



# Scaling Graph Clustering with Distributed Sketches

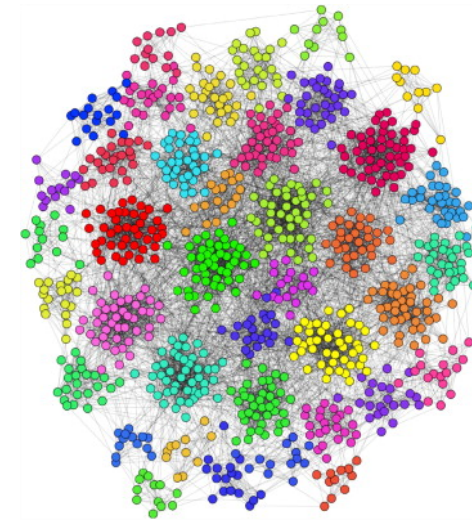
Alec Dunton  
Computing/CASC  
Ben Priest and Geoff Sanders



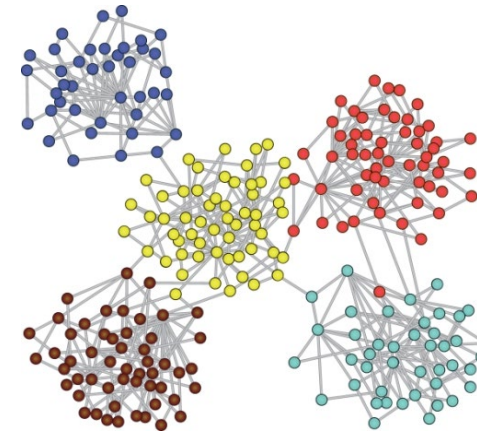


# Graph Clustering

- **Goal:** Identify “like” vertices as groups in a network.
- **Applications:** Contact Tracing!
- **Accepted Approach (Spectral):**
  - Embed graph matrix in low dimensional space spanned by subset of its eigenvectors.
  - Feed the embedding to a clustering algorithm.
- **Challenges:**
  - Distributed: bulk synchronous communication.
  - Streaming: spectral methods slow to update.
- **Proposed Solution:**



