

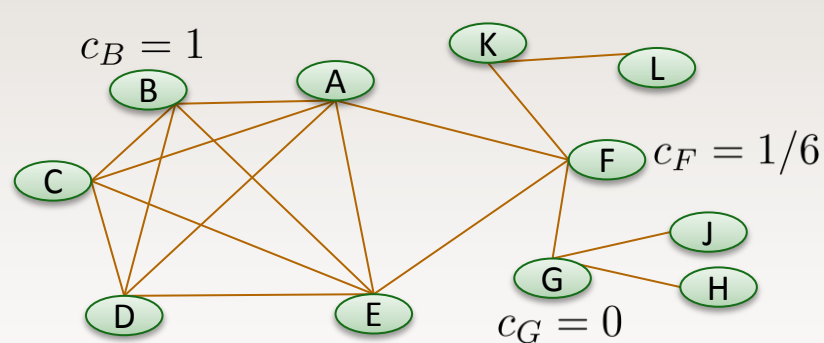
Fast algorithms for Evolving graphs via Assays, Sampling, & Tensors (FEAST)



Characterizing Networks

Undirected

The **transitivity** or **clustering coefficient** measures the rate of wedge closure.



$$c_i = \frac{\# \text{ closed wedges centered at node } i}{\# \text{ wedges centered at node } i}$$

$$c_d = \frac{1}{n_d} \sum_{i \in V_d} c_i = \text{mean for degree } d$$

$$c = \frac{3 \times \# \text{ triangles in graph}}{\# \text{ wedges in graph}}$$

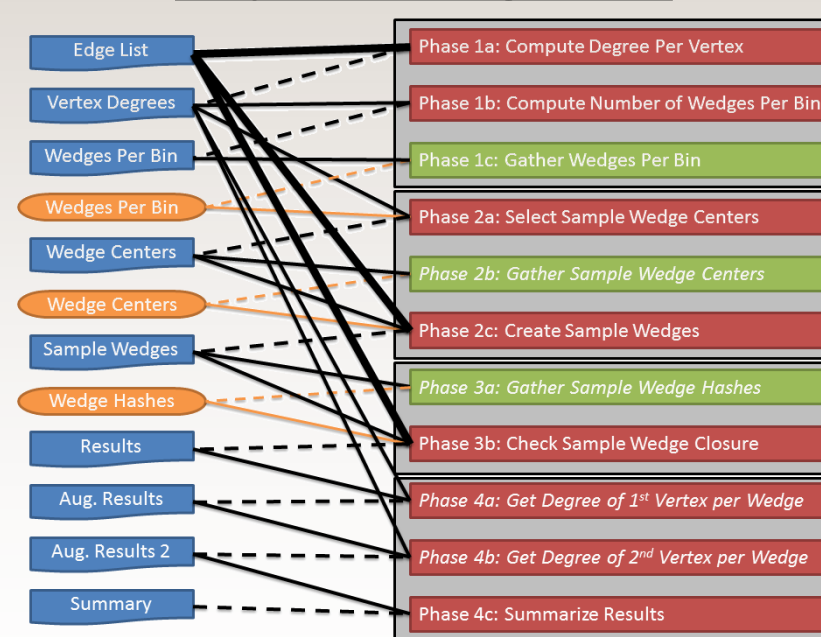
Triangles play a crucial role in network characterization. We discuss:

- Sampling methods for transitivity in very large graphs
- The extension of transitivity for directed networks
- The characterization of network traffic using graph assays

