Graph Clustering Approaches using Quantum Annealing

QUBITS 2018

D-Wave Users Conference

S. M. Mniszewski, smm@lanl.gov
C. F. A. Negre, cnegre@lanl.gov
H. Ushijima-Mwesigwa, hushiji@g.clemson.edu

September 24-27, 2018

Operated by Los Alamos National Security, LLC for the U.S. Department of Energy's NNSA

LA-UR-18-28859

Introduction

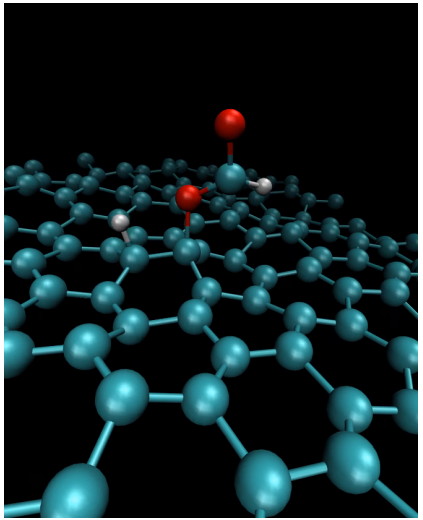
- **Motivated by graph-based methods for quantum molecular dynamics (QMD) simulations**
- **Explored graph partitioning/clustering methods and implementations that run on the D-Wave**
  - *Initially, 2-partitioning, 2-clustering*
  - *Recursive methods*
  - *k*-Concurrent methods
  - Iterative multi-level graph partitioning with D-Wave refinement
- **Used *sapi* and hybrid classical-quantum *qbsolv* software tools**
- **Demonstrated "proof of principle" results on benchmark graphs, example graphs and electronic structure graphs**
- **Results are shown to equal or out-perform current "state of the art" methods**

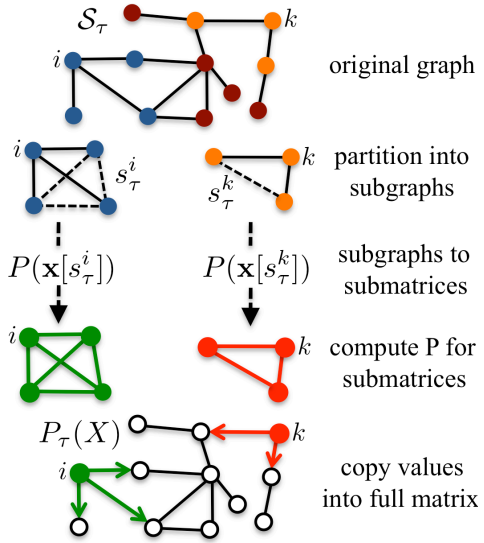
Graph partitioning/clustering implementations on the D-Wave.

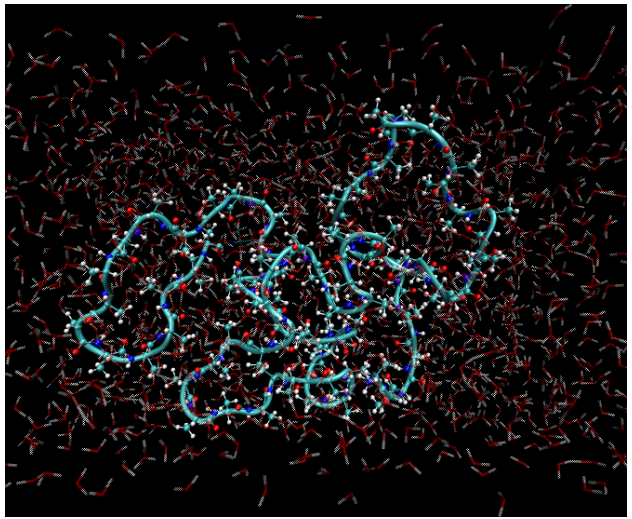
Motivation

- **Multi-year research program on Next Generation Quantum Molecular Dynamics**
- **Quantum-based models capture the making and breaking of covalent bonds, charge transfer between species of differing electronegativities, and long-range electrostatic interactions - reactions**
- **Graph-based methods for quantum molecular dynamics (QMD) simulations**

  A. M. N. Niklasson et al, Graph-based linear scaling electronic structure theory, *J. Chem. Phys.* **144**, 234101 (2016).
- **Density matrix generated each timestep from many small sub-matrices (or sub-graphs)**
- **Shown to be equivalent to traditional methods (ex. diagonalization)**





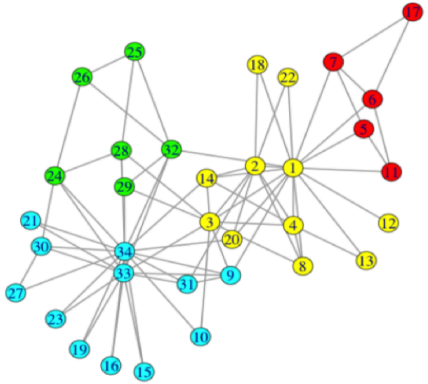


Graph Clustering - Description

- **Definition:** Unsupervised machine learning data classification
  - Identify structure based on a similarity measure

    Given a graph G = (V, E)

    V ~ nodes, E ~ edges (possibly weighted)
- **Goal:** Partition V into up to k clusters/communities
- **NP-hard:** Uses heuristics and approximation algorithms
- **Methods:** Modularity, K-Means, Spectral, Hierarchical, etc.
- **Approaches:** 2-clustering, recursive bisection, k-concurrent, multi-level with refinement