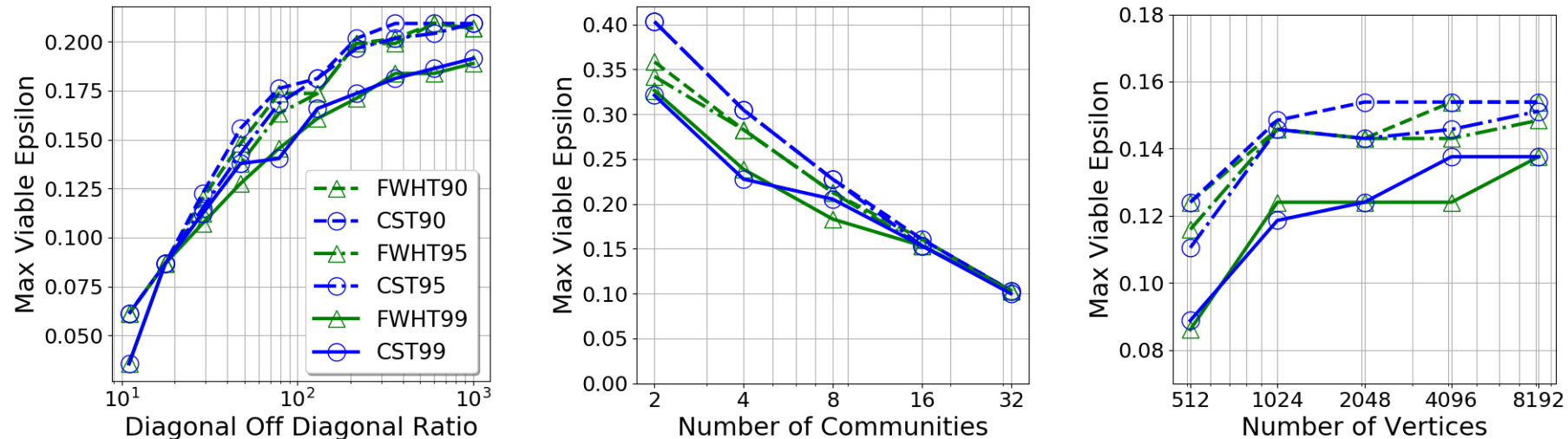
*sciencedirect.com*



*github.com*



# Matrix Sketching: Overview and Results



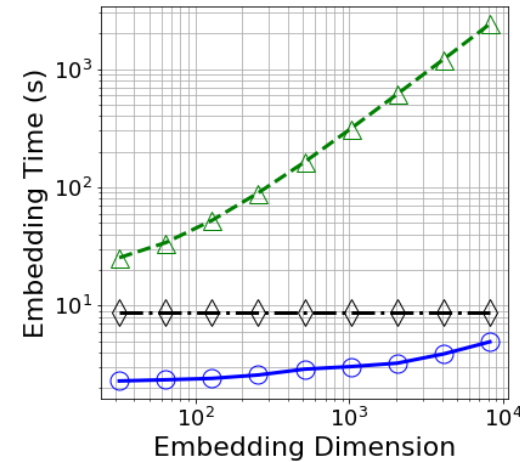
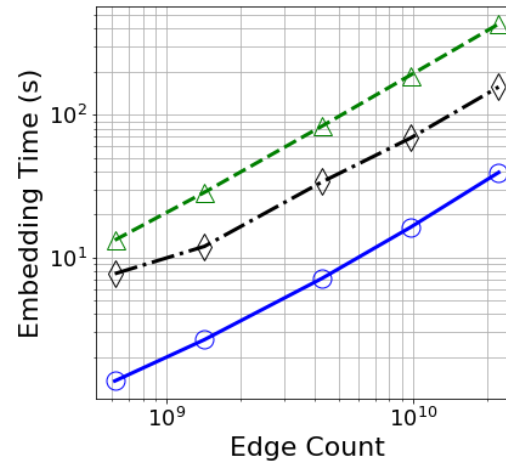
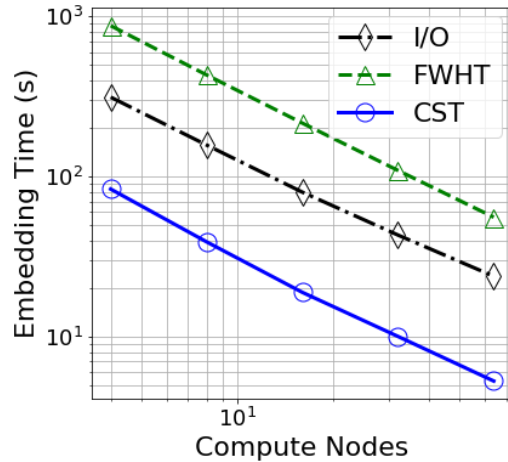
**Y-axes:** Required geometric fidelity of sketch. **X-axes:** (left) ratio of intra-group connectivity to inter-group connectivity; (center) number of true groups, fixed size; (right) size of graph, fixed number of groups.

- Randomly embed a graph matrix in a lower dimension, preserve important geometric structure.
  - Fast Walsh-Hadamard Transform (FWHT): theoretically sound, harder to implement.
  - CountSketch (CST): less rigorous, but faster and easily implemented in distributed setting.
- Use Stochastic Block Models (SBMs) in our numerical experiments.
  - SBMs function as surrogates for real world networks.
  - Randomly assign vertices to groups; high connectivity within groups, low between groups.



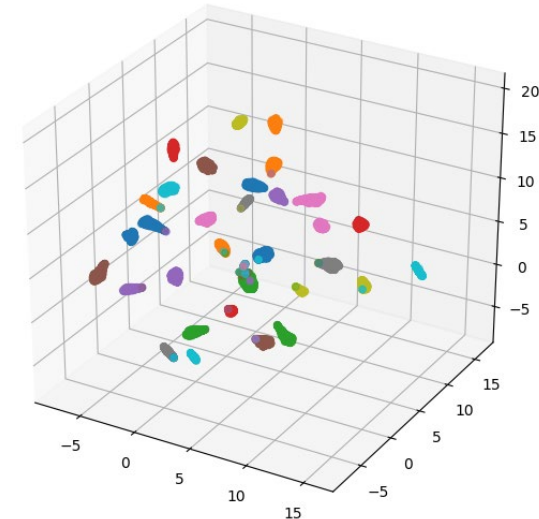


# Results, Next Steps, Questions?



$ \mathcal{V} $	Pair Precision, Recall > 0.9 Parameters				
	# Partitions	$\epsilon$	PP	PR	Accuracy
500	8	0.1	0.96983	0.97719	0.986
1000	11	0.1	0.95991	0.95301	0.976
5000	19	0.05	0.97103	0.97395	0.9878
20000	32	0.018	0.91305	0.90455	0.9588
<b>50000</b>	44	0.01	0.55959	0.12414	0.73773

- CST is faster than I/O, as accurate as FWHT!
- Next: Optimization of Clustering Algorithms
- See our paper: <https://arxiv.org/abs/2007.12669>





#### **Disclaimer**

This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes.