CATEGORY: **VISUALIZATION - IN-SITU & SCIENTIFIC - VS04**

POSTER **P5200**

CONTACT NAME **Karan Sapra: ksapra@g.clemson.edu**

GPU TECHNOLOGY CONFERENCE

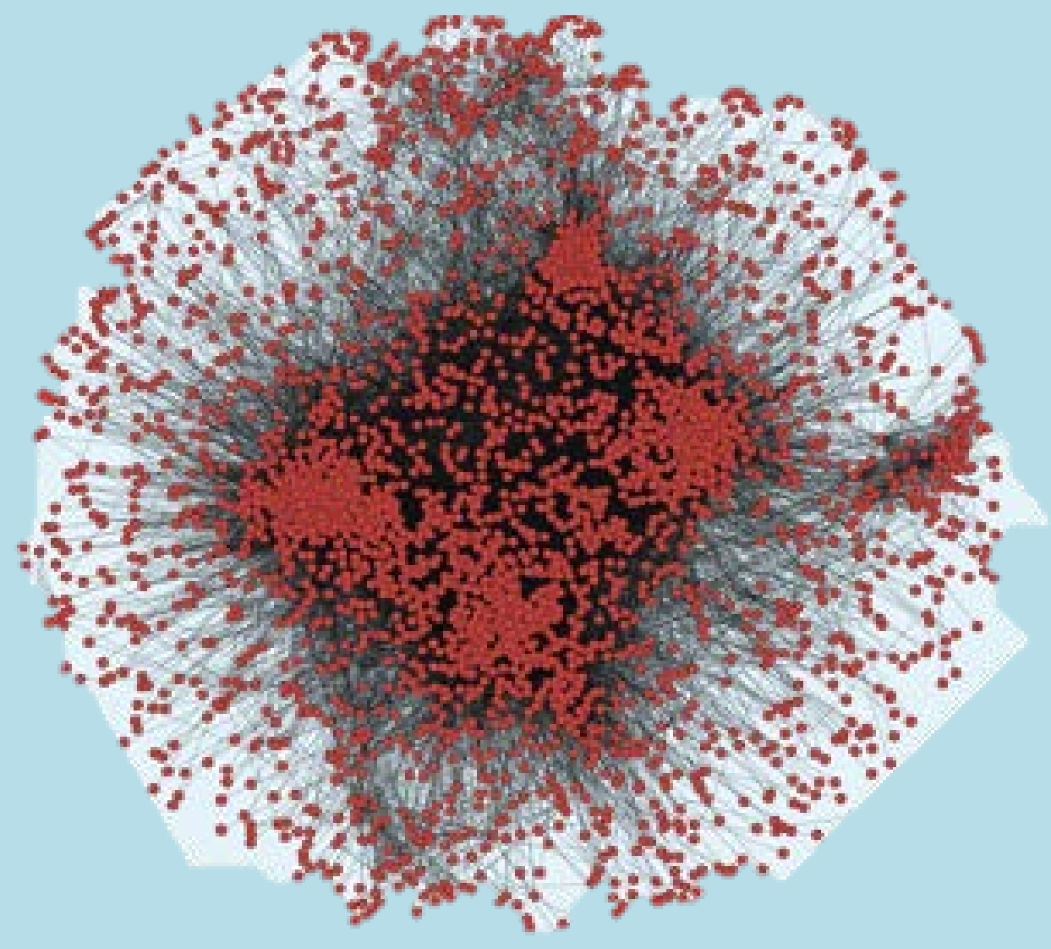
G^{3}NA: A GPU Optimized Global Gene Network Alignment Tool.

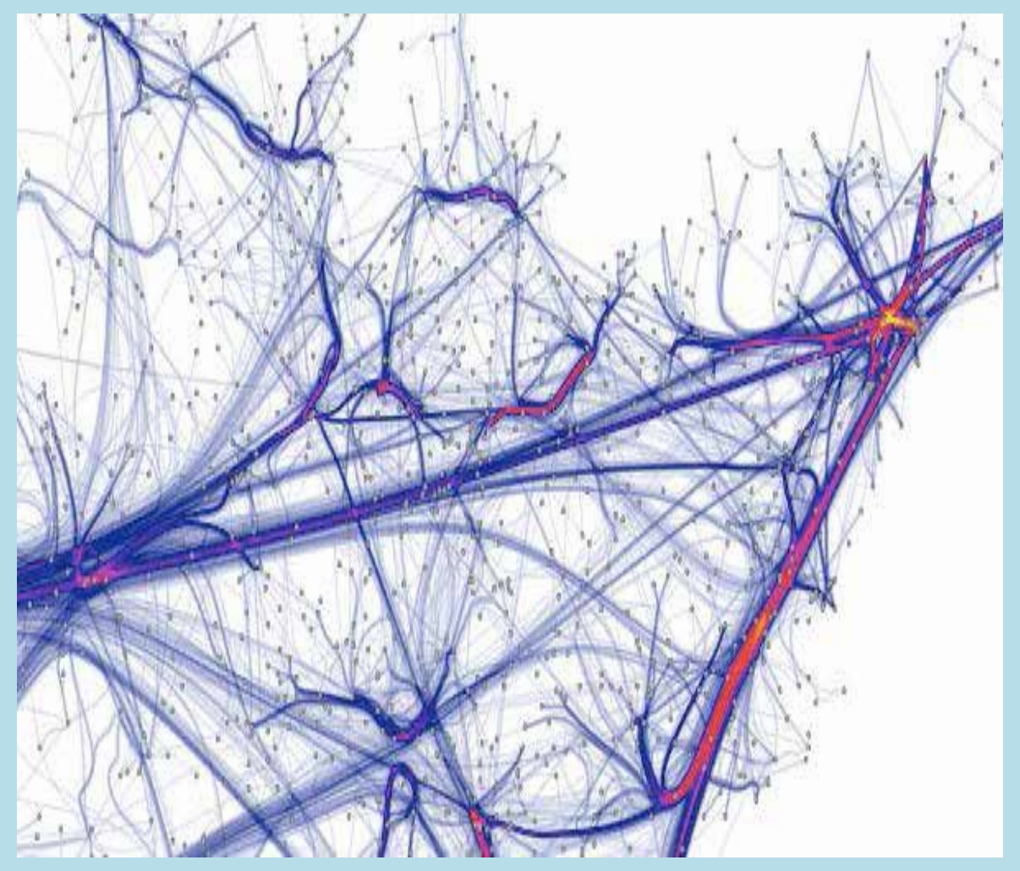
Karan Sapra[1], Melissa C. Smith[1], F. Alex Feltus[2], Asher Sampong[3], Joshua A. Levine[4], Anagha Joshi[1]
Departments of [1]Electrical and Computer Engineering and [2]Genetics & Biochemistry, Clemson University, SC, 29634, USA;
[3]Fort Valley State, Fort Valley, GA 31030, USA; [4]School of Computing, Clemson University, SC, 29634, USA.

Abstract

Gene interaction graphs help genomics researchers discover gene expression patterns across multiple tissues and experimental conditions. Alignment of these graphs reveals conserved sub-graphs that show evolutionarily or condition conserved sets of genes controlling biological processes. However, alignment of medium sized graphs (2000+ nodes) is computationally expensive and visually overwhelming, and for large graphs alignment can become unrealistic without the use of advanced computing resources. Thus, we have developed a GPU based optimized framework to perform multi-genomic network alignment and in-situ based visualization using GPUs. We aligned two gene interaction networks from rice and maize (each ~2,000 nodes/~20,000 edges) on the Palmetto Supercomputer using NVIDIA K20 GPUs. Maize-rice alignment took 152 seconds compared to 8 hours with IsoRankN. For larger random network (8,000 nodes/60,000 edges), we observed a speed-up of 51.3x.We examine the results of alignment using a visualization tool. Furthermore, we provide visualization of these networks using OpenGL on Palmetto and WebGL on a web-browser.

MOTIVATION

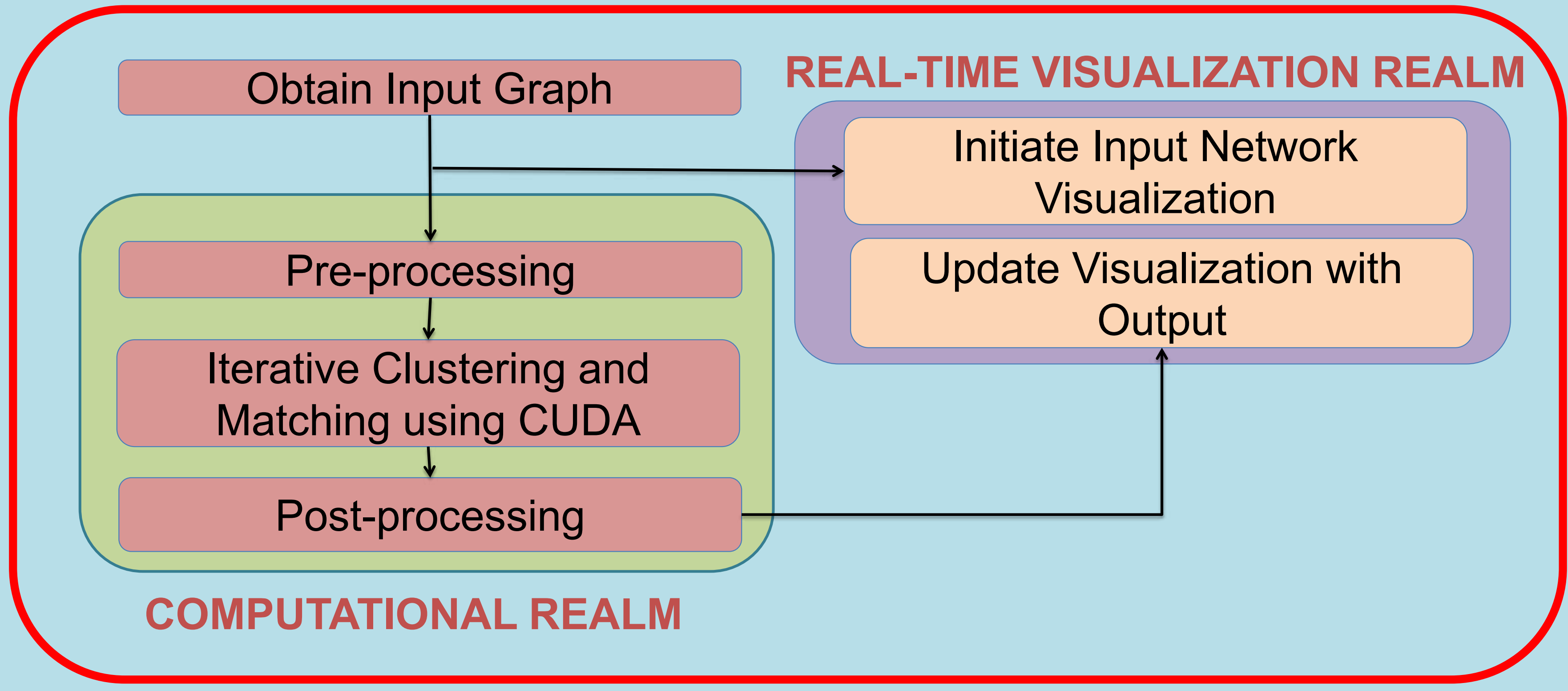




- **Medium Networks such as Rice and Maize go up to 2000 Nodes and 40,000 Edges**
- **Large Scale Networks such as for Humans can go up to 10,000 Nodes and 200,000 edges**
- **Analysis and Visualization of Large-scale Network**
- **Real-time In-situ visualization is tough with large-scale**
- **GPUs are really good at Computation and Visualization**
- **Existing Approaches do not utilize the GPU or all the resources on a node completely.**

OUR APPROACH

G3NA Workflow



REAL-TIME VISUALIZATION REALM

Obtain Input Graph

Initiate Input Network Visualization

Update Visualization with Output

Pre-processing

Iterative Clustering and Matching using CUDA

Post-processing

COMPUTATIONAL REALM

Input: {***Graph A***, ***GRAPH B***, ***Homology Graph A–B***}
Primary Output: {***Interaction Graph A-B***}
Secondary Output: {***Web-Browser based Visualization***, ***GPU Based Visualization***}

Computation Realm

- CUDA code performs the alignment of Graph A and Graph B on Palmetto using NVIDIA K20s
- Python scripts for Pre- and Post-Processing
- Can perform multi network-alignment using pair-wise alignment

Visualization Realm

- Real-time *OpenGL and CUDA based variation of* force-directed layout visualization on Palmetto using NVIDIA K20s
- Web-Browser based visualization using *WebGL*
- Interaction and attribute based layout in 2D and 3D

PERFORMANCE RESULTS

GPU versus Serial Implementation

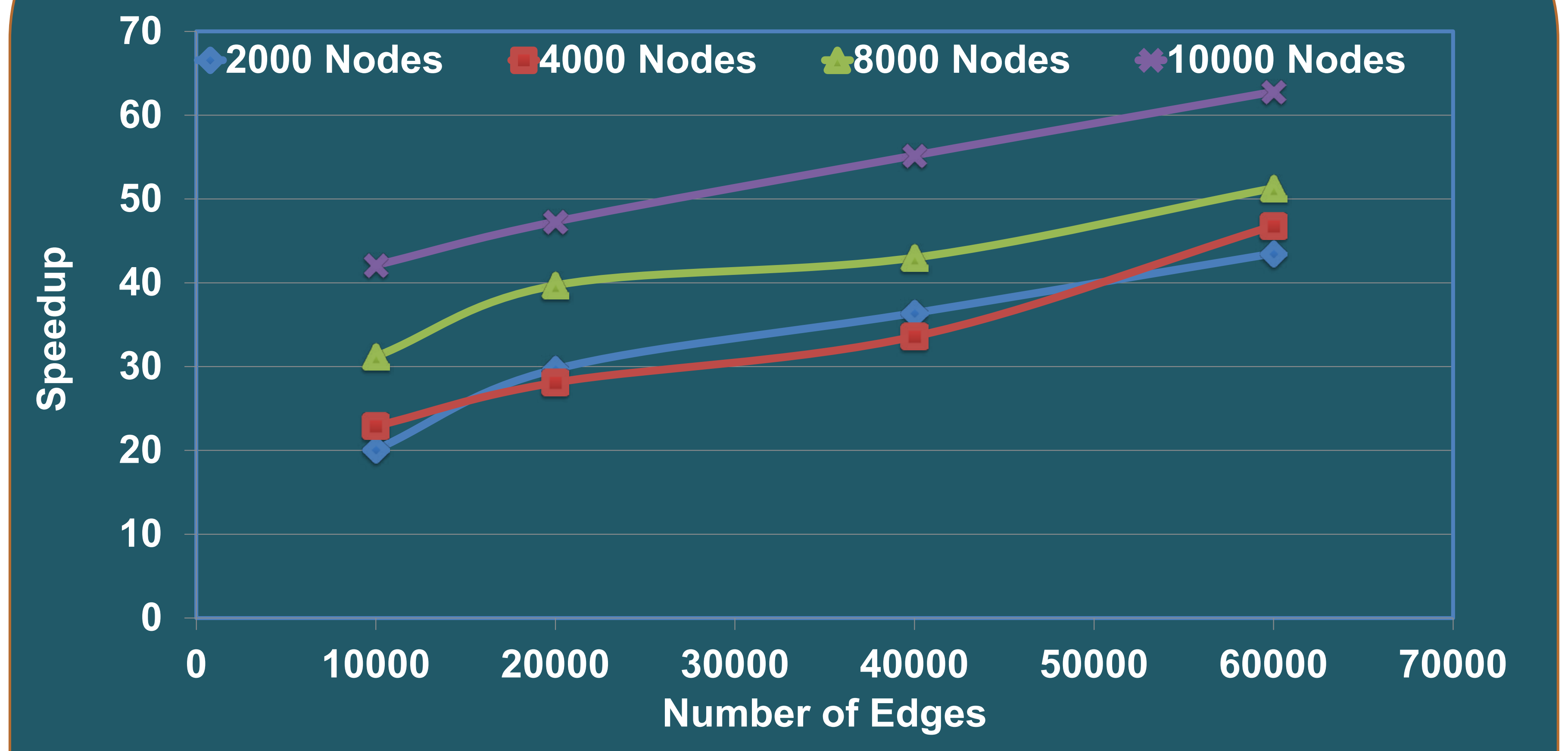


Figure 1. Speedup *for varying of nodes and edges in each of the two input graphs.*

Visualization

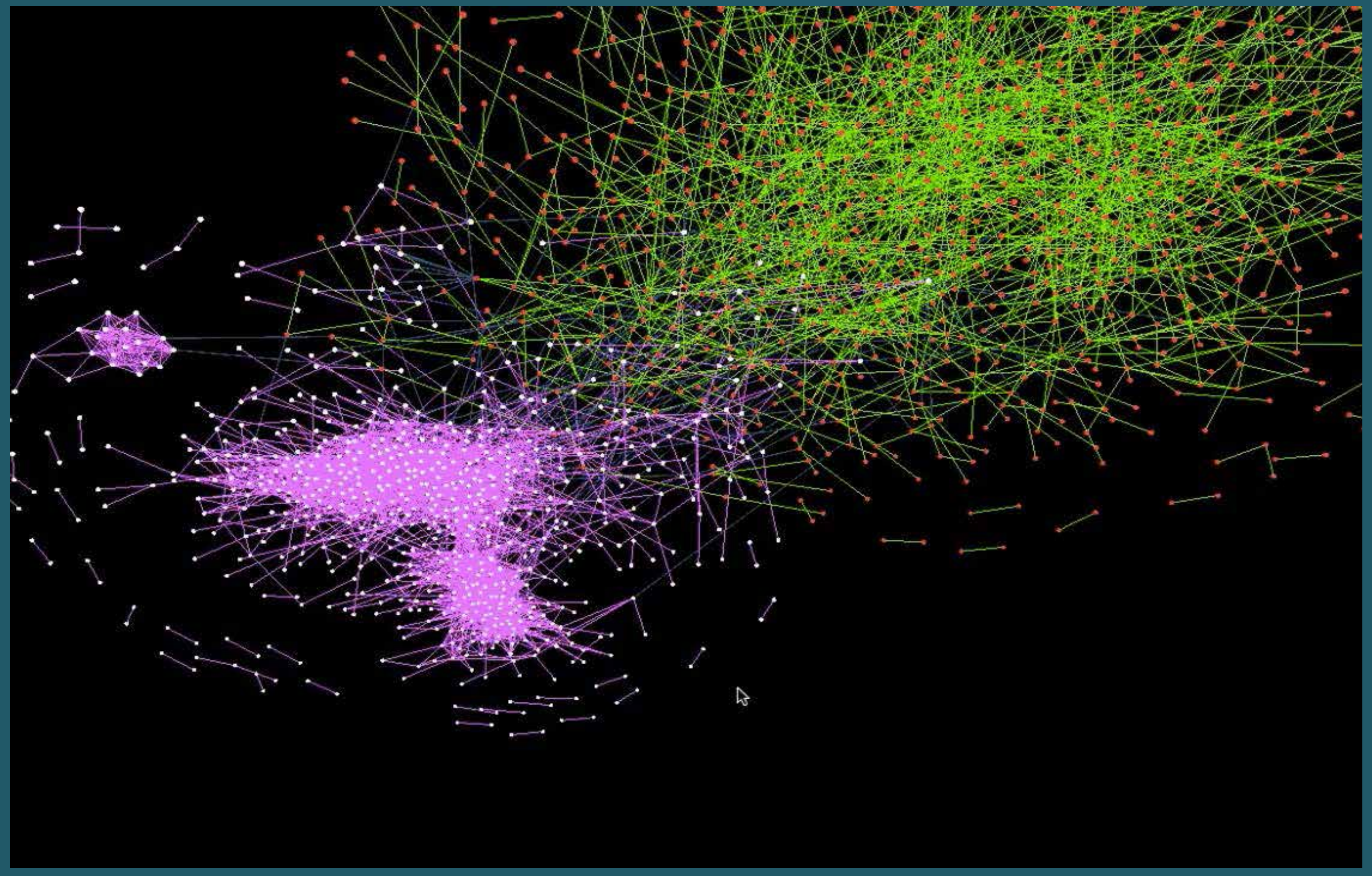


Figure 2. Palmetto based Network Alignment visualization using CUDA and OpenGL.

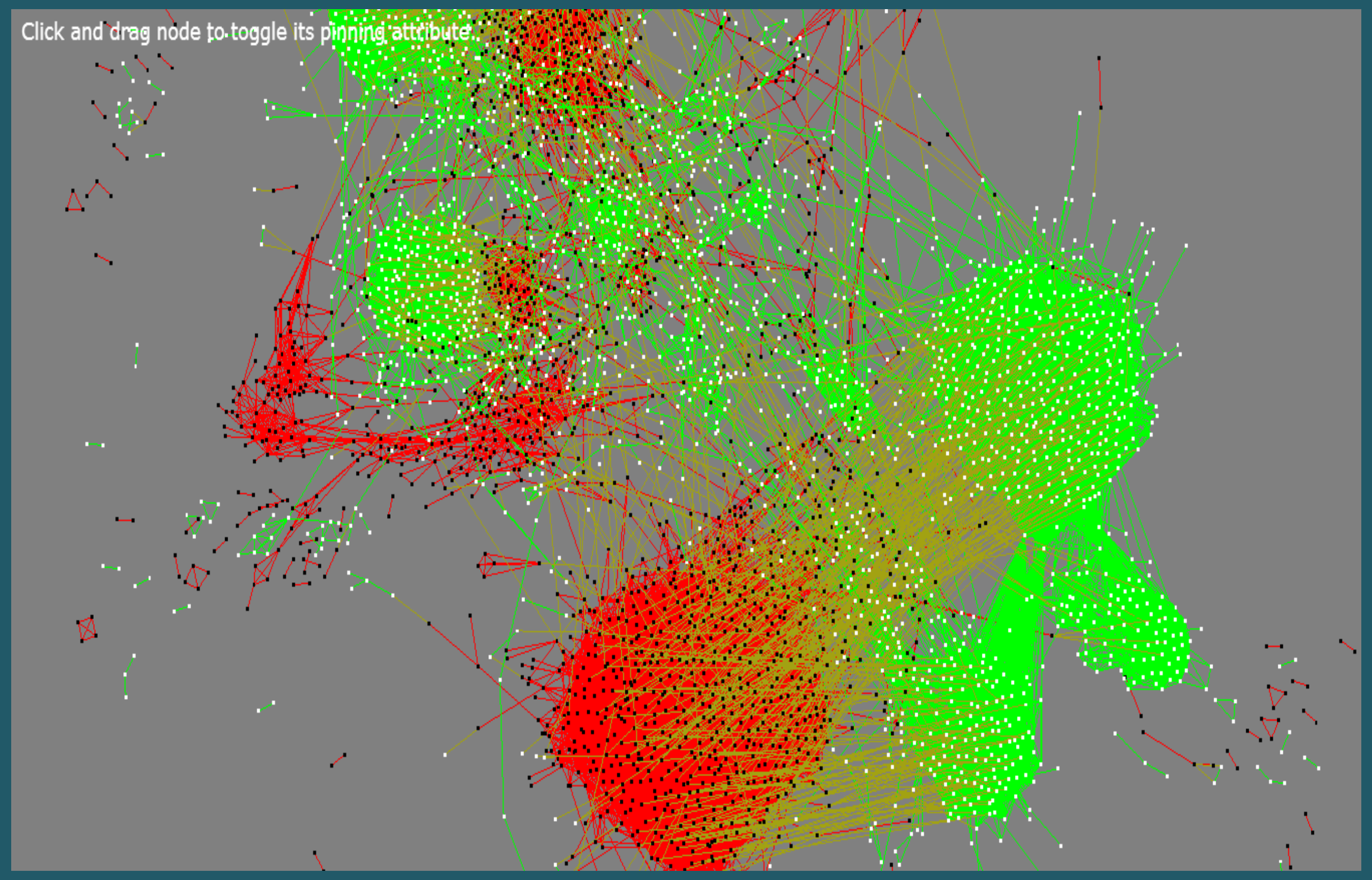


Figure 3. Web-Browser based Network Alignment visualization using WebGL and VivaGraphJS.

CONCLUSION

- We developed a framework that makes the node as a self-contained entity by perform GPU based alignment and visualize multiple gene interaction networks on the Palmetto Supercomputer on an NVIDIA K20s in real-time.
- Both Maize and Rice Network are ~2,000 nodes networks with ~20,000 edges in the network. The alignment of Maize and Rice network takes 152 seconds compared to 8 hours of IsoRankN [Liao et al. 2009] and obtain 32.1x speed-up compared to a serial version for the maize and rice network. For larger random network containing 8,000 nodes and 60,000 edges, we observed a speed-up of 51.3x.
- We have developed a web-interface for submission of alignment jobs for external research community.