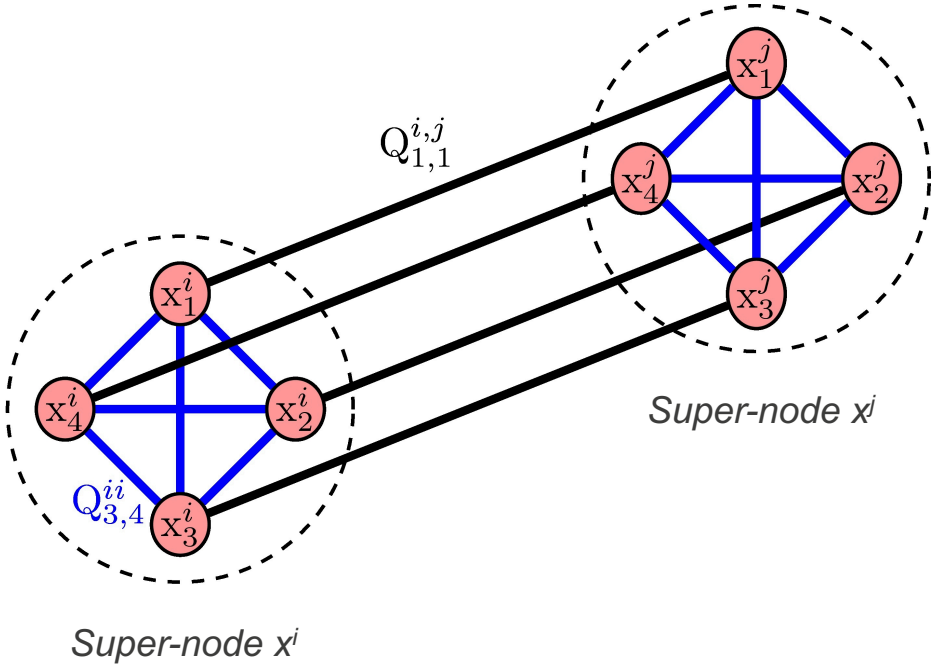
- **Applications:**
  - Graph-based data decomposition for distributed HPC simulations
  - Document analysis, text classification
  - Image segmentation
  - Networks – computer, communications, physical, VLSI circuit, telephone
  - Load balancing - minimize total communication between processors
  - Biosystems, social networks, cyber networks

Graph Clustering using Community Detection

- **Using modularity or community network analysis for natural clusters**
  - M. E. J. Newman, Modularity and community structure in networks, 2006, PNAS, vol. 103, no. 23, 8577-8582.
- **Identifying communities in social networks**
- **Identifying secondary structures in proteins**
  - I. Rivalta, M. M. Sultan, N.-S. Lee, G. A. Manley, J. P. Loria, V. S. Batista, Allosteric pathways in imidazole glycerol phosphate synthase, *PNAS,* vol. 109, no. 22, pp. 1428-1436 (2011).
- **Maximize modularity metric:** $Q = 1/4m\ s^T B s$
- **2-Clustering fits naturally on D-Wave machine, no reformulation required**
- **New *k*-Concurrent community detection formulation**

k-Concurrent Clustering – Multiple parts in parallel

- Cluster into *k* parts in parallel
- Uses super-node concept
- Unary encoding
- *k* logical qubits per vertex
- New formulation requires a *kN x kN* QUBO
- Results in 1 of *k* qubits set on for each vertex
- Similar to graph coloring problem
- Useful for graph partitioning and community detection



Super-node Concept

QUBO/Ising Modularity Clustering Formulation

- **Construct adjacency matrix *A***

$$A_{ij} = \begin{cases} 0, & \text{if } i = j \\ w_{ij}, & \text{if } i \neq j \end{cases}$$

Where w_{ij} is the weight on the edge between nodes i and j

- **Construct modularity matrix *B***

$B_{ij} = A_{ij} - g_i g_j / 2m = A_{ij} - g_i g_j / \sum(g_i)$, where g_i is the degree of node i

- ***Maximize objective for the optimal modularity for QUBO or Ising***

Ising: $\max_s Q(s) = s^T B s$, where $s_i \in \{-1, 1\}$ or

QUBO: $\max_x Q(x) = x^T B x$, where $x_i \in \{0, 1\}$

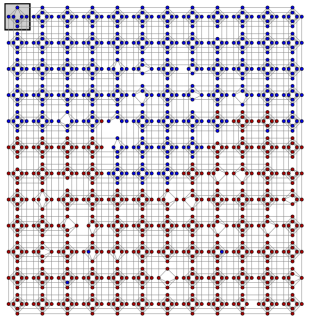
- ***Calculate modularity metric to evaluate a set of clusters***
  - $Q_{metric} = \sum(B_{ij} \phi(c_i, c_j)) / 2m$, where $\phi(c_i, c_j)$ is 1 when $c_i = c_j$, else 0

Data and Examples

- **Determine at most k communities, maximizing modularity in parallel**
- **Similar formulation as k-concurrent graph partitioning**
  - *kN x kN* QUBO
  - Constraint: each node in only 1 community
  - Penalty constants, β=1, γ dependent on data
- **Using *qbsolv* on D-Wave 2X and 2000Q**
- **Examples and results**
  - Newman's benchmark social networks (*www-personal.umich.edu/~mejn/netdata/)*
  - Alex Arena's network data sets (*deim.urv.cat/~alexandre.arenas/data/welcome.htm)*
  - The Koblenz Network Collection (konect.uni-koblenz.de)
  - Molecular electronic structure – Phenyl dendrimer, Peptide 1 aft
  - Protein structure – IGPS enzyme
- **Comparison metrics** - # of communities and modularity
- **Next, apply to more bio-systems and social networks**

Using D-Wave's *qbsolv on 2X and 2000Q*

- **Hybrid classical-quantum approaches are required for problems that are too large for embedding into D-Wave's *Chimera* Graph.**
- **D-Wave's *qbsolv* is a tool that solves large quadratic unconstrained binary optimization (QUBO) problems by partitioning into subproblems for execution on the D-Wave quantum annealer.**
- **A QUBO is generated for the full problem to be used as input.**
- **SubQUBO size:** 46 for D-Wave 2X and 64 for D-Wave 2000Q.



