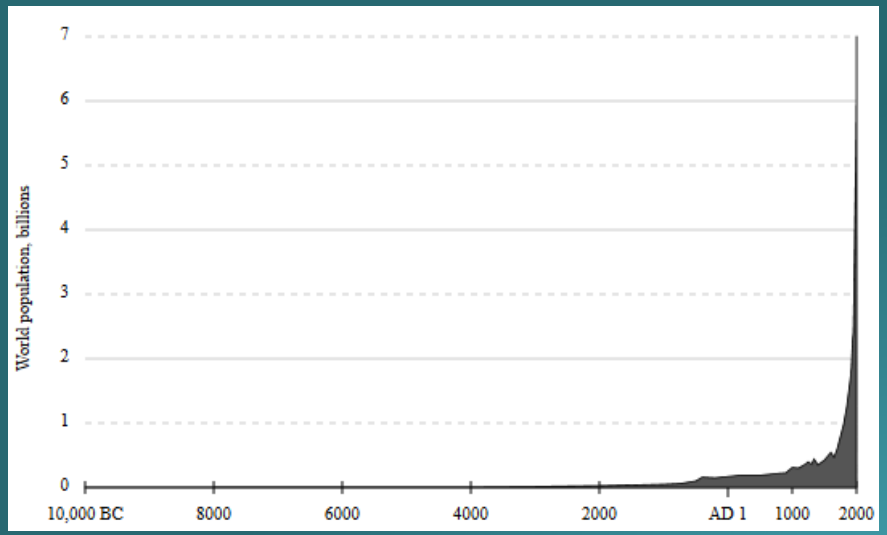
AN NVIDIA POWERED BIOGRAPH
ALIGNMENT AND VISUALIZATION TOOL

G3NA-V

Alex Feltus, Melissa C. Smith, Karan Sapra

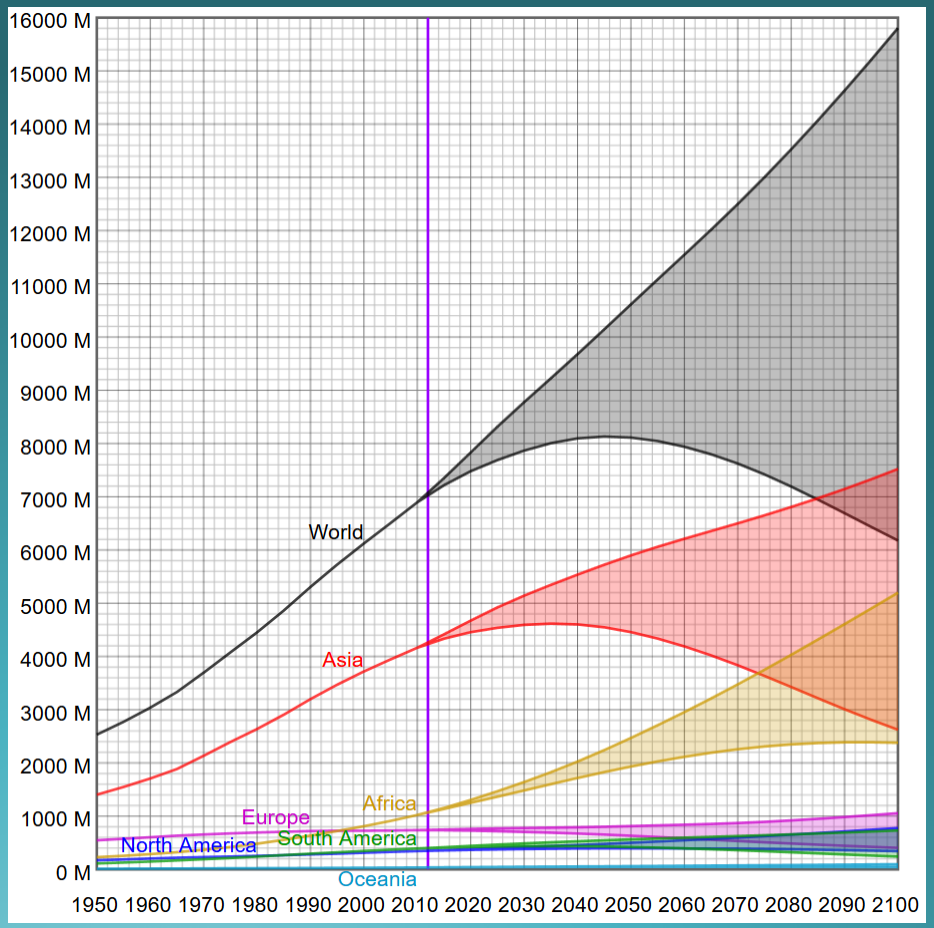
The world's population is growing…



1776: New York City (25,000), Boston (15,000).
2014: New York City (8,336,697), Boston (636,479).

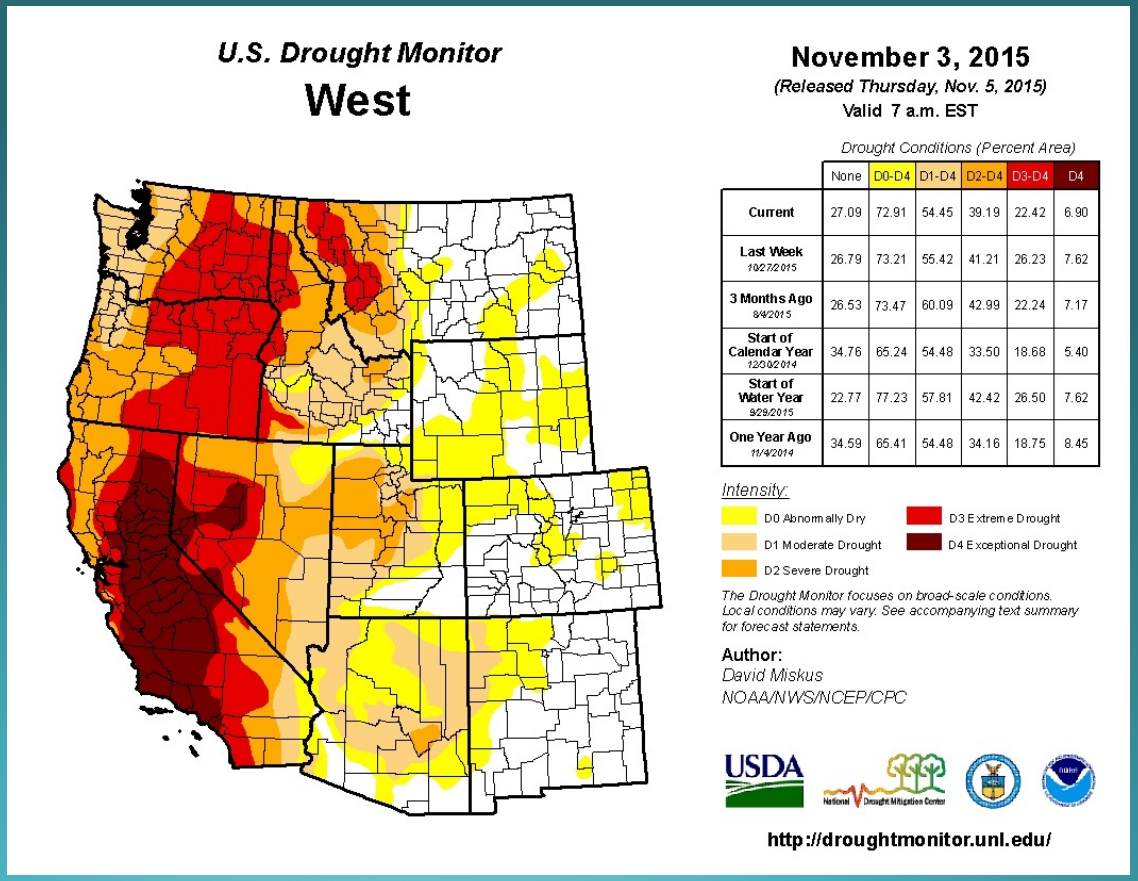
Current World Population

7,379,817,607



www.census.gov; Wikipedia; www.worldometers.info (11/9/2015)

