Parallel Algorithm and Analysis for Understanding Evolving Community Structures in Temporal Graphs

Naw Safrin Sattar[1], Shaikh Arifuzzaman[1], Khaled Ibrahim[2], Aydin Buluc[2]

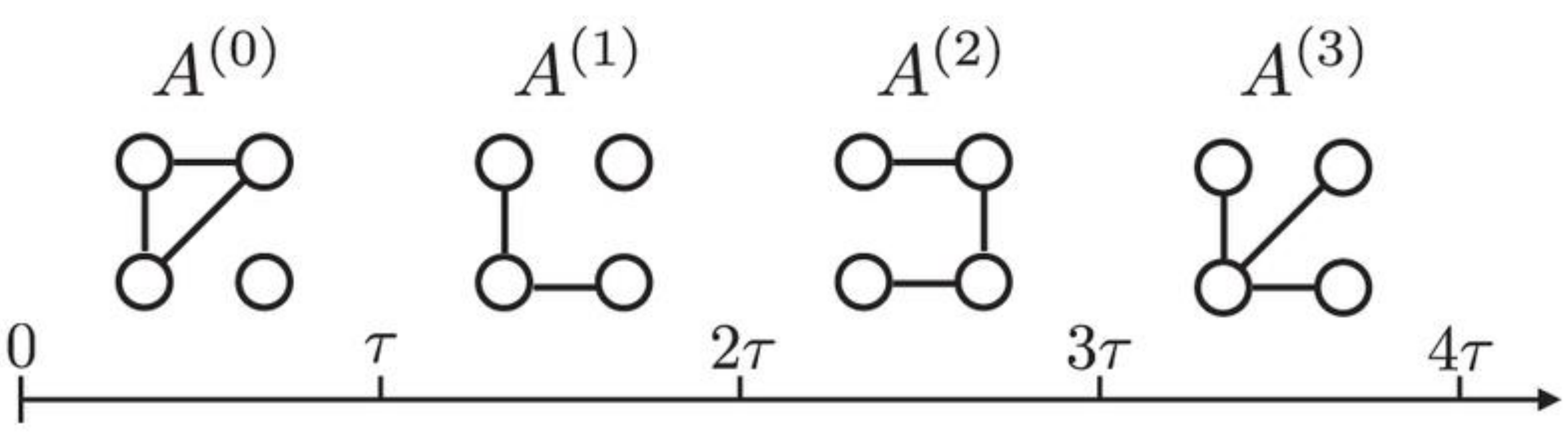
{nsattar, smarifuz}@uno.edu, {kzibrahim, abuluc}@lbl.gov

[1]Big Data and Scalable Computing Research Group, Dept. of Computer Science, University of New Orleans

[2]Performance and Algorithms Research Group, Computational Research Division, Lawrence Berkeley National Laboratory

Introduction

Temporal Network



- Different real-world complex networks: social, biological, bibliographic, communication, computer systems, etc.

Evolving Temporal Community

- A temporal community is defined as a partition of the set of {node, time} pairs in the network
- Nodes moving from one community to another, makes some communities grow, and others shrink
- Social Network: User's preference for music, politics, food habit, etc., changes
- Bibliographic Network: Authors change topic of work (Data Science, Machine Learning)

Methods

Parallel Community Detection

- Evolutionary Clustering Method
- Graph metric for optimization: Permanence [3]

Analysis for Community Evolution

Graph Metrics	Function
Intra Community Edges	E_{intra}
Internal Density	$\frac{2E_{intra}}{n_c(n_c-1)}$
Average Degree	$\frac{2E_{intra}}{n_c}$
Fraction over median degree (FOMD)	$\frac{\lvert\{u: u \in C, \lvert\{(u,v): v \in C\}\rvert > d_m\}\rvert}{n_c}$
Inter Community Edges	E_{inter}
Expansion	$\frac{E_{inter}}{n_c}$
Cut Ratio	$\frac{E_{inter}}{n_c(n-n_c)}$
Conductance	$\frac{E_{inter}}{2E_{intra}+E_{inter}}$
Normalized Cut	$\frac{E_{inter}}{2E_{intra}+E_{inter}} + \frac{E_{inter}}{2(E-E_{intra})+E_{inter}}$
Separability	$\frac{E_{intra}}{E_{inter}}$
Clustering Coefficient	$CC(v) = \frac{n_i}{\binom{D(v)}{2}}$
Permanence	$Perm(v) = \left[\frac{I(v)}{D(v)} \times \frac{1}{E_{max}(v)}\right] - [1 - c_{in}(v)]$