- **Call directly through D-Wave Ocean API**

```
response = QBSolv().sample(bqm_qubo, solver=EmbeddingComposite(DWaveSampler())
```

- **Or from the command line**

```
qbsolv –i example.qubo –o (-m) dwave_output.out
```

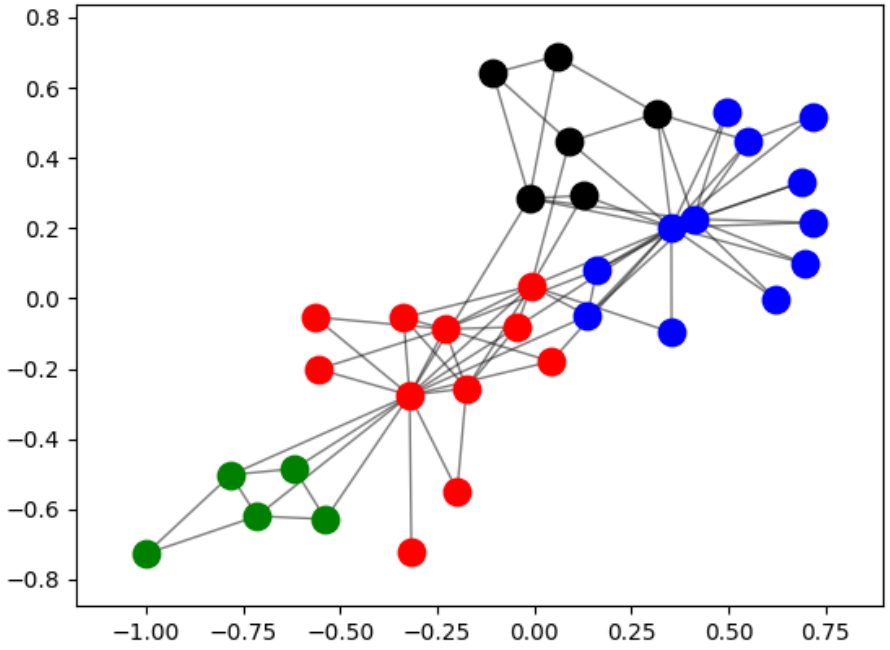
- **Resulting bitstring of 0's and 1's is translated based on the optimization problem's representation**

Benchmark Social Network – Karate Club

- **Social network of friendships between 34 members of a karate club at a US university in the 1970s.**
  - W. W. Zachary, An information flow model for conflict and fission in small groups, *Journal of Anthropological Research* 33, 452-473 (1977).
- **Matches best known results for communities and modularity (Blondel et al.)**

# Communities	Modularity
2	0.3717949
3	0.4020381
4	0.4197896

Karate club graph (N = 34, E = 78)
4 communities, modularity = 0.419789

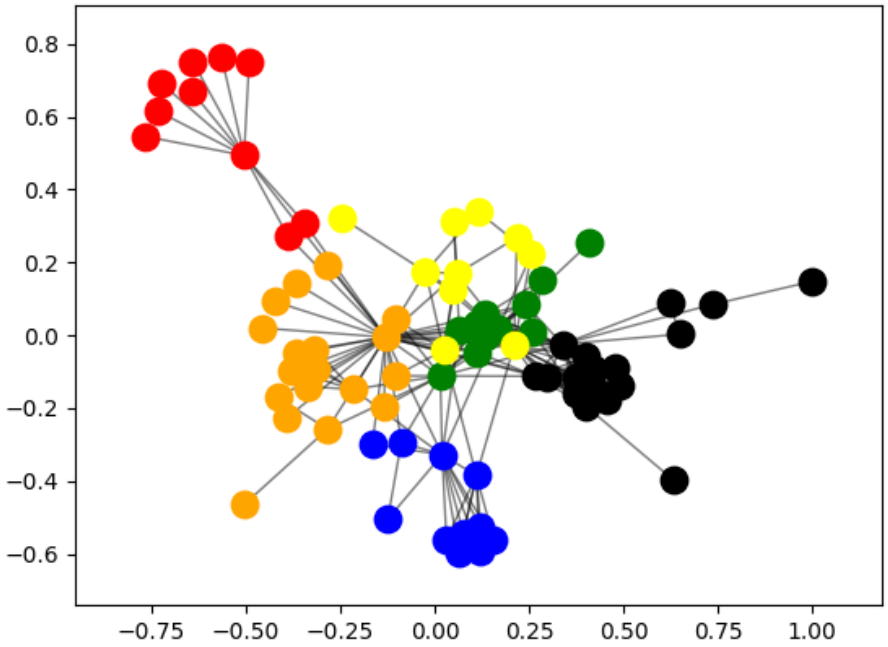


Benchmark Social Network – Les Miserables

- **Coappearance network of characters in the novel *Les Miserables*.**
  - D. E. Knuth, *The Stanford GraphBase: A Platform for Combinatorial Computing*, Addison-Wesley, Reading, MA (1993).
- **Matches best known results for communities and modularity**