The world's climate is changing…



Wikipedia;NOAA; "Tuvalu - Funafuti - Approach" by mrlins - http://flickr.com/photos/mrlins/304273866

Problem Statement

Current World Population

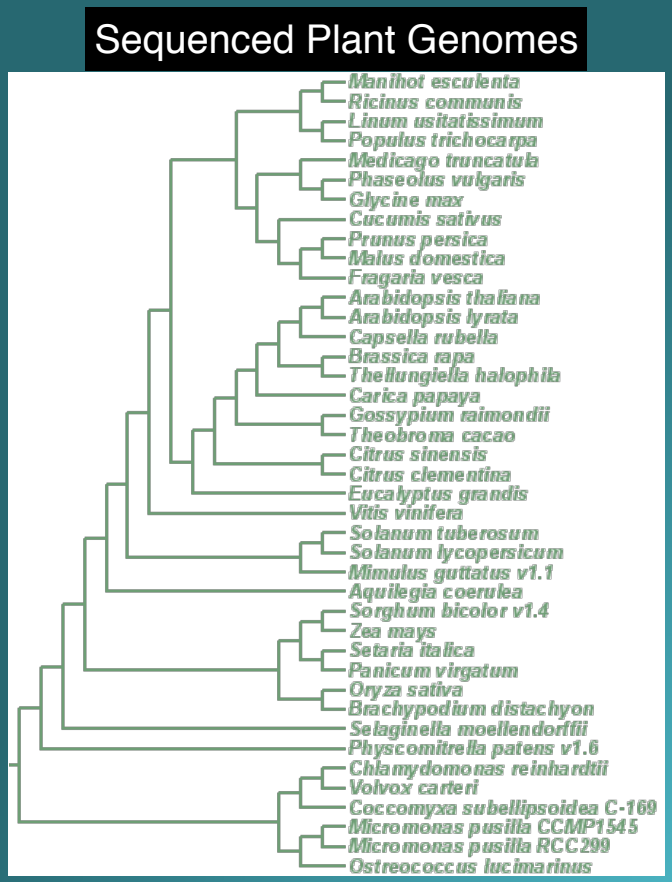
7,379,817,607

***Billions** of people need a lot of Food, Fiber, Fuel, Plastics, Medicines.*

***Faster Crop Development** required.*

***Climate history** may not predict the future.*

Accelerating crop development means understanding genomes…



Theobroma cacao
(Cocoa Bean)

Carica papaya
(Papaya)

Sorghum bicolor
(Sorghum)

Gossypium hirsutum
(Cotton)

Wikipedia; http://phytozome.net

CLEMSON UNIVERSITY

I ♥ Crop Systems Genetics

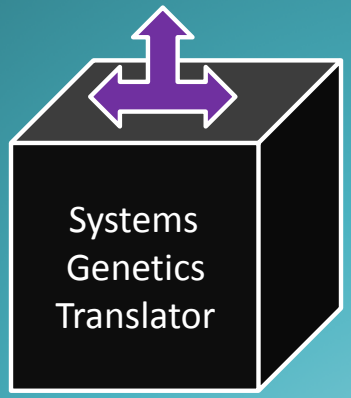
Rice

Sugarcane

Sorghum

Maize

Brachypodium



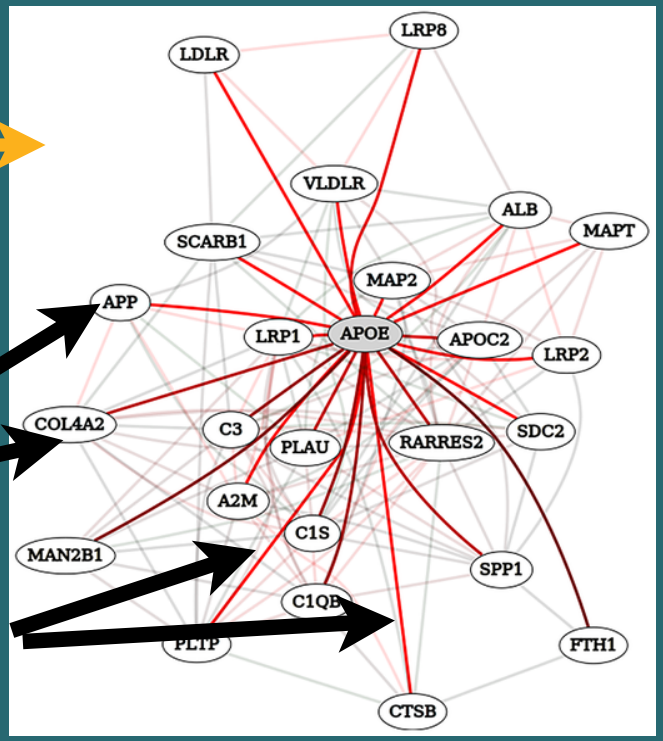
Switch grass

Miscanthus

*Images from www.wikipedia.org

Complex biological systems can be modeled as graphs…

Alzheimer's (Plaque in Brain)

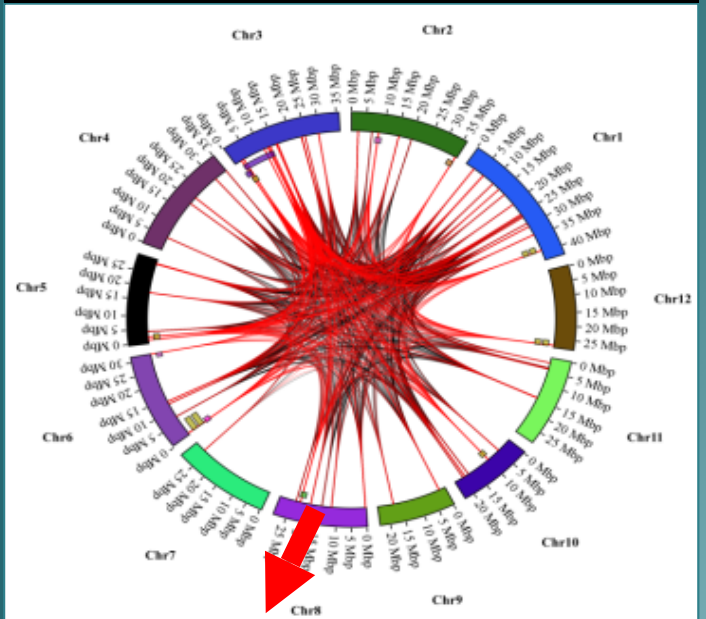


Node (Gene)

Edge (Gene Interaction)