Real-world Networks [1,2,3]

Dataset	Snapshots	Nodes	Edges
Cumulative Co-authorship	17	708,497	1,166,376
Non-Cum Co-authorship	17	708,497	1,166,376
College Msg	7	1,899	59,835
Primary School	6	242	77,602

LFR Synthetic Networks

Dataset	Snapshots	Nodes	Mixing coefficient
Syn-1	20	3500	0.2
Syn-2	30	1000	0.2
Syn-3	10	1000	0.2

Score of the Graph Metrics for Community Quality

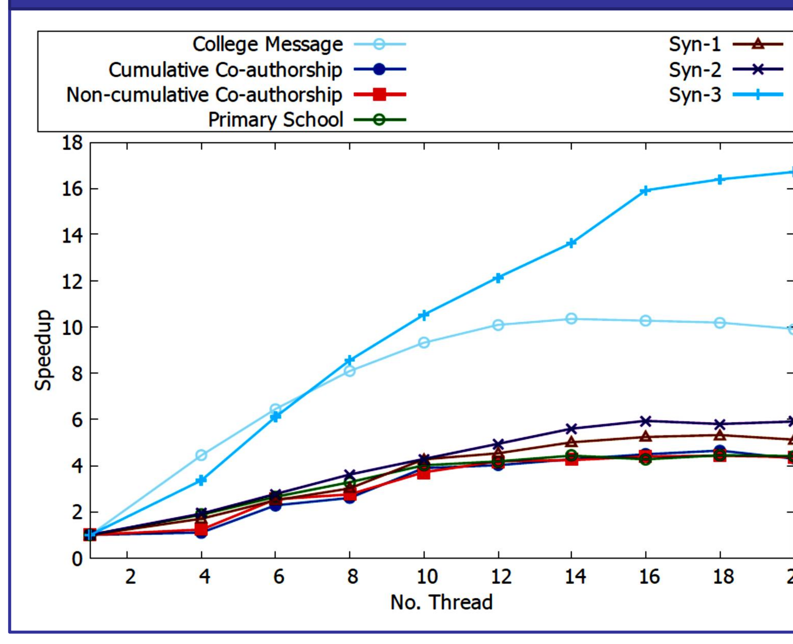
- Value Increased/(Decreased) over snapshots: Positive Score +1
- Value Decreased/ (Increased) over Snapshots: Negative Score -1
- Conductance, Normalized Cut, Separability, Clustering Coefficient and Permanence depends on both internal and external connection, so scoring value is doubled [+2 / -2]

Score	Maximization (1)				Minimization (1)			Minimization (2)			Maximization (2)				
Network	Intra Community Edges	Internal Density	Average Degree	FOMD	Inter Community Edges	Expansion	Cut Ratio	Conductance	Normalized Cut	Separability	Clustering Coefficient	Permanence	Total +ve, -ve Score	Total Score without CR and CC	Community Quality
Primary School	0	1	1	1	-1	-1	-1	-2	-2	2	2	0	+7,-7	-1	Weak
College Msg	0	0	0	1	0	-1	-1	0	0	2	0	0	+3,-2	2	Strong
Cum CoA	1	-1	0	-1	-1	-1	1	0	0	0	0	0	+2,-4	-3	Weak
Syn-1	0	0	0	1	-1	-1	-1	-2	-2	2	0	-2	+3,-9	-5	Weak
Syn-2 /Syn-3	0	0	0	1	-1	-1	-1	-2	-2	2	-2	-2	+3,-11	-5	Weak

Cut Ratio (CR) and Clustering Coefficient (CC): Minimal Change in Value => Same Community Quality