# Communities	Modularity
2	0.3827887
3	0.4973274
4	0.5428343
5	0.5555388
6	0.5586134

Les Miserables graph (N = 77, E = 254)
6 communities, modularity = 0.5586134

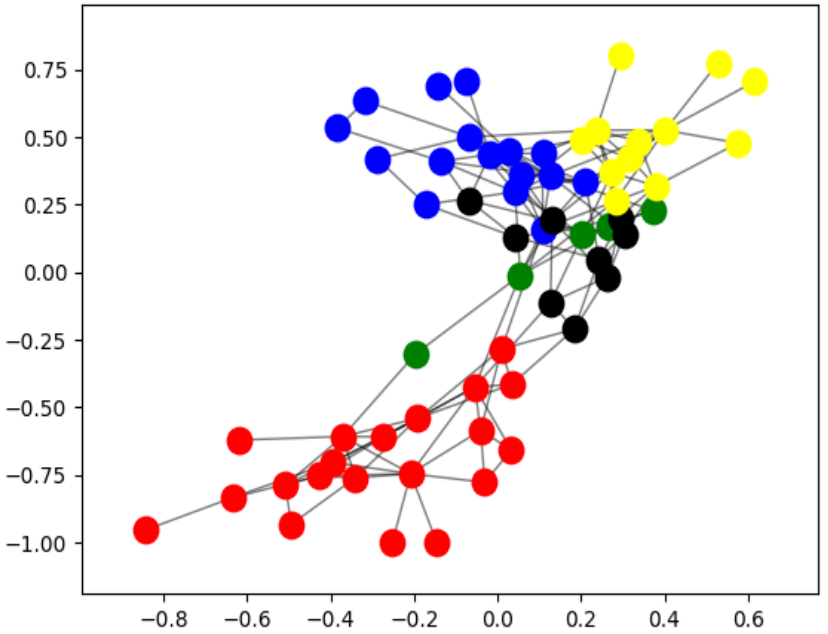


Benchmark Social Network – Dolphins

- **An undirected social network of frequent associations between 62 dolphins in a community living off Doubtful Sound, New Zealand.**
  - D. Lusseau, K. Schneider, O. J. Boisseau, P. Haase, E. Slooten, and S. M. Dawson, *Behavioral Ecology and Sociobiology* 54, 396-405 (2003).

# Communities	Modularity
2	0.4027333
3	0.4941853
4	0.5267987
5	0.5285194

Dolphins graph (N = 62, E = 159)
5 communities, modularity = 0.5285194



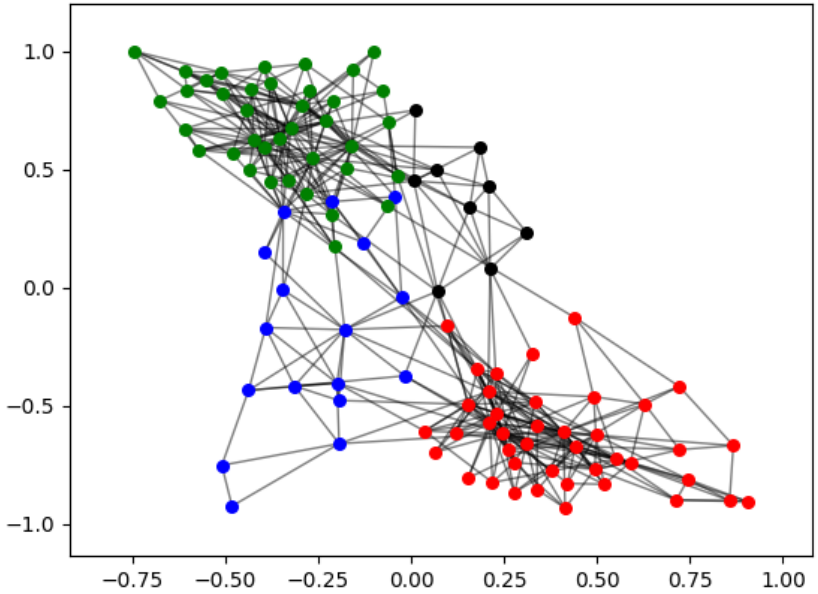
Matches best known results for communities and modularity.

Benchmark Social Network – Political Books

- **Book co-purchasing network. Nodes represent books about US politics and edges represent frequent co-purchasing of books.**
  - M. E. J. Newman, Modularity and Community Structure in Networks, *PNAS*, vol. 103, no. 23, 8577-8582 (2006).

# Communities	Modularity
2	0.45687
3	0.52207
4	0.52555

Political books graph (N = 105, E = 441)
4 communities, modularity = 0.52555



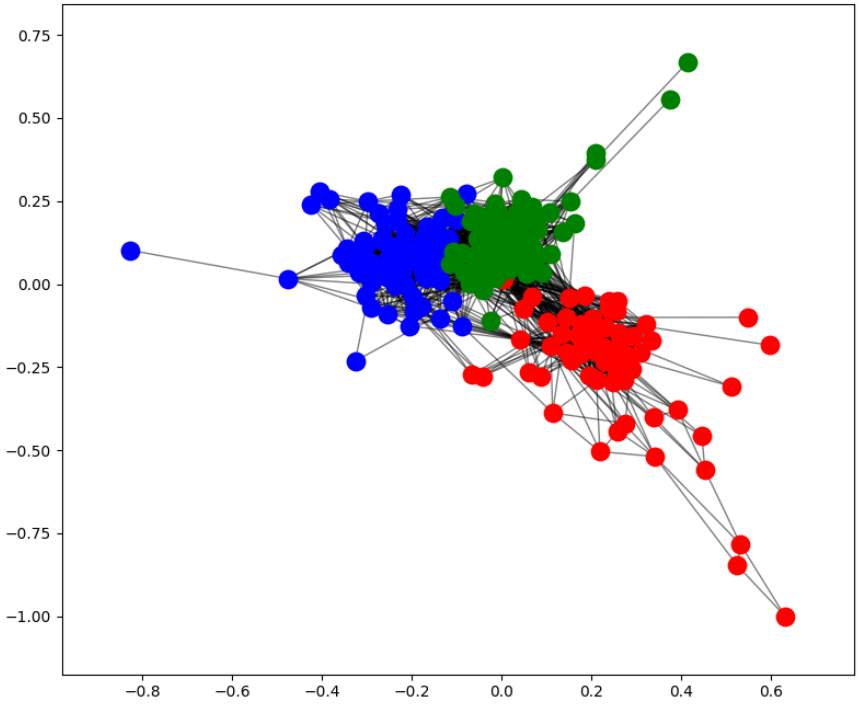
Matches best known results for communities and modularity.

Benchmark Social Network – Jazz Musicians Network

- **Collaboration network of Jazz musicians.**
  - P.Gleiser and L. Danon, Community Structure in Jazz, Adv. Complex Syst.6, 565 (2003).
- **Comparable results for communities and modularity**
- **Race and recording location determine communities**

# Communities	Modularity
2	0.3206093
3	0.4444694

Jazz graph (N = 198, E = 2742)
3 communities, modularity = 0.4444694

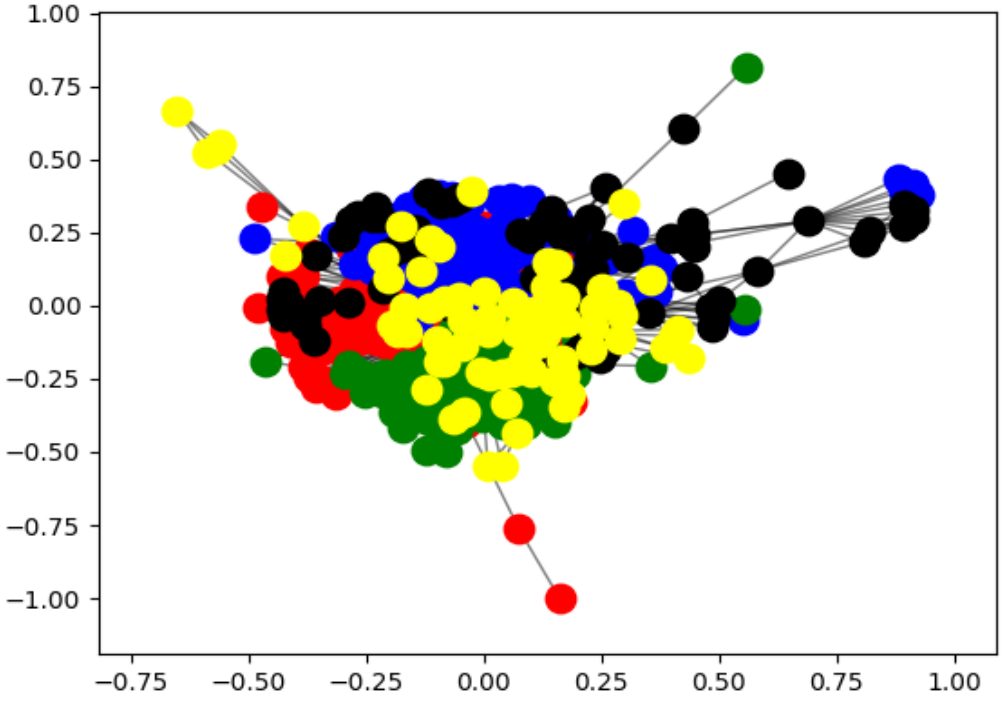


Benchmark Social Network – C. elegans metabolic network

- **Metabolic network of the nematode C. elegans. Nodes are proteins, edges are interactions.**
  - J. Duch and A. Arenas, Community identification using Extremal Optimization, Physical Review E , vol. 72, 027104, (2005).
- **Comparable results for communities and modularity.**

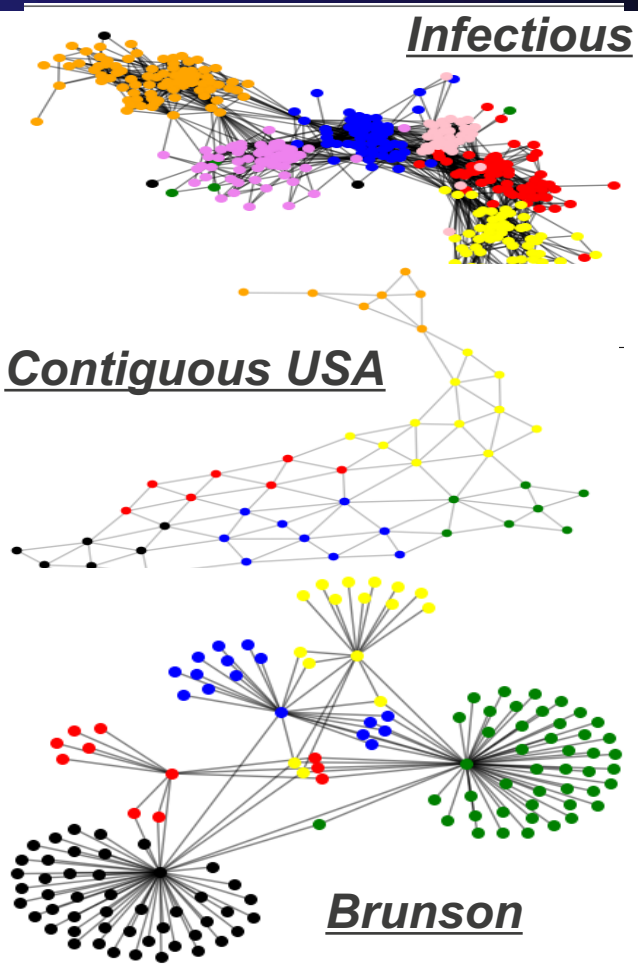
# Communities	Modularity
2	0.3199541
3	0.3929653
4	0.4091737
5	0.4172796

C. elegans graph (N = 453, E = 2040)
5 communities, modularity = 0.4172798



Koblenz Networks – k-Concurrent Clustering

Dataset	Nodes	Edges	Comms	Mod
Cont. USA (spatial)	49 state	107 border	6	0.5970
Infectious (social net)	410 visitor	2765 contact	8	0.7067
Contact (cell phone contacts)	274 person	2124 phone contact	4	0.1315
Brunson (affiliation network)	136 person	159 membership	5	0.5795
Moreno_oz (friendship)	217 person	1839 friend	6	0.4282

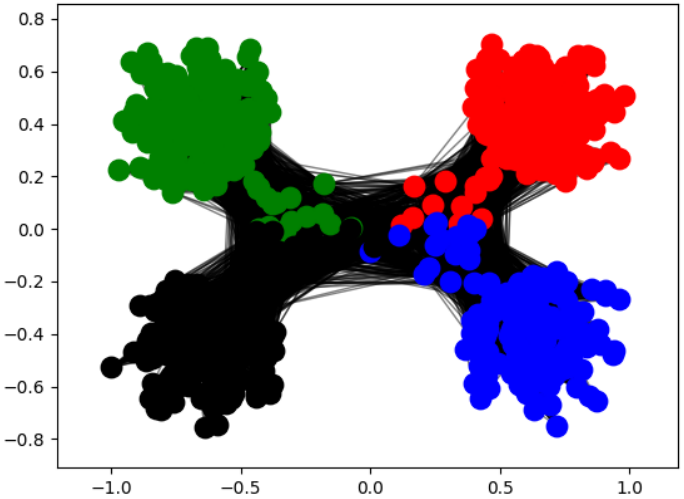


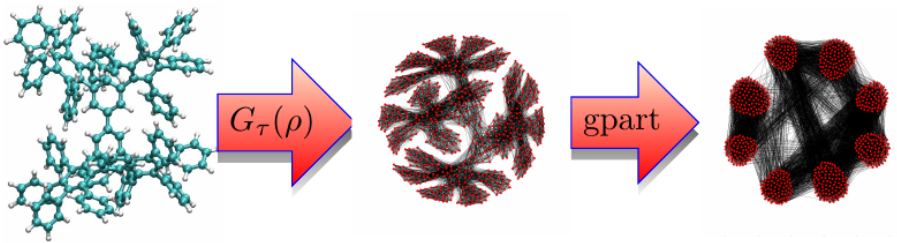
Molecule electronic structure – Phenyl Dendrimer

- **Molecular electronic structure graph for Phenyl Dendrimer protein.**
  - H. N. Djidjev, G. Hahn, S. M. Mniszewski, C. F. A. Negre, A. M. N. Niklasson, V. B. Sardeshmukh, Graph partitioning Methods for Fast Parallel Quantum Molecular Dynamics, SIAM Workshop on Combinatorial Scientific Computing 2016 (CSC16).
- **Density matrix used for adjacency**

# Communities	Modularity
2	0.4773614
3	0.5724155
4	0.6651407

Phenyl dendrimer graph (N = 730, E = 31877)
4 communities, modularity = 0.6651407

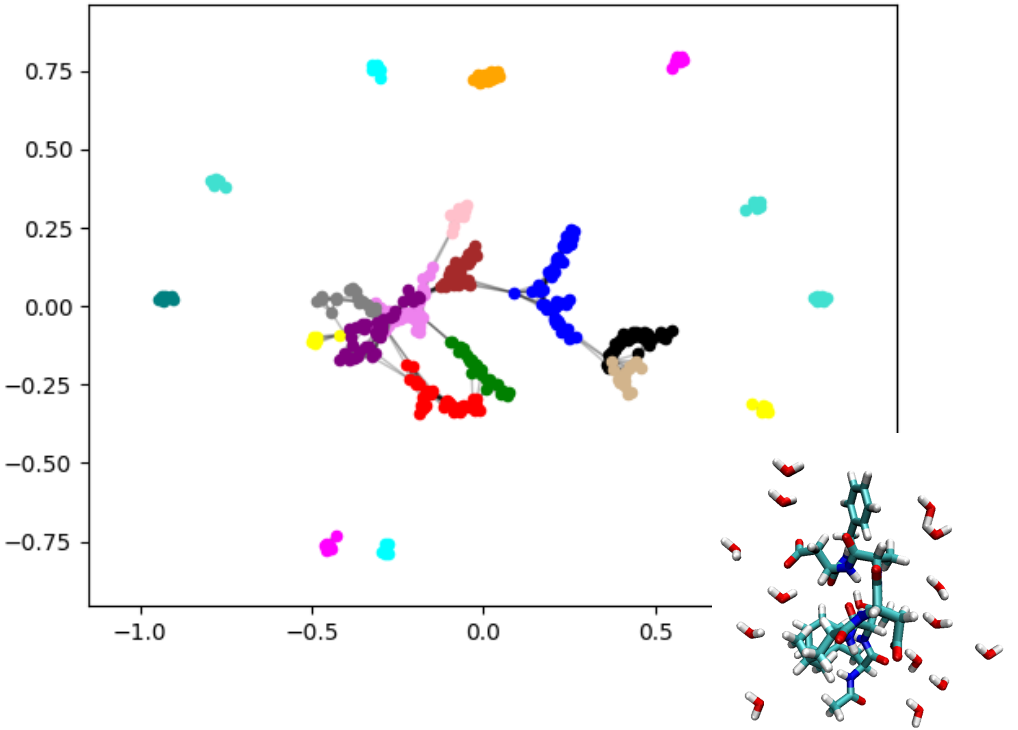




Molecule electronic structure – Peptide 1 aft + H_2O

- **Molecular electronic structure graph for Peptide 1 aft protein.**
  - H. N. Djidjev, G. Hahn, S. M. Mniszewski, C. F. A. Negre, A. M. N. Niklasson, V. B. Sardeshmukh, Graph partitioning Methods for Fast Parallel Quantum Molecular Dynamics, SIAM Workshop on Combinatorial Scientific Computing 2016 (CSC16).
- **Amino acid sequence**
  - PHE-ASP-ALA-ASP-LEU-THR-PHE
    - (PHEnylalanine, ASPartic acid, LEUcine, THyRosine)
- **Solvated in water**
- **LATTE QMD after SCF convergence**
- **Orbitals as nodes**
- **Density matrix used for adjacency**