Genenet Engine:
sysbio.genome.clemson.edu

Rice graph mapped to genome



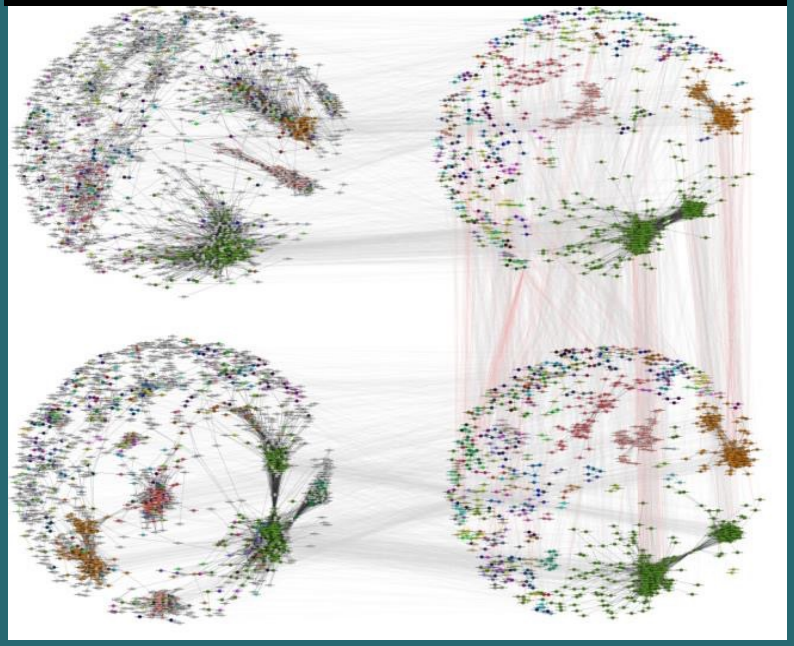
Higher Yield!!!!

Paleogenomics: Conserved subgraphs can be detected by graph alignment…

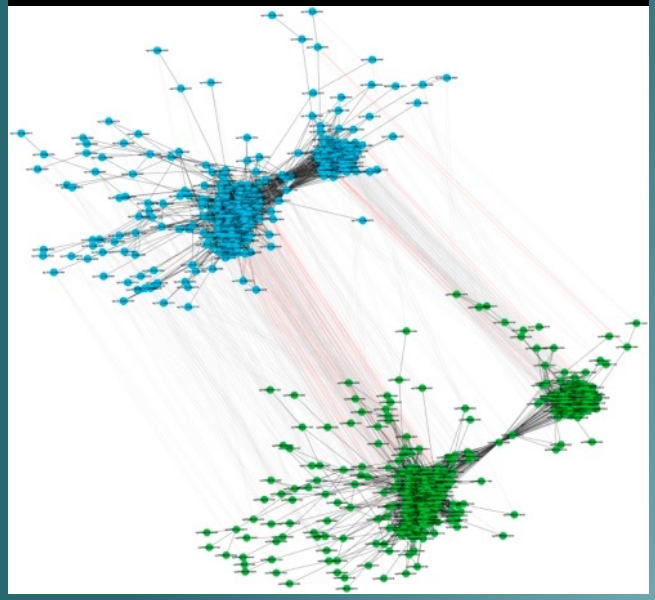
Maize

Rice

Aligned Graphs



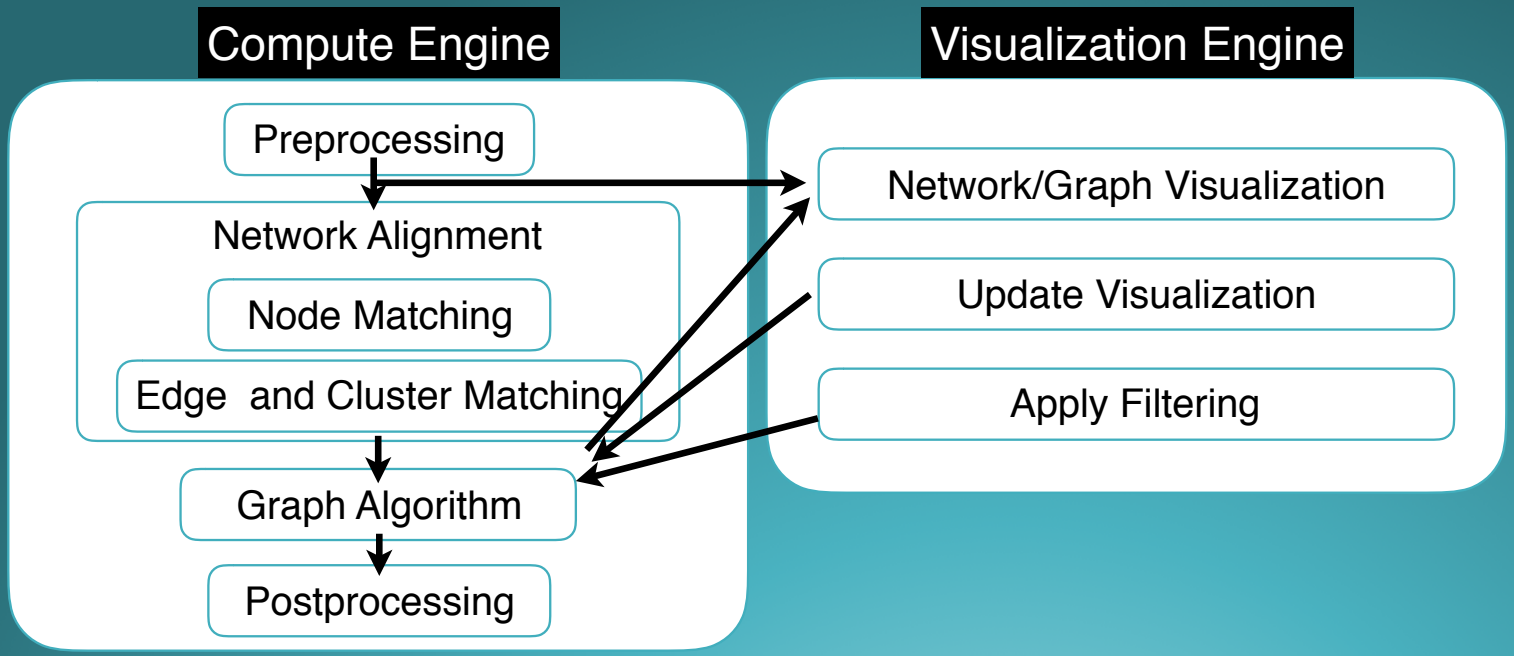
Conserved Subgraphs



Evidence: Maize-Rice Ancestor Shared Similar Gene Interaction Patterns 50-70 Millions of Years Ago

Ficklin & Feltus "Gene coexpression network alignment and conservation of gene modules between two grass species: maize and` rice." Plant Physiology 156:3 (2011)

