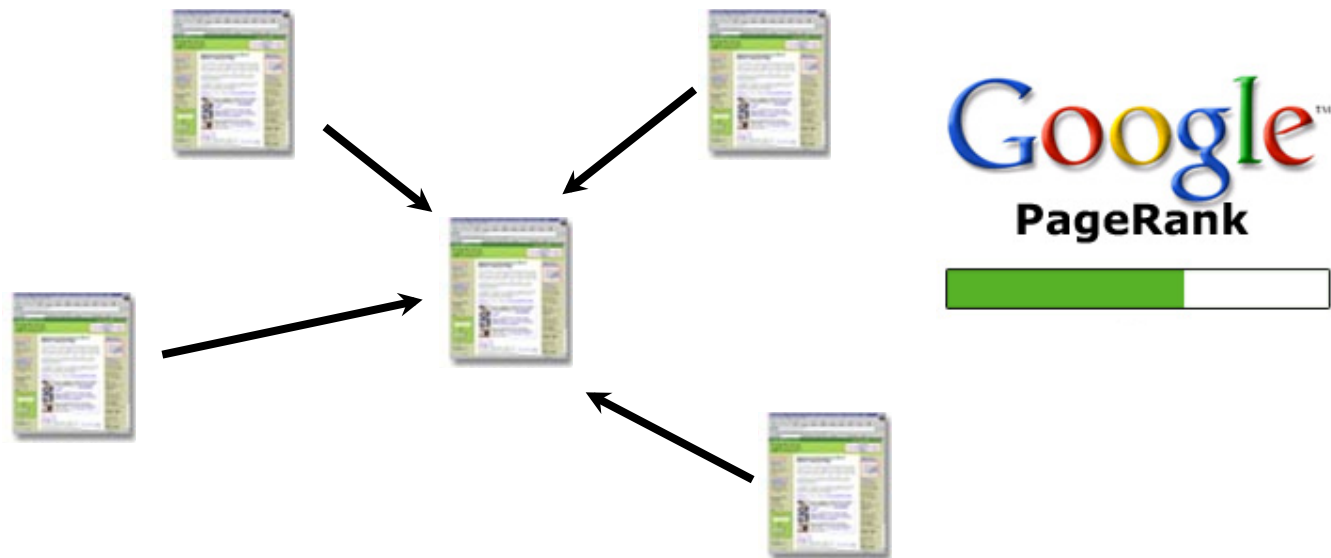


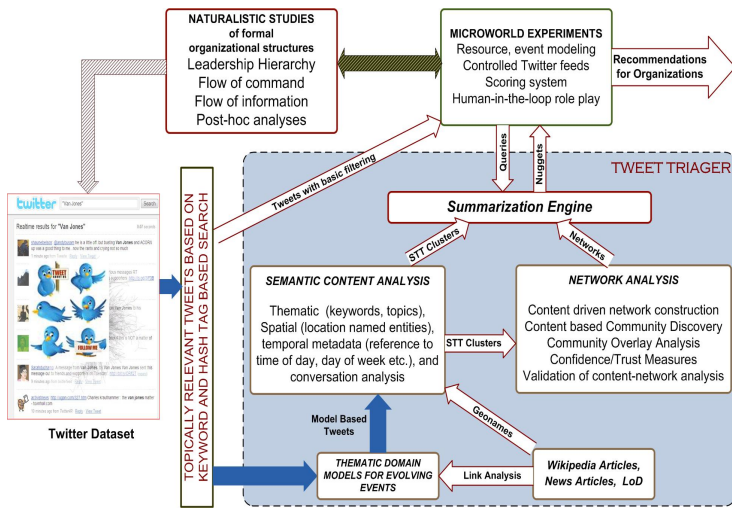
---

Data does not exist in isolation.

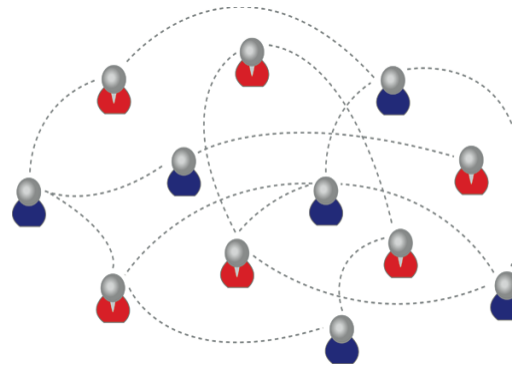


Data almost always exists in connection with other data – integral part of the value proposition.

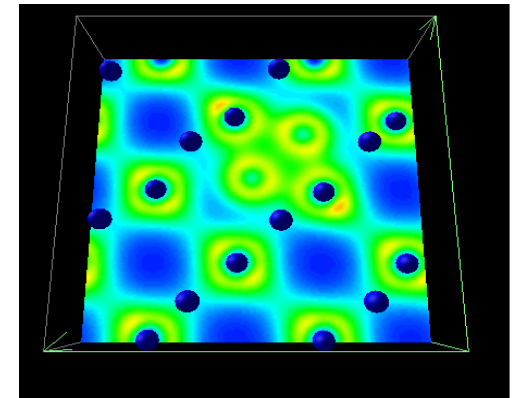




Crisis Response and Surveillance

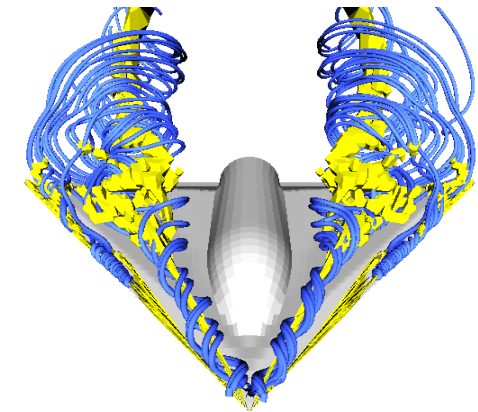
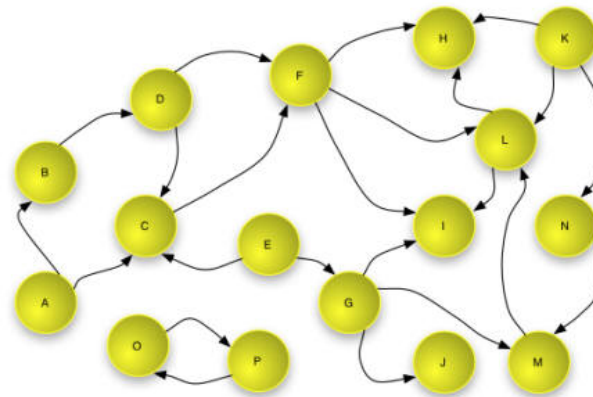
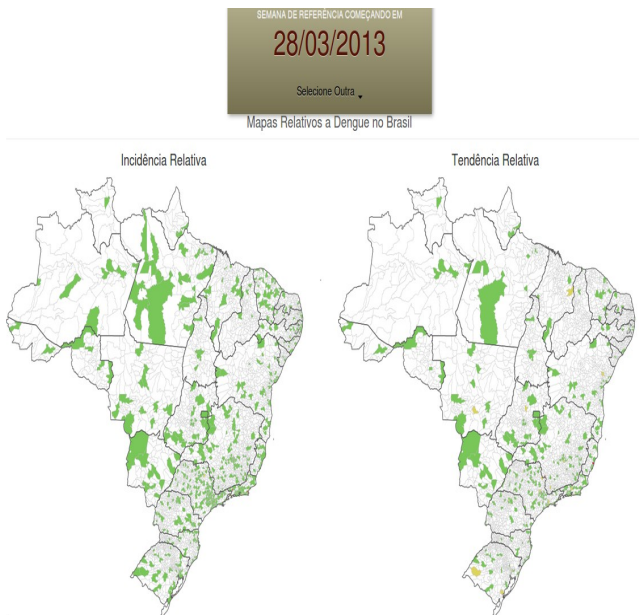


Social Networks



Mining Simulation Data

### Data (I/O) Dependency Analysis



---

# Challenges with Graph Analytics

- Scale
- Noise
- Topological Constraints
- Dynamics

Graph sketching under a streaming model

# Models of Graph Streaming

- Snapshot model [Muthukrishnan 2003, Asur 2007]
  - Graphs are presented in sequence in logical batches
  - Sketch -- Lossless compression [Buehrer 2008, 2015]
- **Semi-streaming model** [Feigenbaum 2005]
  - Memory budget proportional to node set; edges are streamed in.
  - Sketch – Lossy
- Fully streaming model [Muthukrishnan 2003]
  - Memory budget is fixed. Sketch – Lossy.