Results

Shared-Memory Parallel Algorithm



Multi-threading Environment

- Python Joblib
- Multiprocessing modules

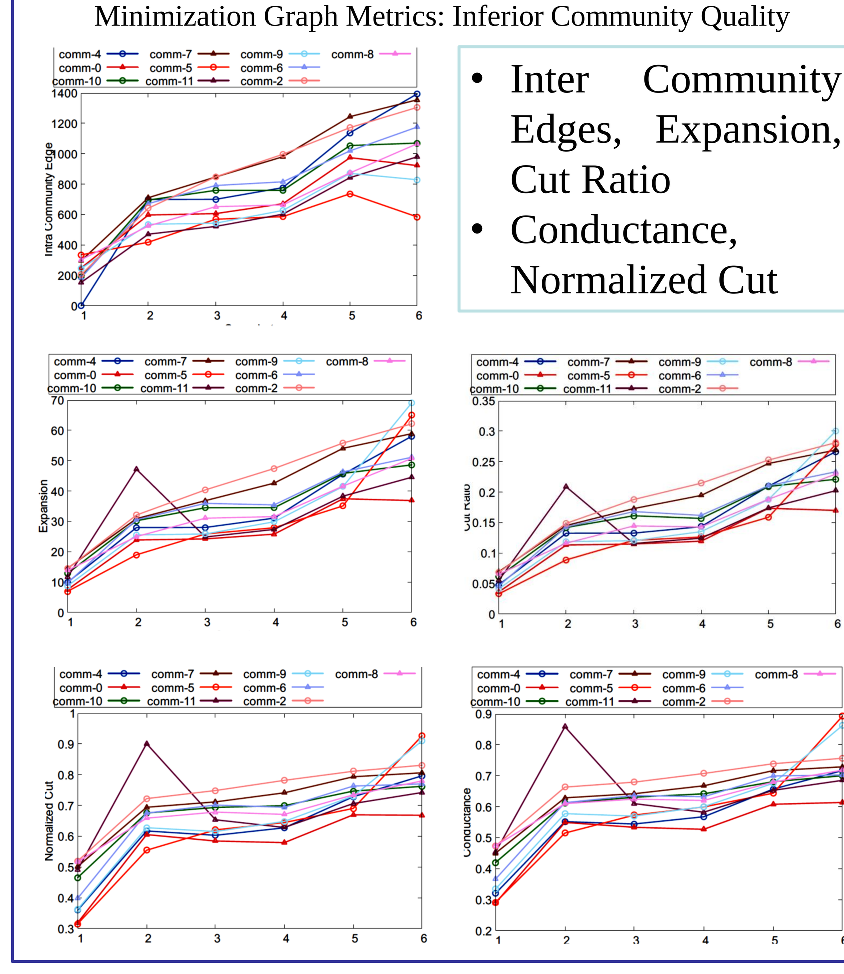
Computing Resources

Louisiana Optical Network Infrastructure (LONI) QB2 cluster: 1.5 Petaflop peak performance @2.8 GHz with over 10,000 Intel Xeon processing cores having 504 compute nodes, 20 cores per node.