G3NA-V Overview

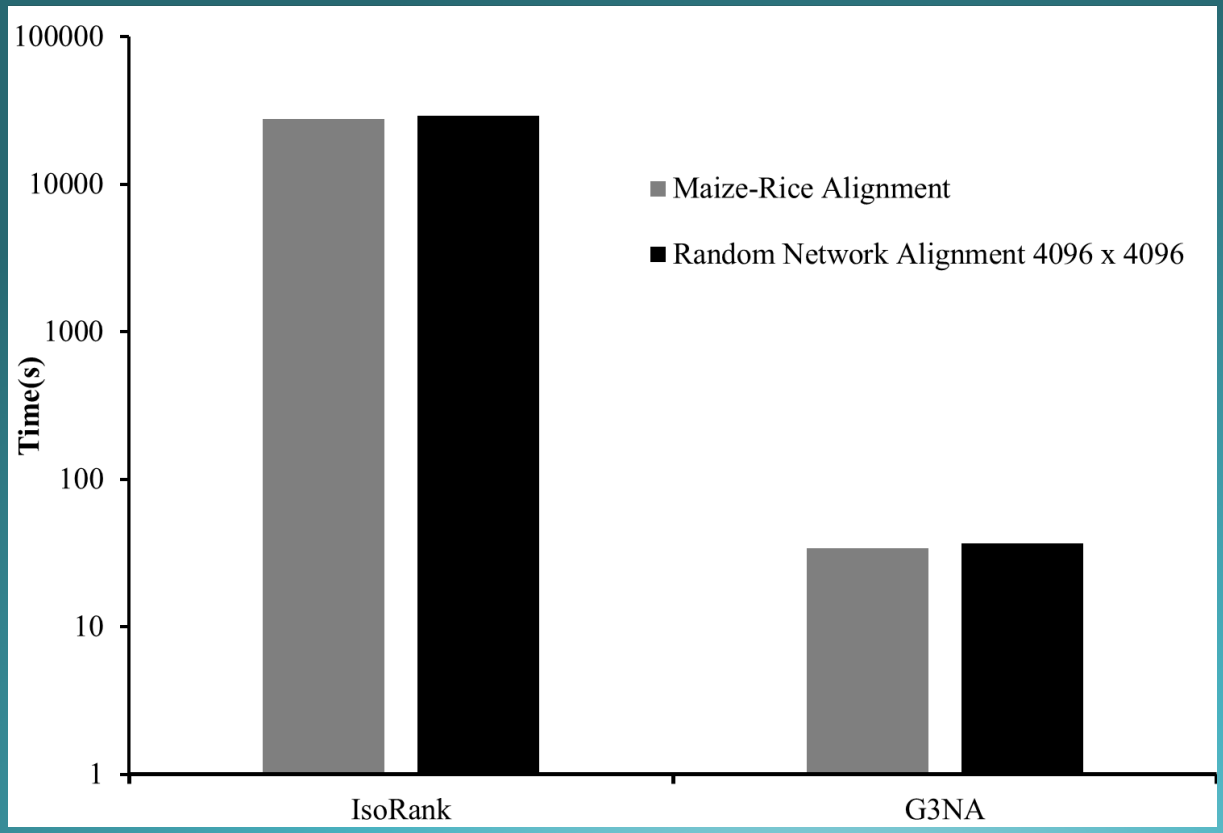


CLEMSON UNIVERSITY

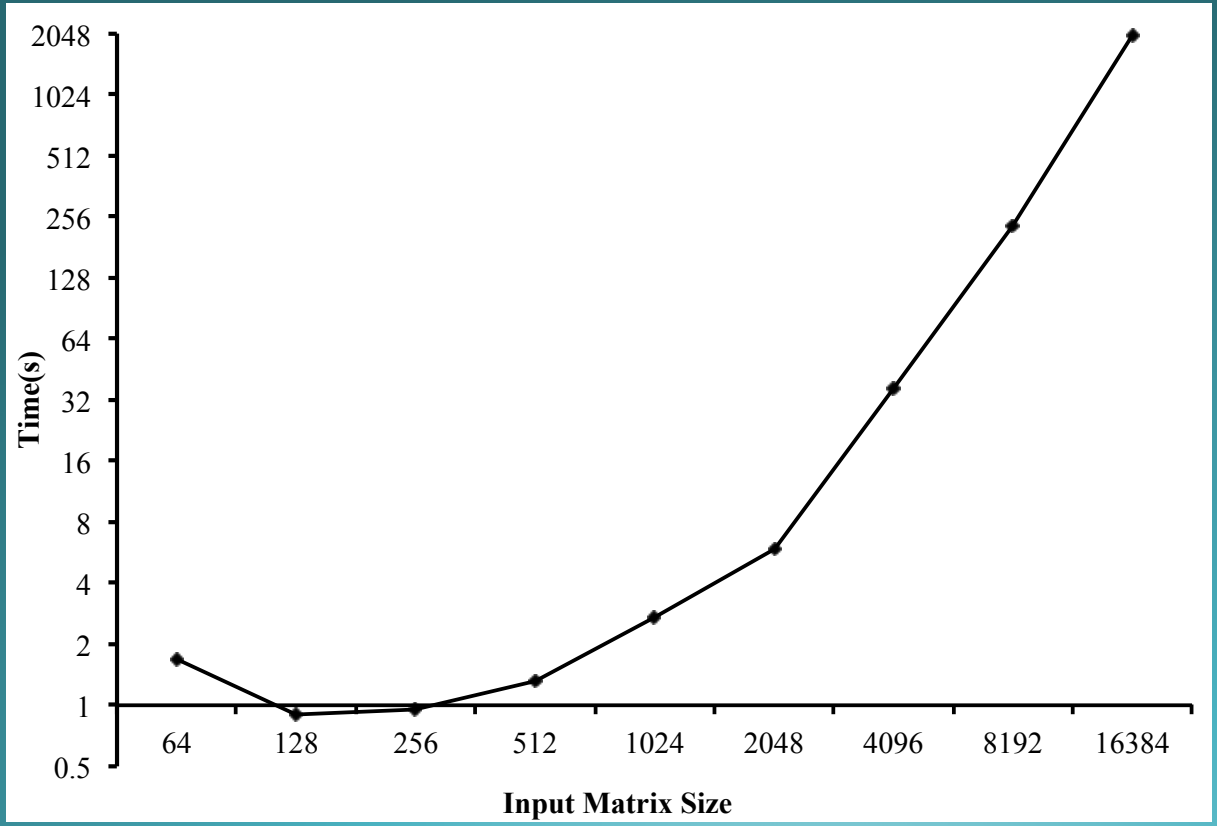
Compute Engine

- CUDA7 enabled global pairwise aligner GPU-enabled Global Gene Network Aligner (G3NA)
- CUDA enabled graph processing libraries
  - Thrust, Map-graph, etc.
- Use Multiple GPUs for alignment of multiple graph
- Utilize various graph algorithms
  - Clustering, Page Ranking, Filtering, Max-flow min-cut, etc.

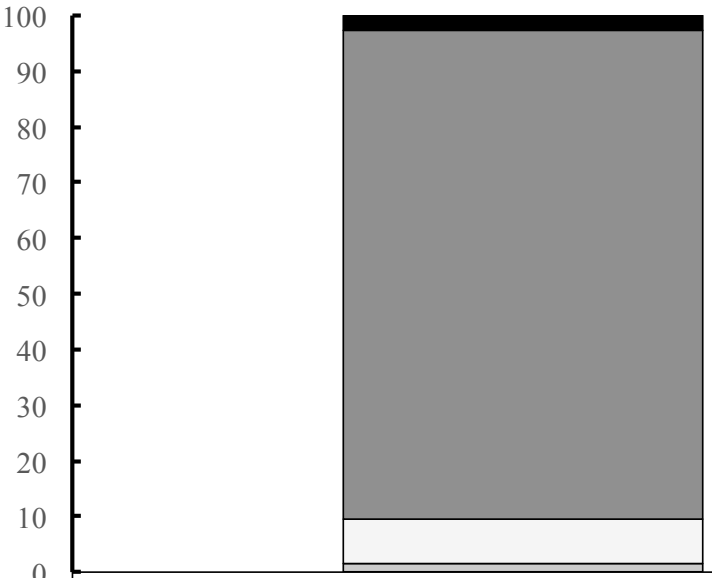
G3NA Comparative Result



G3NA Scaling Result



G3NA Profiling Overview



	G3NA Profile Percentage
Writeup %	2.55
Edge Matching %	88.03
Node Matching %	8.01
Read %	1.41

