

*Exceptional service in the national interest*

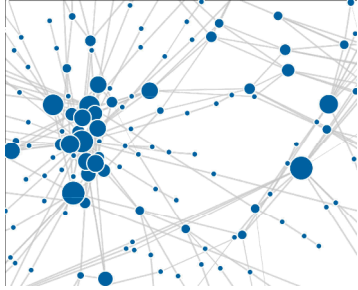


# Algorithms, Data Science, and Cyber Analytics at Sandia National Laboratories

Kevin Matulef (MIT PhD '09)  
MIT Tech Talk, October 2015

# Big Data at Sandia

## Cyber and Network Data



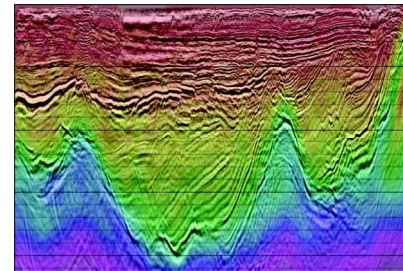
## Modeling and Simulation



## Remote Sensors and Imagery



## Seismic Data



# Big Research Questions

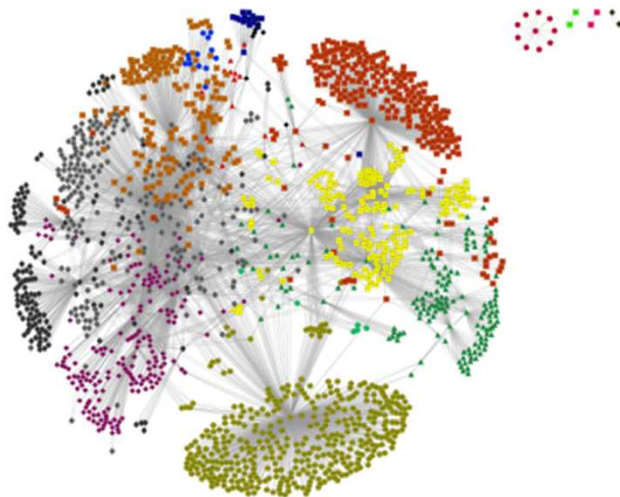
- Graph Modeling
  - How do you model *dynamic* networks?
  - How do you detect *communities or events*?
  - How can you *estimate* properties of networks?
- Machine Learning
  - How do you design ML algorithms against *adaptive* adversaries?
  - How can you find *succinct* data representations?
  - How can you identify the most *relevant* features/variables?
  - How can you *optimize* the computation?

# Some recent work...

- A Space Efficient Streaming Algorithm For Triangle Counting
  - [Madhav Jha](#), Seshadri Comandur, Ali Pinar. KDD 2013 **Best Student Paper!**
- A Scalable Generative Graph Model with Community Structure.
  - Tamara Kolda, Ali Pinar, Todd Plantenga, C. Seshadri. SIAM Journal of Scientific Computing, 2014.
- Finding the Hierarchy of Dense Subgraphs using Nucleus Decompositions.
  - [Erdem Sariyuce](#), C. Seshadhri, Ali Pinar. WWW 2015 **Best paper finalist!**
- A Framework for Practical Parallel Fast Matrix Multiplication
  - [Austin Benson](#) and Grey Ballard. PPOPP 2015.
- Catching the head, the tail, and everything in between: a streaming algorithm for the degree distribution
  - [Olivia Simpson](#), C. Seshadhri, Andrew McGregor. ICDM 2015.

## The Problem:

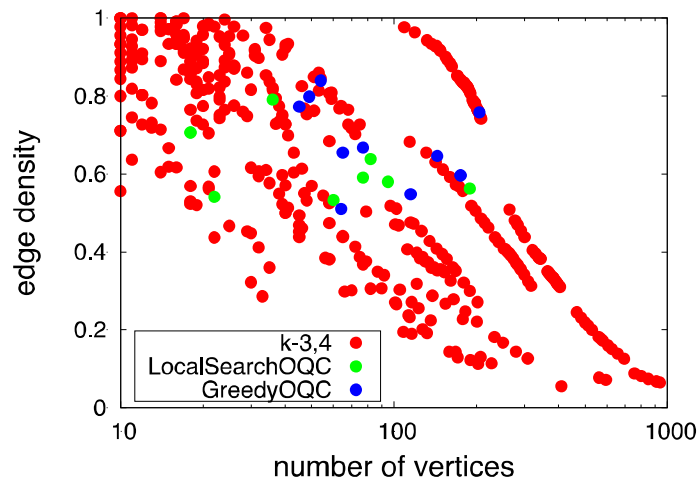
- Graphs in real world are SPARSE, but they contain many DENSE substructures (communities, related genes, etc.)
- Many different definitions of DENSE (K-cores, K-truss, etc.)
- Often hard to find them