# EgoNet: Definition

- EgoNet (one-hop neighborhood of a node)
  - for directed graphs split into in-links and out-links
- EgoNet (L) – L-hop neighborhood
- EgoNet-sf (L) – subset of EgoNet(L) with semantic filtering

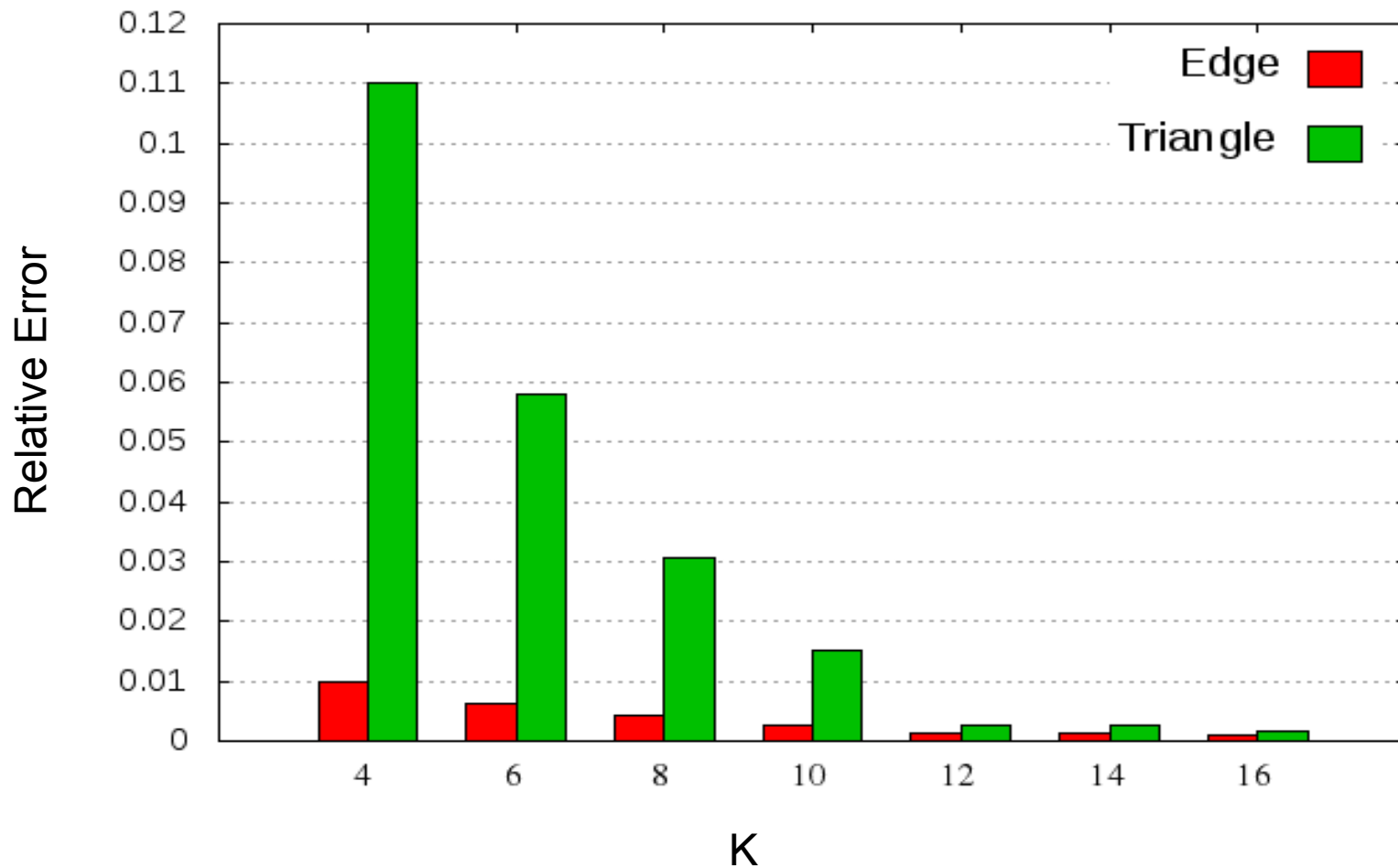


# Sketching EgoNets

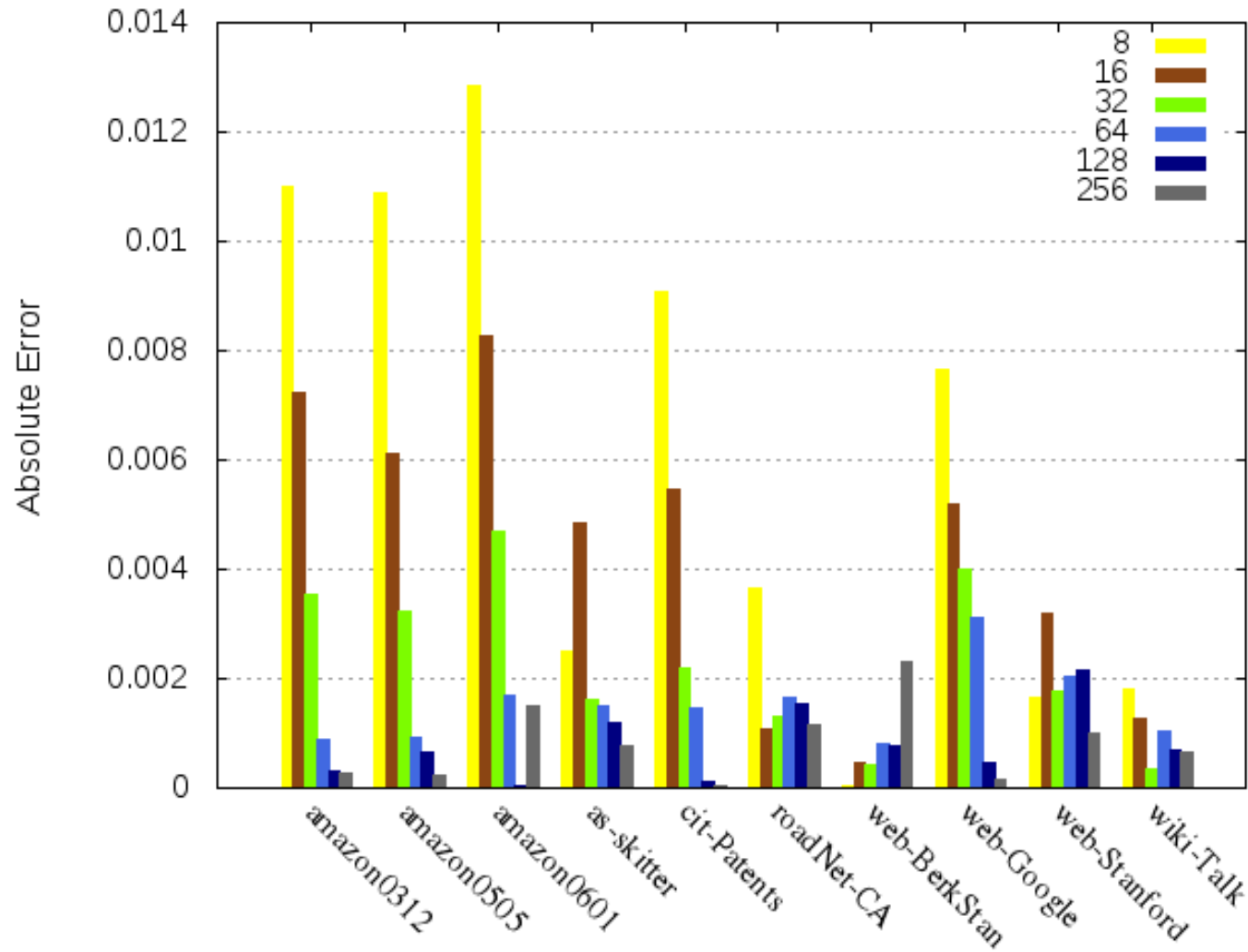
- Key Idea: Leverage Locality Sensitive Hashing [Broder'98, Indyk'99]
  - K minwise independent hashes (K controls sketch size)
- As edges arrive (semi-streaming model), update Sketch associated with nodes incident on each edge
- One pass algorithm to generate sketches – provably equivalent to generating the sketch on snapshot model!
  - Can compute properties of the original graph with strong guarantees