CLEMSON UNIVERSITY

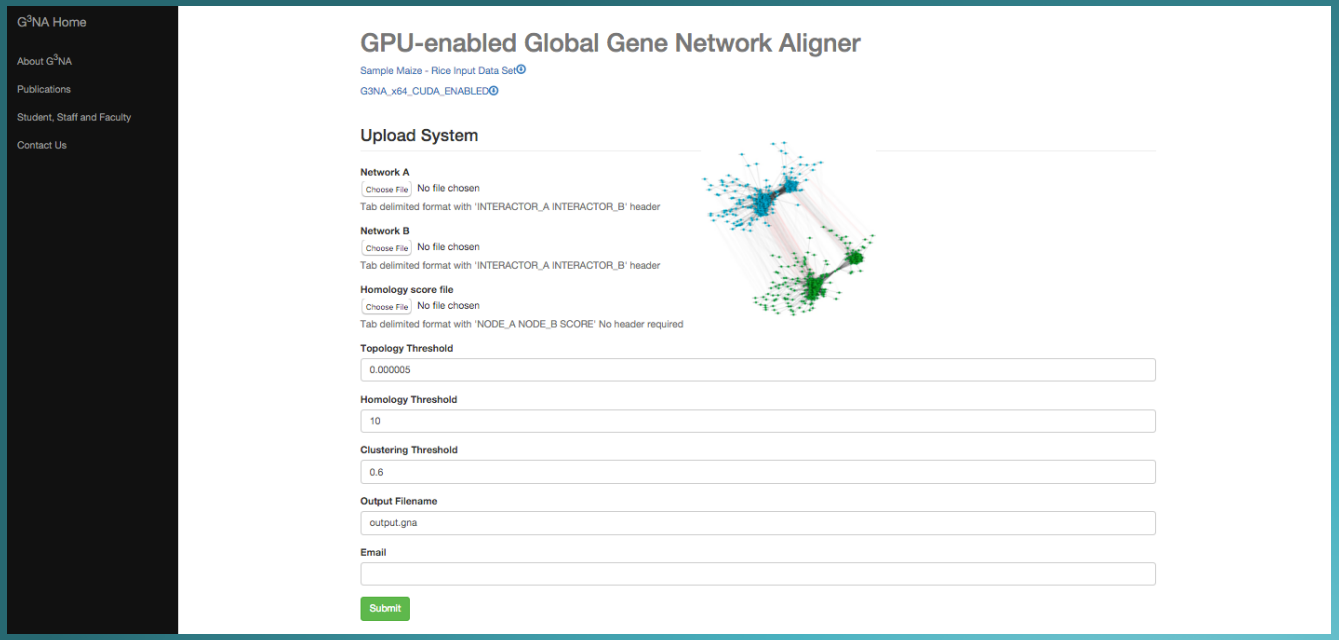
Visualization Engine

- Orientation and Visual Flexibility
- GPU enabled OpenGL and GLUI based visualization
- Support for Multiple View ports and File formats
- CUDA-based Layout algorithms for Graphs and Trees
- Dual GPU Support for Compute and Visualization

Enabling Systems Genetics using HPC

CLEMSON UNIVERSITY

- Enabling anonymous pairwise alignment using Palmetto Supercomputer at Clemson University
- Snappy overview visualization of alignment using WebGL