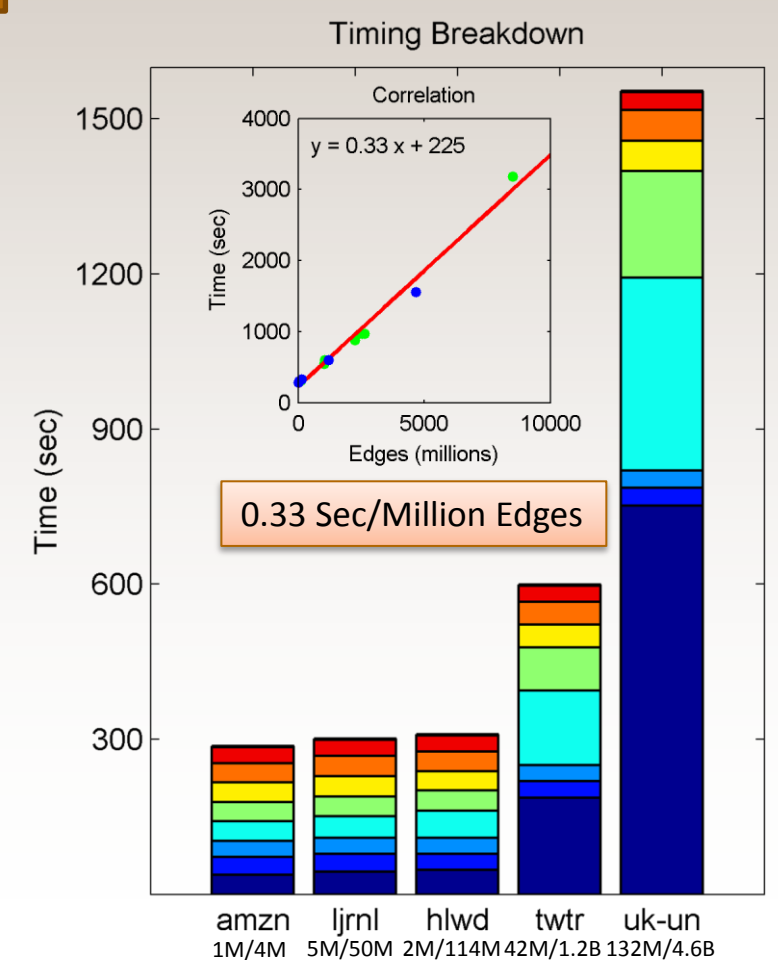
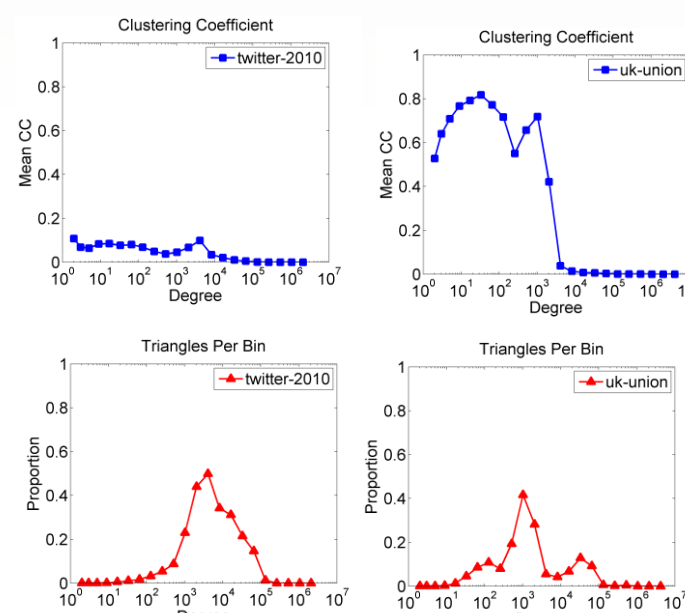
Sampling Approach

- Enumeration checks every wedge for closure: $c = \# \text{ closed wedges} / \# \text{ wedges}$
- Sampling chooses a few wedges (uniformly at random) and checks closure: $c \approx \# \text{ closed sampled wedges} / \# \text{ sampled wedges}$
- 1000X average speedup versus enumeration, $k = 32,000$ ($\epsilon = 0.011$)
- Faster than edge sampling (Doulion) and less variance
- Can also compute clustering coefficient per degree

MapReduce Algorithm



Sample Results

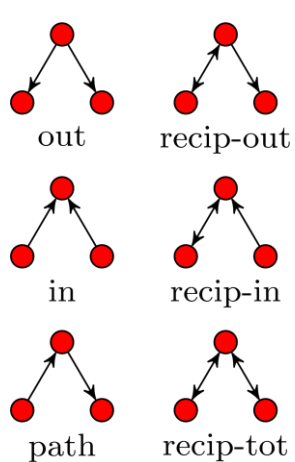


Distributed Server: 32-Node Hadoop Cluster

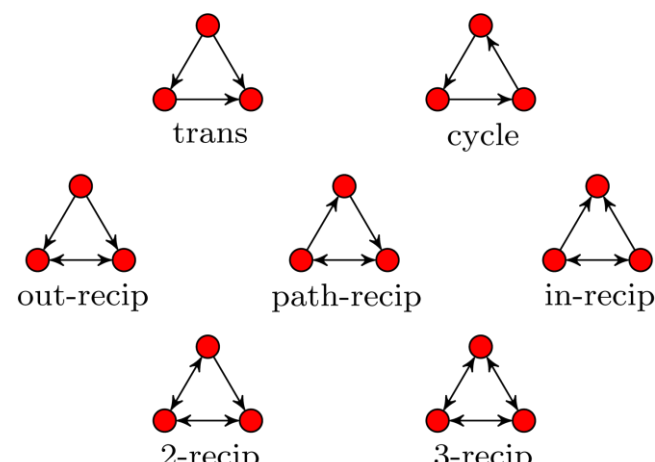
- 32 x Intel 4-Core i7 930 2.8GHz CPU = 128 cores
- 32 x 12GB = 384GB memory
- 32 x 4 2TB SATA disks = 256TB disk storage