Evolution of Community Quality w.r.t. Graph Metrics

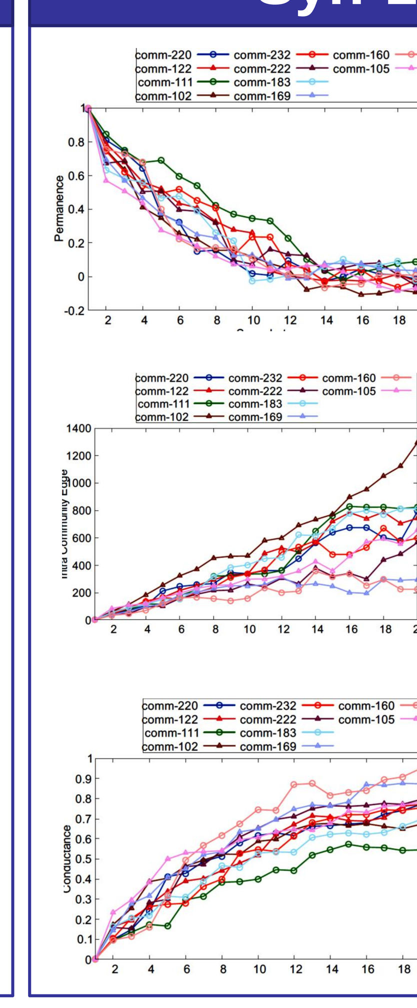
Primary School Network

Minimization Graph Metrics: Inferior Community Quality



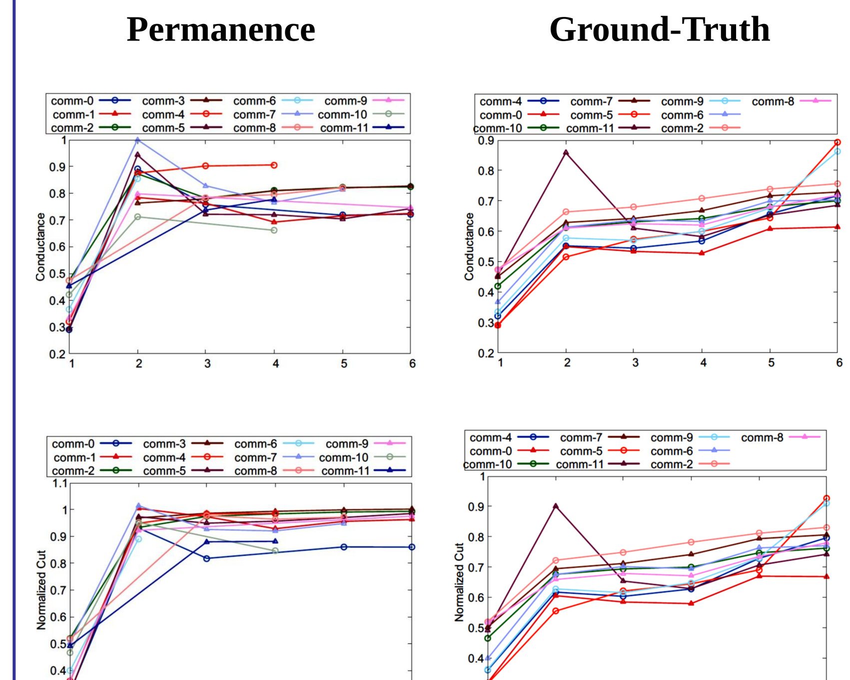
- Inter Community Edges, Expansion, Cut Ratio
- Conductance, Normalized Cut

Syn-1 Network

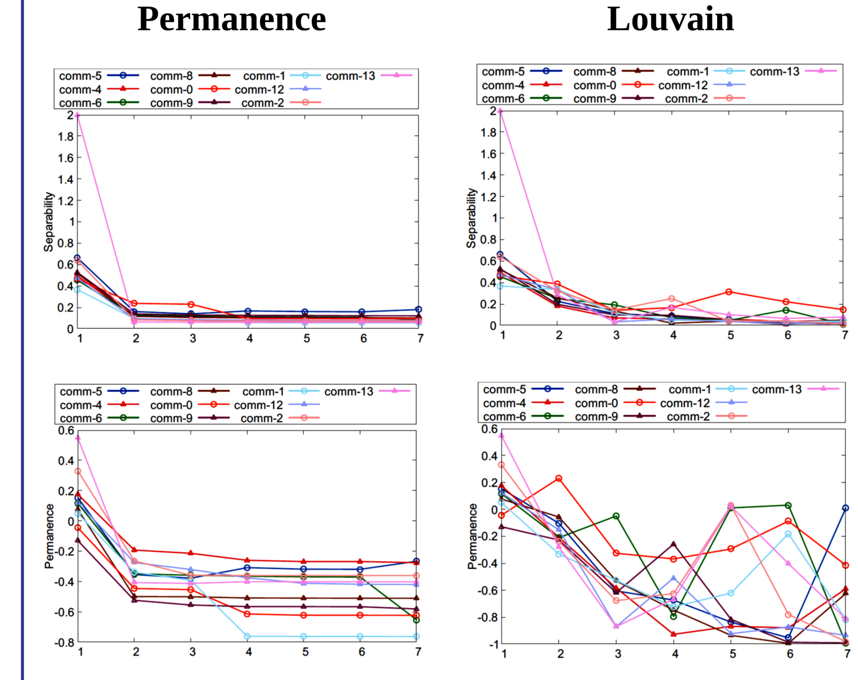


- **Maximization Metric:** Permanence
- **Minimization Metric:** Inter Community Edges, Expansion, Cut Ratio, Conductance, Normalized Cut

Comparison: Ground-Truth



Comparison: Louvain



Conclusions

- ~4-18x speed-up using thread model for shared memory parallelization
- Identification of a subset of graph metrics to understand the evolving community structure quality for different temporal networks
- Permanence-based communities showing similar or better community structure for both Ground-Truth and Louvain-derived [4] community structure
- Spike for the values of graph metrics in Louvain-derived communities indicating frequent change of the community members over communities

References

1. http://snap.stanford.edu/data/index.html#temporal
2. http://www.sociopatterns.org/
3. P. Agarwal, R. Verma, A. Agarwal, and T. Chakraborty, "Dyperm: Maximizing permanence for dynamic community detection," in Pacific-Asia Conference on Knowledge Discovery and Data Mining. Springer, 2018, pp. 437–449
4. V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks, "Journal of statistical mechanics: theory and experiment, vol. 2008, no. 10,p. P10008, 2008.