Palmetto Cluster

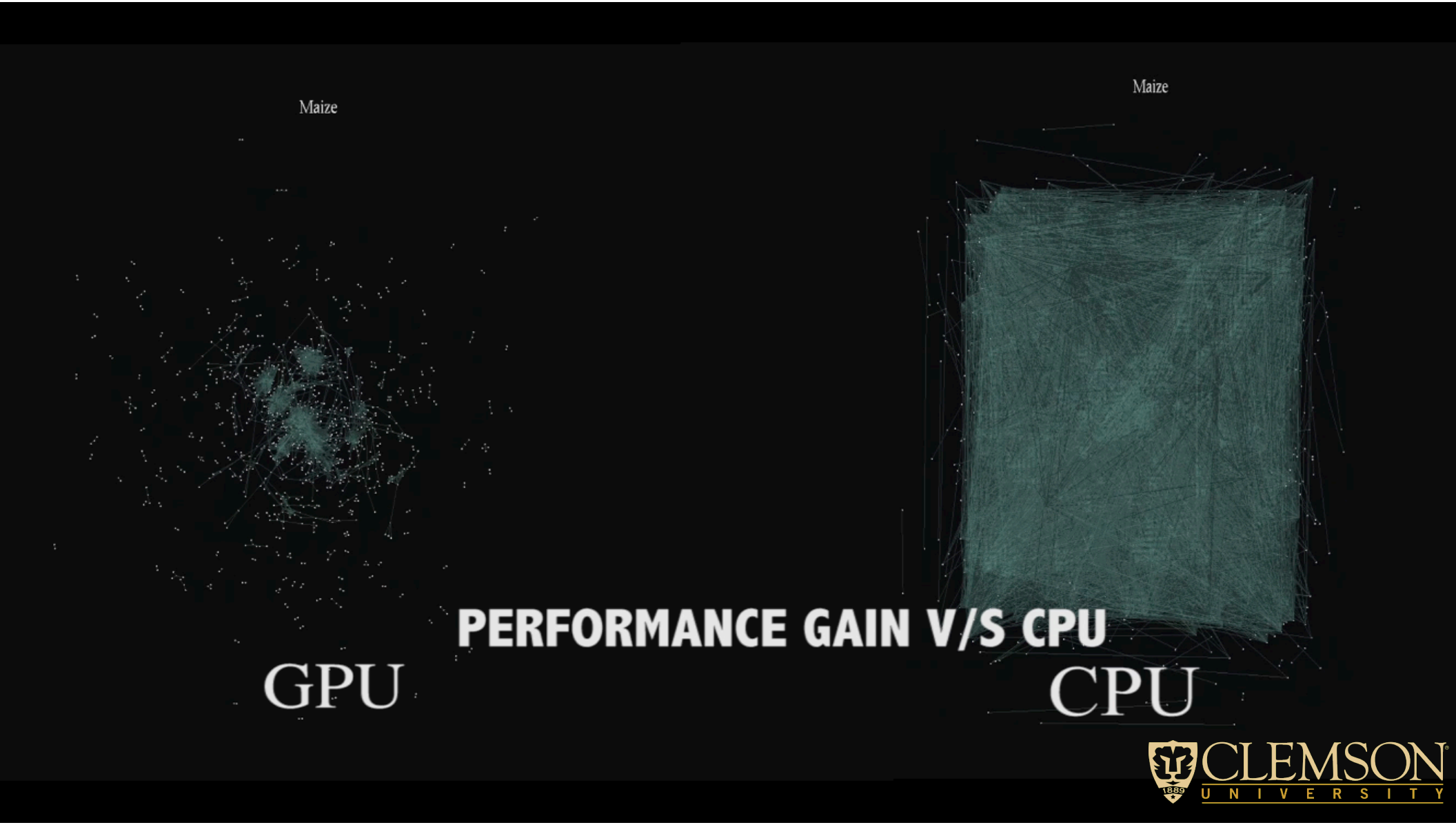
721.9 Teraflops

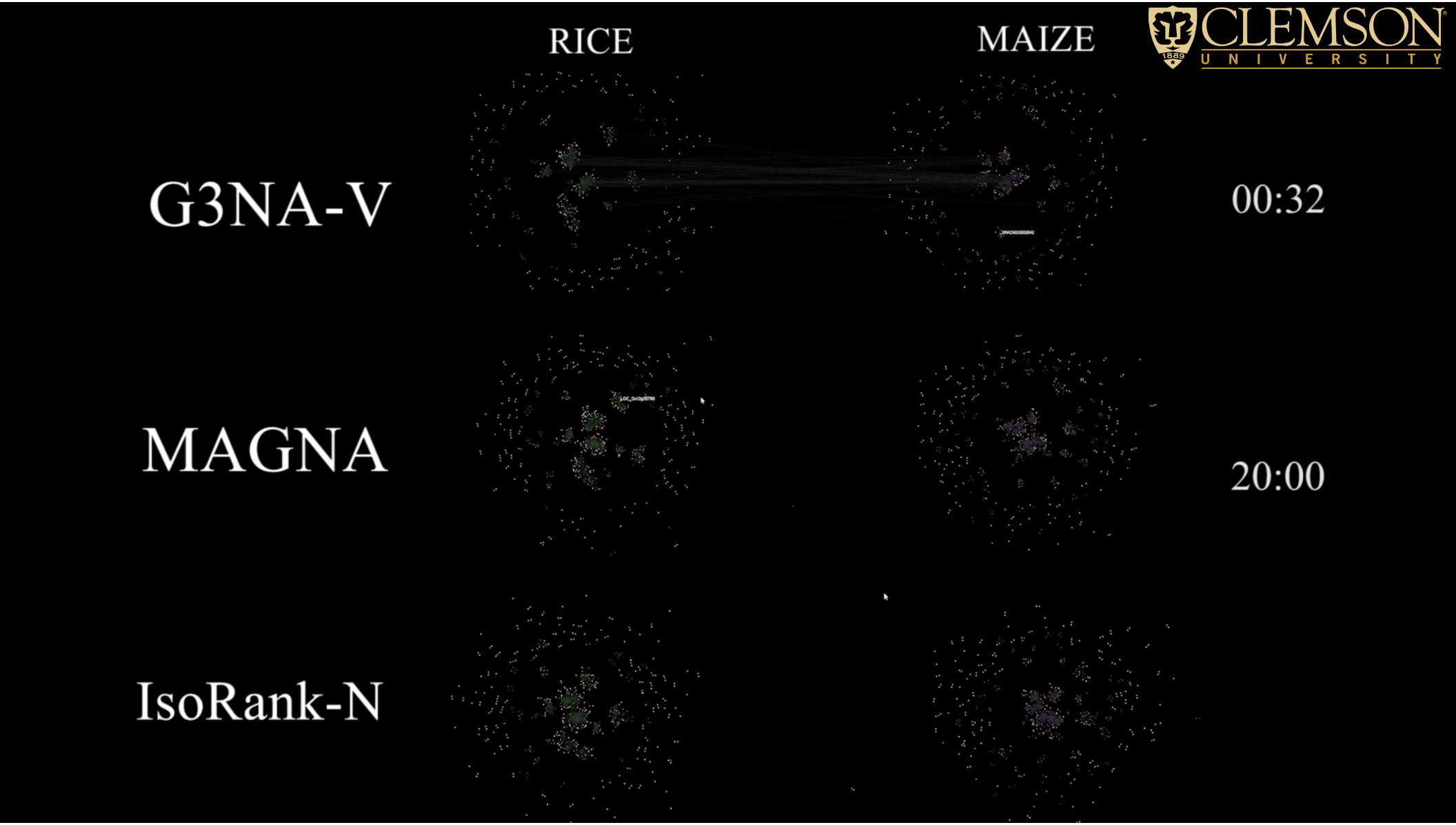
2021 Compute Nodes and 22,336 cores

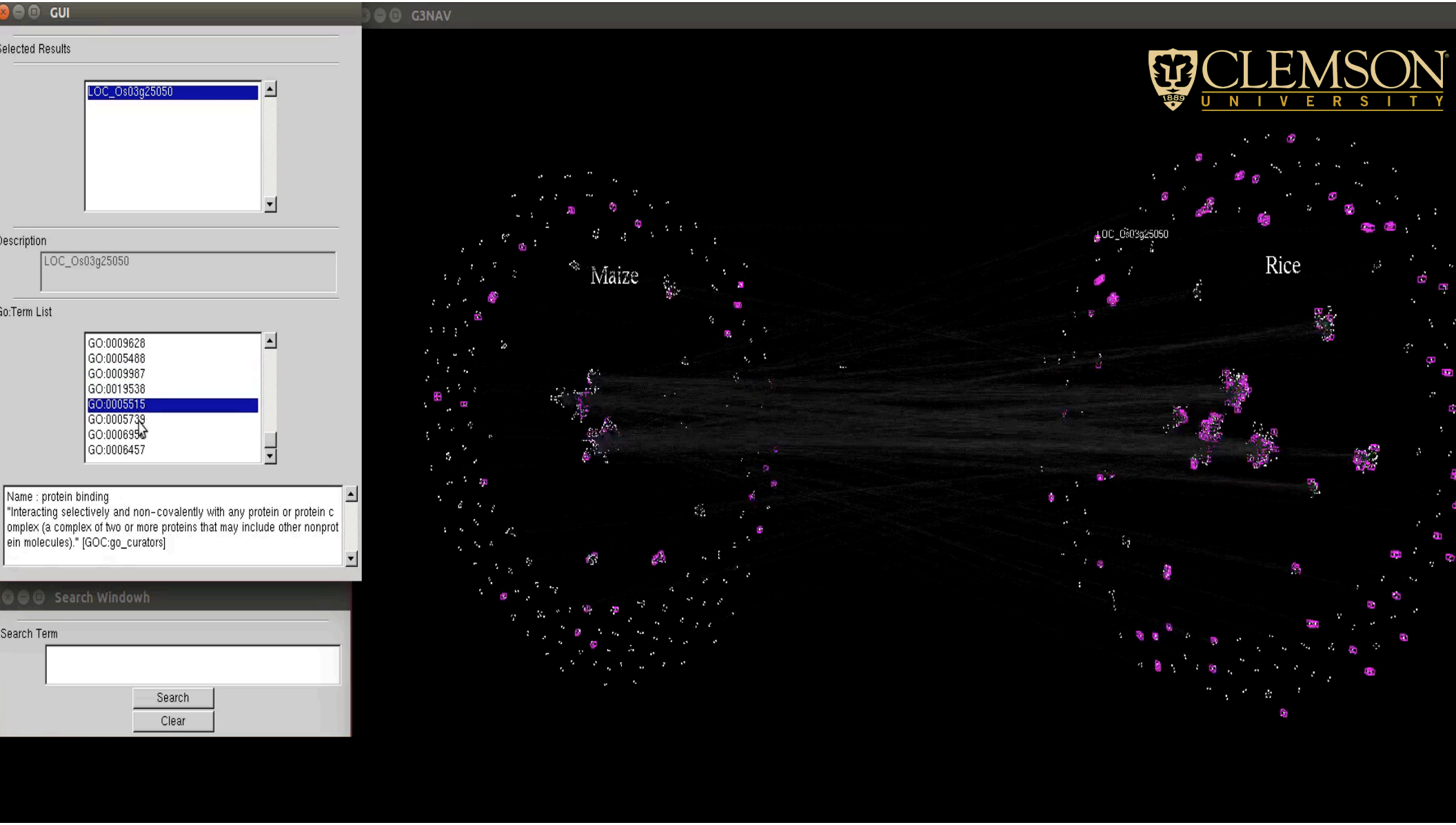
374 nodes with dual K20/K40 for acceleration and visualization