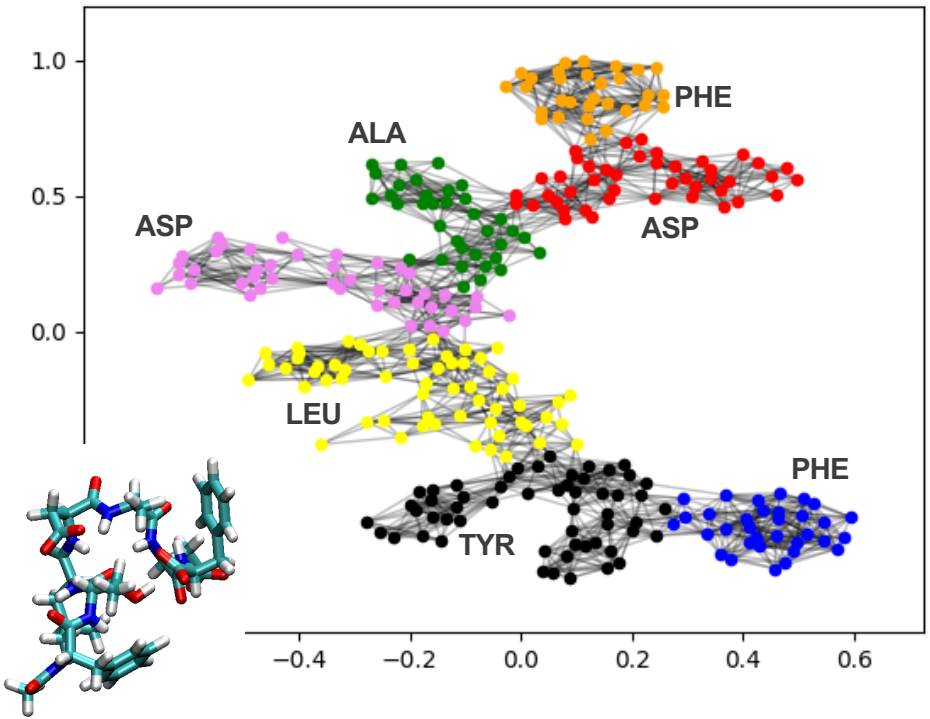
Peptide 1 aft graph (N = 384, E = 2217)
16 communities, modularity = 0.875424



Molecule electronic structure – Peptide 1 aft (No water)

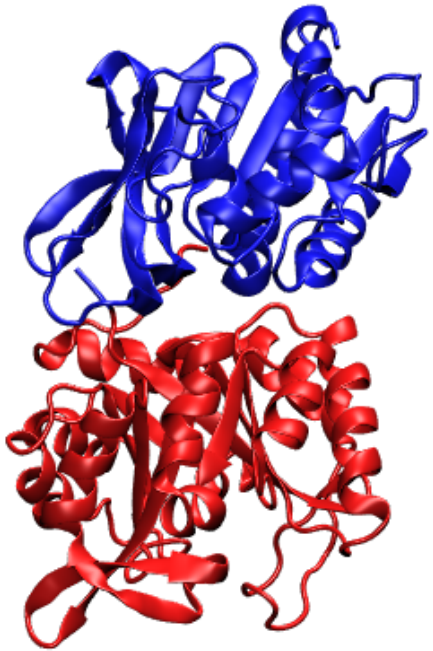
- **Molecular electronic structure graph for Peptide 1 aft protein.**
  - H. N. Djidjev, G. Hahn, S. M. Mniszewski, C. F. A. Negre, A. M. N. Niklasson, V. B. Sardeshmukh, Graph partitioning Methods for Fast Parallel Quantum Molecular Dynamics, SIAM Workshop on Combinatorial Scientific Computing 2016 (CSC16).
- **Amino acid sequence**
  - PHE-ASP-ALA-ASP-LEU-TYR-PHE
    - (PHEnylalanine, ASPartic acid, ALAnine, LEUcine, TYRosine)
- **LATTE QMD after SCF convergence**
- **Communities correspond to amino acids**

Peptide 1 aft graph (N = 300, E = 1794)
7 communities, modularity = 0.765666



Protein Structure Communities in Bio-Systems

- **IGPS is an enzyme in bacteria (454 residues)**
- **Applying community detection using *qbsolv* resulted in communities corresponding to IGPS's 2 molecules**
- **The modularity matrix is calculated from a correlation matrix based on a molecular dynamics (MD) simulation**



IGPS Protein Structure

Correlation matrix (for CNA)

$$\mathbf{r}_{ij}^{MI} = g(\mathbf{I}[\mathbf{x}_i, \mathbf{x}_j])$$

Girvan-Newman

Modularity Matrix

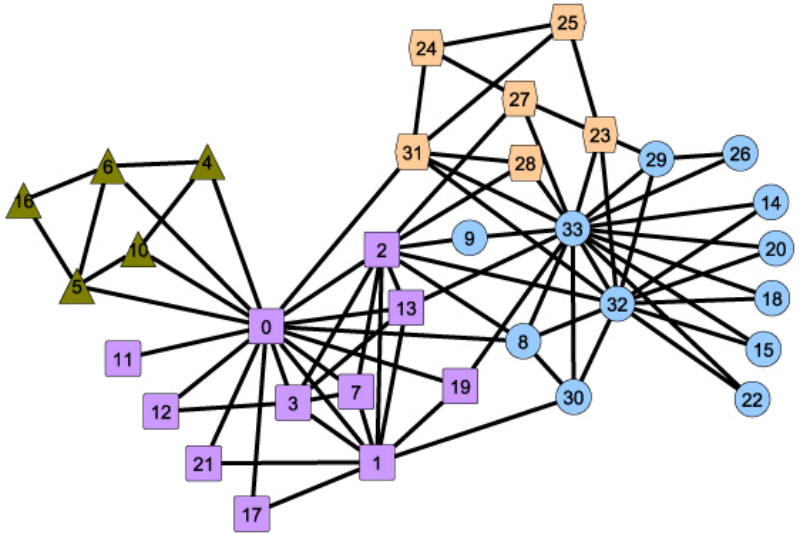
Protein Structure Communities in Bio-Systems