# Measures on EgoNet Sketch I

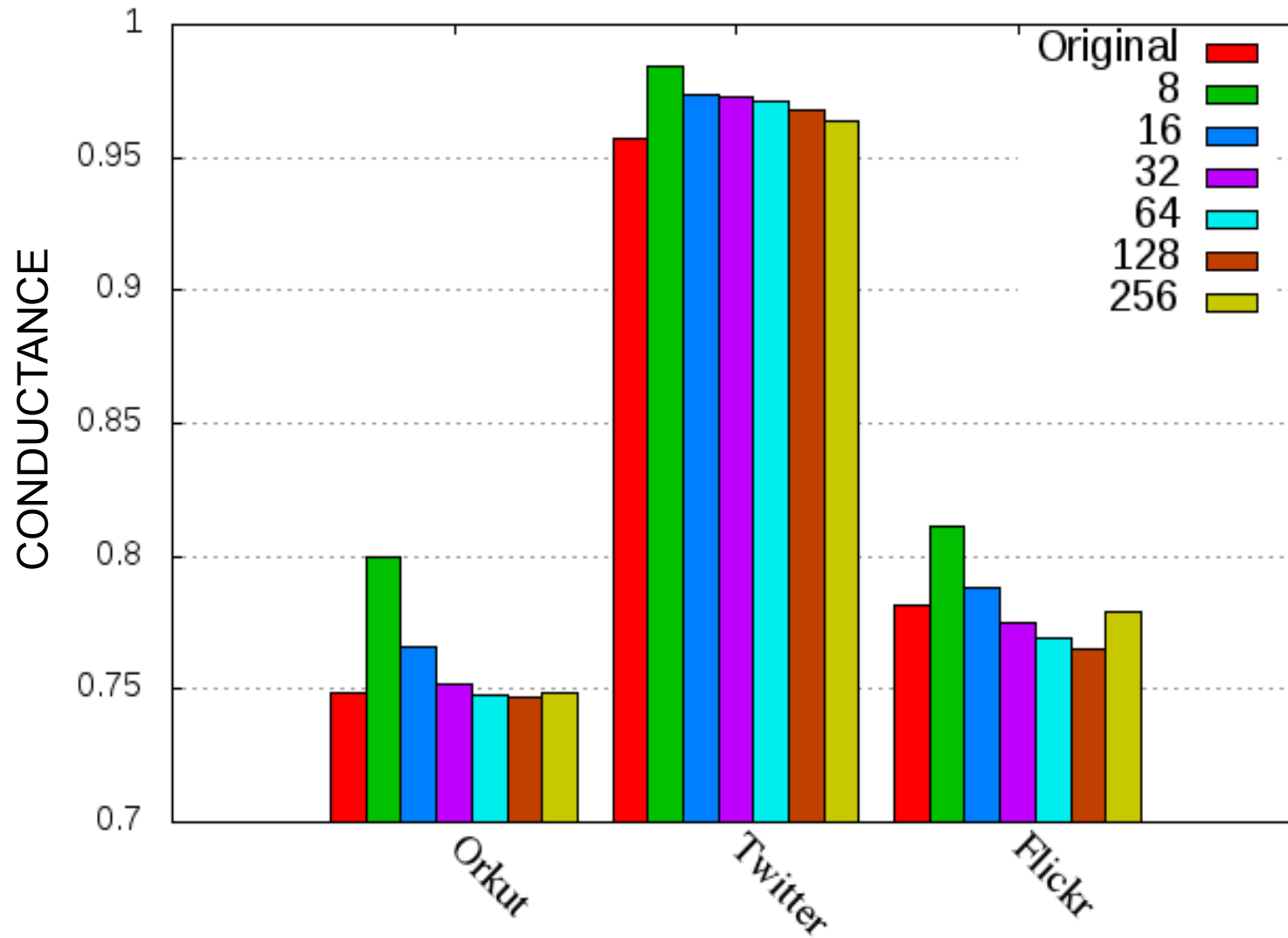
Horvitz-Thompson Estimators from Egonet sketch to estimate edge/  
triangle density