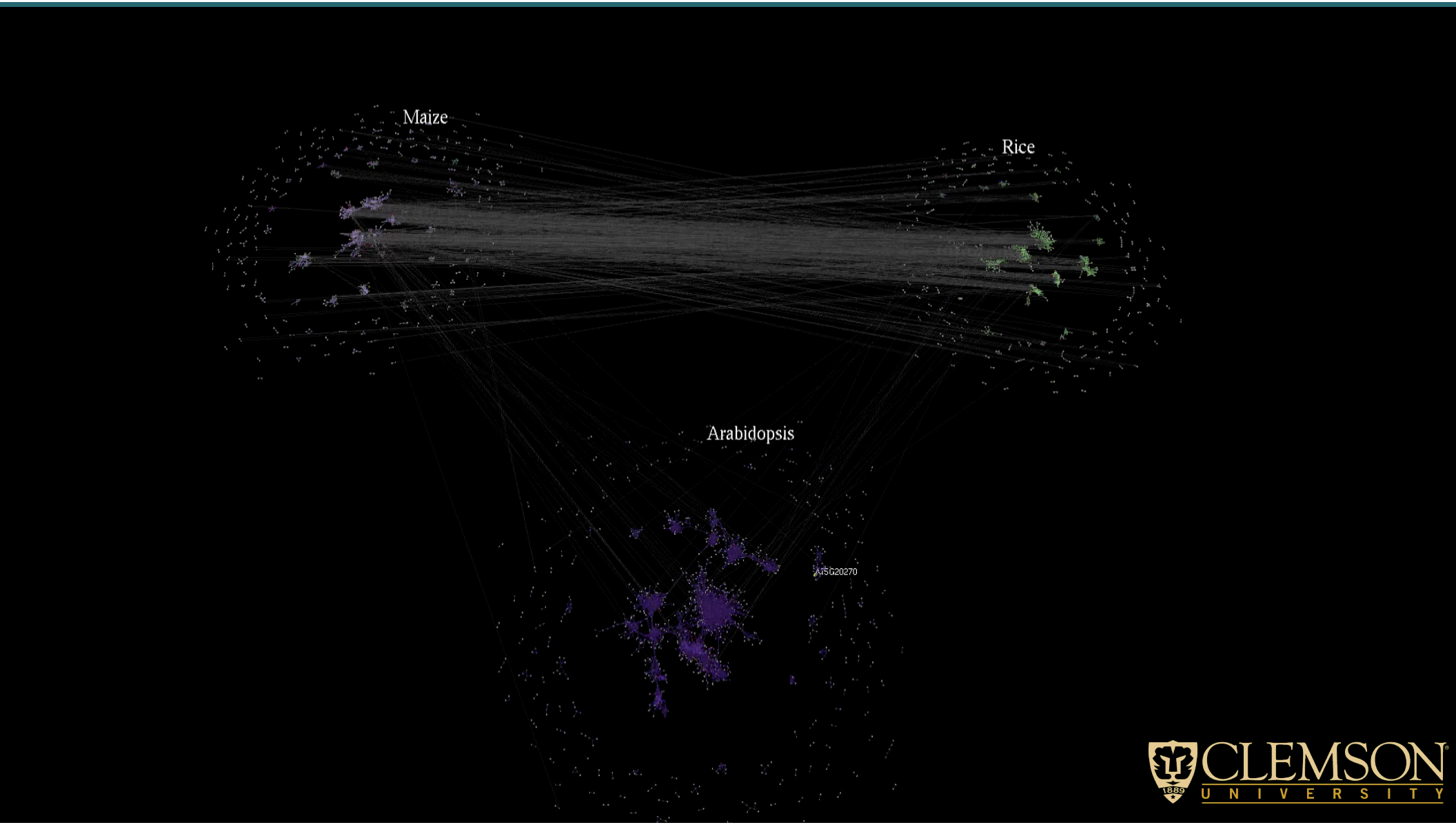
56Gbit interconnect

http://network.genome.clemson.edu


Maize
Maize
PERFORMANCE GAIN V/S CPU
GPU
CPU

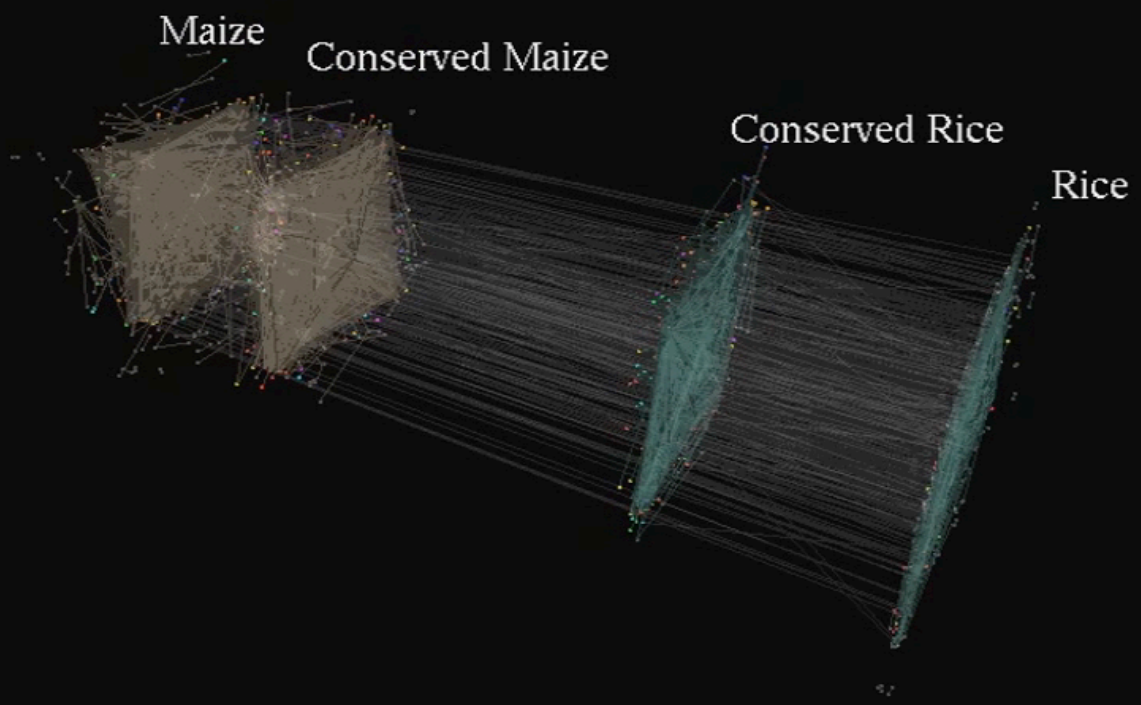

RICE
MAIZE
CLEMSON UNIVERSITY
G3NA-V
00:32
MAGNA
20:00
IsoRank-N


GUI
Selected Results
LOC_Os03g25050
Description
LOC_Os03g25050
Go:Term List
GO:0009628
GO:0005488
GO:0009987
GO:0019538
GO:0005515
GO:0005739
GO:0006950
GO:0006457
Name : protein binding
"Interacting selectively and non-covalently with any protein or protein complex (a complex of two or more proteins that may include other nonprotein molecules)." [GOC:go_curators]
Search Windowh
Search Term
Search
Clear
G3NAV
CLEMSON
UNIVERSITY
Maize
LOC_Os03g25050
Rice

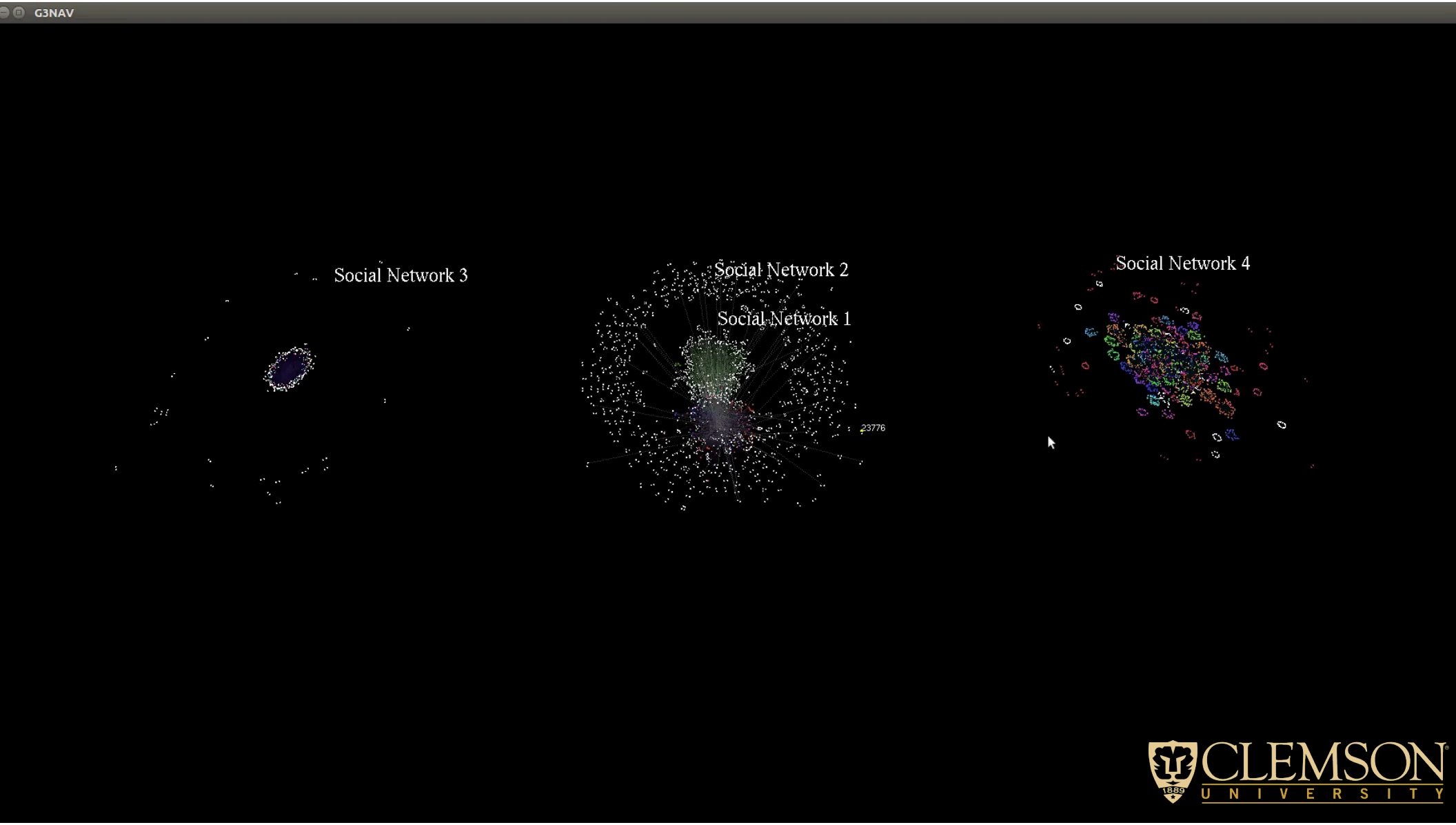


CLEMSON UNIVERSITY

STACKED CONSERVED SUB GRAPHS



CLEMSON UNIVERSITY


G3NAV
Social Network 3
Social Network 2
Social Network 1
23776
Social Network 4
CLEMSON
UNIVERSITY

Conclusion

- NVIDIA powered tool for alignment and visualizations of graphs and networks
- Support for ontologies and supporting information regarding graph and networks
- Supports various file formats
- Multi-view port based visualization

Future Work

- Support for other visualization methods such as matrix layout, etc
- Support for multiple ontology formats
- Auto layout mechanism
- Port to QT for alignment and visualization on mobile devices

It takes a Village…

Feltus Lab
- Will Poehlman (<PhD, G&B)
- Calvin Cox (<BSc, G&B)
- Zach Gerstner (<BSc, Microbiology)
- Leland Dunwoodie (<BSc, G&B)
- Brittany Rosener (<BSc, G&B)
- Nick Watts (Programmer, CCIT)

@Clemson
- Karan Sapra (ECE)
- Melissa Smith (ECE)
- KC Wang (ECE/CCIT)
- Ryan Izard (ECE)
- Juan Deng (ECE)
- Walt Ligon (ECE)
- Nick Mills (ECE)
- Brian Atkinson (ECE)
- Stephen Kresovich (G&B)
- Zach Brenton (G&B)
- Julia Frugoli (G&B)
- Josh Levine (CS)
- Chris Konger & Clemson Networking(CCIT)
- Becky Ligon (CCIT)
- Randy Martin (CCIT)
- Jim Bottum (CCIT)

- Stephen Ficklin Lab (WSU)
- Don Preuss & NLM Networking (NCBI)
- Joe Breen & Utah Networking (Utah)
- Jill Wegrzyn (UCONN)
- Meg Staton (UT-Knoxville)
- Dorrie Main (WSU)
- Tim Gilmanov (IU)
- Maciej Brodowicz (IU)
- Daniel Kogler (IU)
- Alireza Kheirkhahan (LSU)
- Adrian Serio (LSU)
- Hartmut Kaiser (LSU)
- Chris Branton (LSU)
- Florence Hudson (Internet2)

And it takes money…

- **"CAREER: Harnessing Hybrid Computing Resources in PetaScale Computing and Beyond"**
Source: National Science Foundation [1149644] (M Smith, PI)
- ***Genomic and Breeding Foundations for Bioenergy Sorghum Hybrids.***
Source: Plant Feedstock Genomics for Bioenergy [DE-FOA-000041] (S Kresovich, PI)
- ***MRI: Acquisition of a High Performance Computing Instrument for Collaborative Data-Enabled Science***.
Source: National Science Foundation [1228312] (A Apon, PI)
- ***"CC-NIE Integration: Clemson-NextNet"***
Source: National Science Foundation [1245936] (KC Wang, PI)
- **"Big Data Visualization REU"**
Source: National Science Foundation [1359223](V Byrd, PI)
- ***"Tripal Gateway: Platform for Next-Generation Data Analysis and Sharing"***
Source: NSF-DIBBS [1443040] (S. Ficklin, PI)
- ***"BIGDATA: F: DKM: Collaborative Research: PXFS: ParalleX Based Transformative I/O System for Big Data"***
Source: NSF-BIGDATA [1447771] (W. Ligon PI)
- ***"Big Data Analysis Tools for Agricultural Genomics***"
Source: Clemson University Experiment Station (USDA Hatch Project) [SC-1700492] (Feltus, PI)

Clemson Collaborators

CLEMSON UNIVERSITY

Open Science Grid