- **IGPS is an enzyme in bacteria consisting of 2 molecules (454 residues)**
- **Applying quantum community detection using *qbsolv* for up to k = 4 communities**
- **The 4 resulting communities each share common sub-structure**
  - Each molecule is composed of 2 domains
  - Each domain has a specific function
- **Results are comparable with classical methods**
  - **C. F. A. Negre, H. Hendrickson, R. Pal, I. Rivalta, J. Ho, V. S. Batista, Eigenvector Centrality Distribution for Characterization of Protein Allosteric Pathways, arXiv preprint arXiv:1706.02327 [q-bio.BM].**

IGPS Protein Structure

k-Clustering Community Detection with Thresholding

- Using community detection for natural clusters
- Compare for up to 4 clusters
- Threshold modularity matrix weights
- Results in reduction in qubit chains and couplers



Karate club graph (N = 34) using *qbsolv*

Threshold	# weights	# clusters	modularity
0	561	4	0.4197896
0.02	544	4	0.4197896
0.05	411	4	0.4197896
0.06	334	4	0.4197896
0.07	300	4	0.4151052
0.08	244	3	0.3990795
0.10	227	3	0.3990795
0.15	169	2	0.3717948
0.25	110	2	0.3717948

K-Means Clustering of Numerical Data

- **Given *X* data points partition into *k* disjoint subsets**
  - Zero mean data
  - Clusters defined by centroids
- **Even 2-means clustering is NP-hard**

- **Typically minimize the <u>*within cluster scatter*</u>** – $|| x - mean(x_i)||^2$
- **Alternatively maximize the <u>*between cluster scatter*</u>** - $|| mean(x_i) - mean(x_j)||^2$
- **Reduce to an Ising Model formulation**
  - Minimize $-\sum Q_{ij} s_i s_j$, *where* $Q = X^T X$ *which is a Gram Matrix*
  - Gram matrix elements are dot products of the *X* data pairs

- Bauckhage C., Brito E., Cvejoski K., Ojeda C., Sifa R., Wrobel S. (2018) Ising Models for Binary Clustering via Adiabatic Quantum Computing. In: Pelillo M., Hancock E. (eds) Energy Minimization Methods in Computer Vision and Pattern Recognition. EMMCVPR 2017. Lecture Notes in Computer Science, vol 10746.

2-Means clustering of Iris Data

- **Iris Data – UCI Machine Learning Repository**
- Fisher, R.A. "The use of multiple measurements in taxonomic problems" Annual Eugenics, 7, Part II, 179-188 (1936); also in "Contributions to Mathematical Statistics" (John Wiley, NY, 1950).
- **Best known database to be found in the pattern recognition literature**
- **150 instances with 4 features**
  - sepal length, sepal width, petal length, petal width
  - 3 classes: Iris Setosa, Iris Versicolour, Iris Virginica
- **Processing required:**
  - Zero mean centered data
  - Calculate Gram matrix - dot product of pairs of data
- **Results using Quantum Annealing**
  - Partitioned into Iris Setosa and Iris Versicolour + Iris Virginica

Summary

- Quantum annealing for modularity-based community detection results in equal or comparable results to existing classical approaches
- Demonstrated a hybrid classical-quantum approach for existing benchmark graphs, example graphs, and electronic structure graphs
- Quantum-based k-means clustering is possible as demonstrated for 2-means clustering
- Future plans include exploring quantum computing approaches to other clustering methods

Publications

- **S. M. Mniszewski, H. Ushijima-Mwesigwa, C. F. A. Negre, 2017, Graph Partitioning using the D-Wave for Electronic Structure Problems, SIAM Annual Meeting, Minisymposium MS43: Identifying Computational Methods for early Benefit from Quantum Computing,.**
- **P. Goddard, S. M. Mniszewski, F. Neukart, S. Pakin, S. Reinhardt, 2017, How Will Early Quantum Computing Benefit Computational Methods? *SIAM News*, Vol. 50, No. 10.**
- **H. Ushijima-Mwesigwa, C. F. A. Negre. S. M. Mniszewski, 2017, Graph Partitioning using Quantum Annealing on the D-Wave System, arxiv preprint arxiv.org/abs/1705.03082.**
- **H. Ushijima-Mwesigwa, C. F. A. Negre, S. M. Mniszewski, 2017, Graph Partitioning using Quantum Annealing on the D-Wave System, *Proceedings of the 2nd International Workshop on Post Moore's Era Supercomputing (PMES),* 22-29, SC 2017.**
- **S. M. Mniszewski, C. F. A. Negre, Ushijima-Mwesigwa, 2018, Graph Clustering Approaches using Near-term Quantum Computing, Argonne Quantum Computing Workshop.**
- **H. Ushijima-Mwesigwa, C. F. A. Negre, S. M. Mniszewski, I. Safro, 2018, Multilevel Quantum Annealing for Graph Partitioning, Argonne Quantum Computing Workshop.**
- **C. F. A. Negre, H. Ushijima-Mwesigwa, S. M. Mniszewski, 2018, Community Detection using Quantum Annealing on the D-Wave System, (in preparation).**

The End

Thank You!