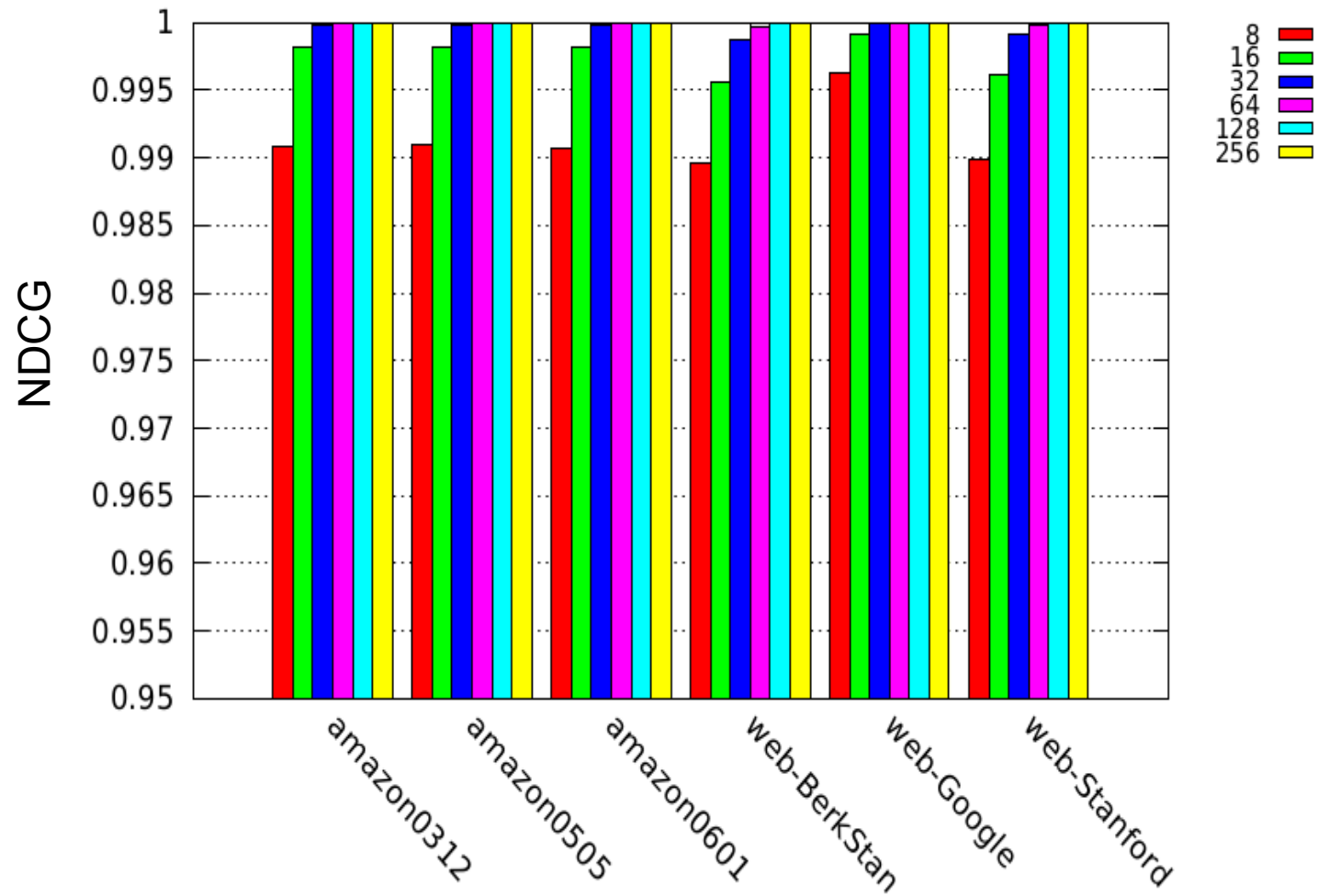
# Measures on EgoNet Sketch II: Computing Clustering Coefficient



# Measures on EgoNet Sketch III: Conductance Estimation



# Analytics on EgoNet Sketch: Page Rank



# Concluding Thoughts

- **Described a Topographical Sketching Algorithm for Graph streams under semi-streaming model**
  - Strong theoretical guarantees (statistical and spectral)
  - Can perform analytics directly on the sketch
  - Scales orders of magnitude faster
- **Easy to parallelize under a privatize and reduce model**
  - Can be accommodated on an accelerator
- **Can couple with real-time content analysis (e.g. Twitter)**
  - For disease surveillance and disaster response.

---

# Questions