Directed

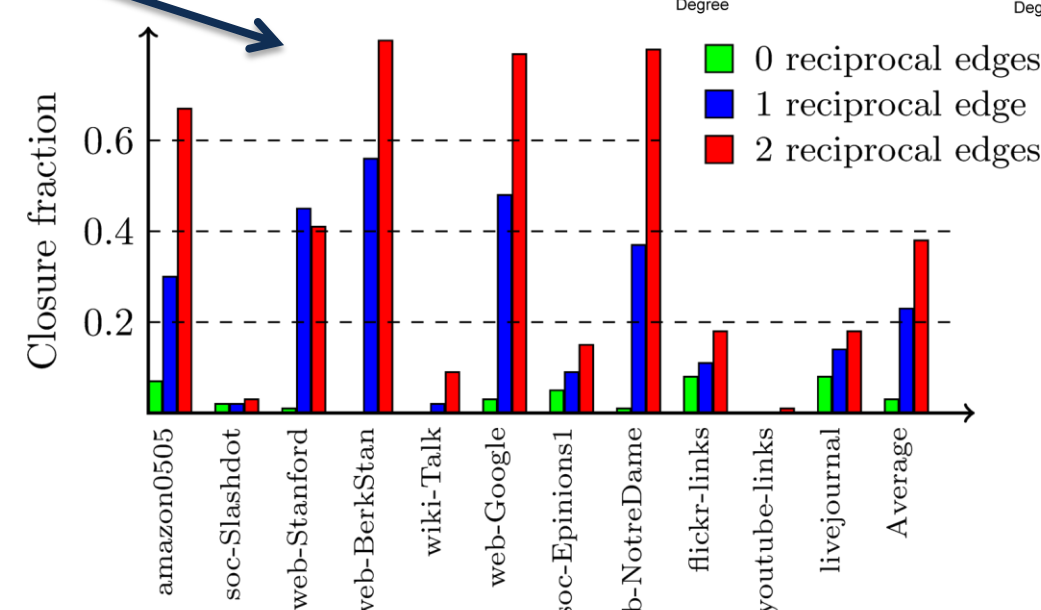
Directed Wedges



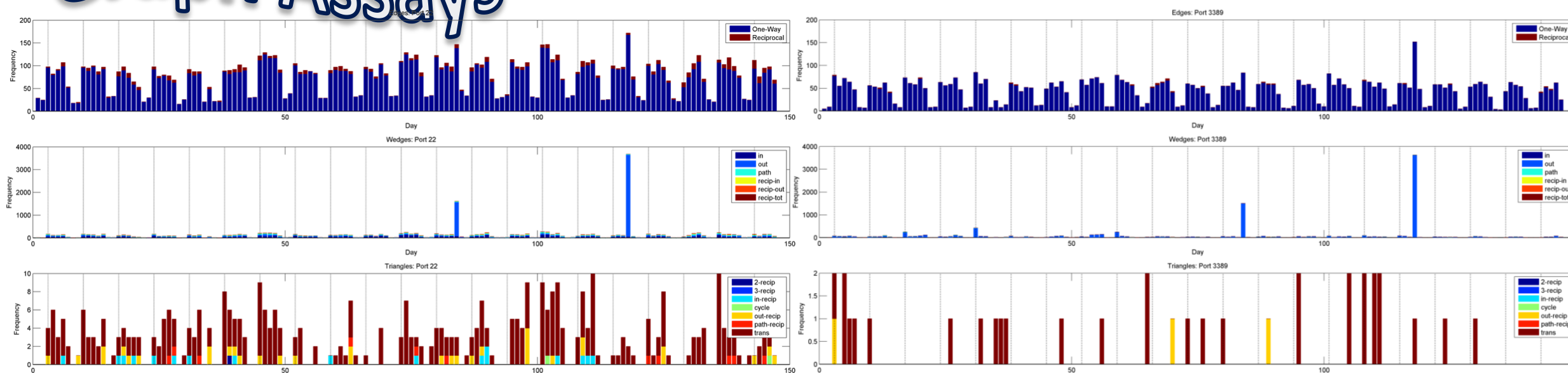
Directed Triangles



Wedges with reciprocal edges are much more likely to close in social and web networks.



Graph Assays



References

- C. Seshadhri, A. Pinar, and T. G. Kolda. **Triadic Measures on Graphs: The Power of Wedge Sampling**, SDM13: 2013 SIAM Int'l Conf on Data Mining, pp. 10-18, 2013
- T. G. Kolda, A. Pinar, T. Plantenga, C. Seshadhri and C. Task, **Counting Triangles in Massive Graphs with MapReduce**, SIAM J. Scientific Computing, accepting pending minor revision (preprint: arXiv:1301.5887)
- C. Seshadhri, A. Pinar and T. G. Kolda, **Wedge Sampling for Computing Clustering Coefficients and Triangle Counts on Large Graphs**, arXiv:1309.3321, 2013
- C. Seshadhri, A. Pinar, N. Durak and T. G. Kolda, **Directed Closure Measures for Networks with Reciprocity**, arXiv:1302.6220, 2013
- See also streaming work in other poster!

Assay Image credit: M. Rocklin/T. Kolda