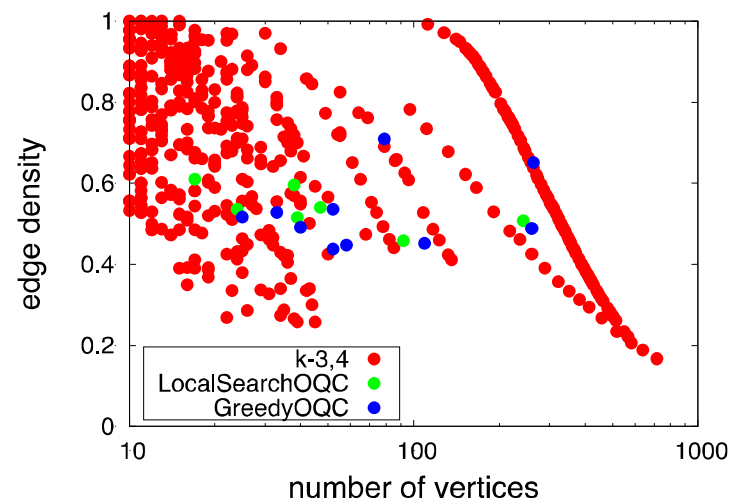
Source: <http://www.complexworld.net/virthulab/ongoing-projects-main>

## Main Contribution:

- A **new definition** of dense substructures, the *nucleus*, generalizes k-core, k-truss, and more
- A practical **algorithm** that finds a *hierarchy* of dense structures.
- Finds many more dense structures of all sizes.



Facebook



Soc-Epinions

# A Scalable Generative Graph Model with Community Structure [KPS14]

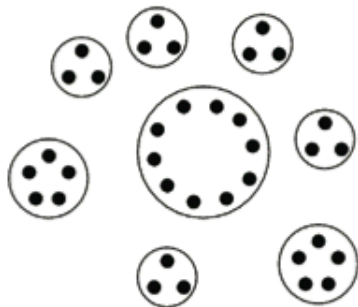
## The Problem:

- Real world social networks have unique, common properties
  - Power law degree distribution
  - High “clustering coefficient” (fraction of your friends that are friends with each other)
- Simple random graph models do not usually have these!
  - Erdos-Renyi does not generate power law degree distribution
  - Chung-Lu model does not generate high clustering coefficient

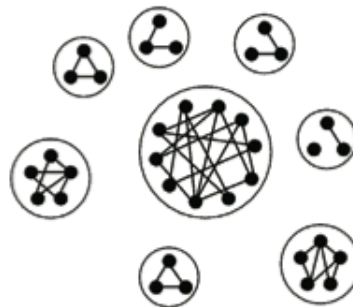
# A Scalable Generative Graph Model with Community Structure [KPS14]

## Main Contribution:

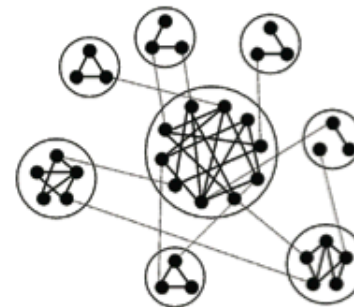
- A **new generative random graph model** that will match any given degree distribution and clustering coefficients, in expectation
- A **scalable, parallel implementation** that can generate graphs with billions of edges.
- Can be used a generator for benchmarking graph algorithm



(a) Preprocessing:  
Distribution of nodes into  
affinity blocks.



(b) Phase 1: Local links  
within each affinity block.



(c) Phase 2: Global links  
across affinity blocks.

# Some current projects...

- Modeling *temporal* graphs streams
- Graph generators for *benchmarking* streaming algorithms
- *Anomaly detection* in network graph streams
- *Sublinear sampling* methods applied to networks
- *Deep learning* on network data
- *Tensor methods* for compression of simulation data

and many more!

# People



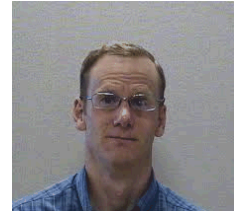
Tammy Kolda



Ali Pinar



Philip  
Kegelmeyer



Jon Berry



Cindy Philips



Kevin Matulef



Erdem Sariyuce



Grey Ballard

Looking for more full-time staff, postdocs and interns!