Toward Efficient Community Detection on Time-Varying Attributed Graphs

Nelson Aloysio Reis de Almeida Passos[1,2][0000−0003−1869−2976], Emanuele Carlini[2][0000−0003−3643−5404], and Salvatore Trani[2][0000−0001−6541−9409]

[1] University of Pisa, Dept. of Computer Science, 56127 Pisa PI, Italy
[2] National Research Council, 56124 Pisa PI, Italy

nelson.reis@phd.unipi.it
{emanuele.carlini,salvatore.trani}@isti.cnr.it

Abstract. Community detection on temporal graphs must ideally satisfy two competing demands: recovering ground-truth structure faithfully, and doing so efficiently at scale. This research charts a path toward both. We summarize contributions spanning the problem's three interconnected challenges: software and datasets for temporal graph representation; generative synthetic benchmarks for controlled model evaluation; and two complementary detection strands: a GPU-accelerated algorithmic route that brings spectral clustering and Leiden optimization to the CuPy and NVIDIA RAPIDS ecosystems — the latter yielding ∼2.7× to ∼978× speedup over the CPU reference depending on graph density and snapshot count, turning previously intractable cases into routine ones — and a neural route that folds spatio-temporal signals into a scalable, end-to-end differentiable modularity objective. We conclude by outlining the problems and next steps defining the remainder of this doctoral research.

Keywords: temporal graphs, community detection, neural networks

1 Introduction

Systems of interacting objects — networks — are commonly represented as graphs, with nodes (vertices) connected by edges (arcs). From molecules and food webs to transportation grids and the World Wide Web, this abstraction underpins the interdisciplinary study of complex systems known as Network Science [13]. A recurring observation across domains is that nodes organize into communities: groups characterized by distinct topological and/or attribute-based properties. Recovering these mesoscale structures — the task of *community detection* — is among the field's most studied problems [4,5], addressed over the previous decades by algorithmic, statistical-inferential, and, more recently, neural approaches [12].

Most real-world systems, however, are *dynamic*: social ties form and dissolve, routes are rerouted, and biological processes unfold over time [8]. They are frequently *attributed* as well, carrying node- and edge-level information that provides essential context to their modeling and analysis, such as knowledge graphs. Static representations discard both, risking oversimplified or misleading descriptions

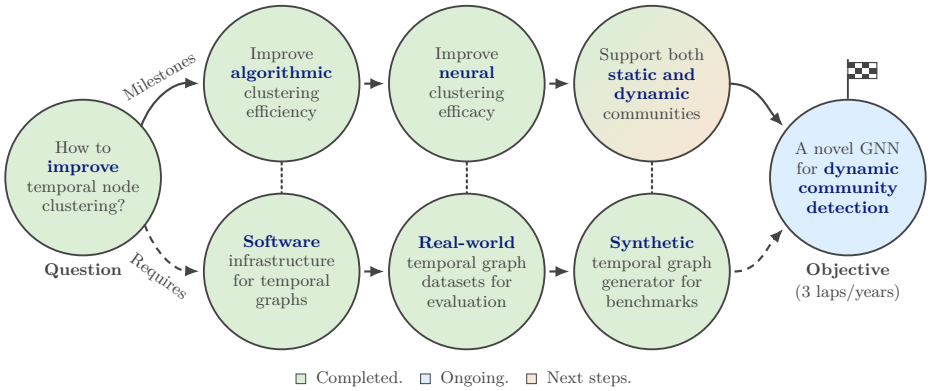

Fig. 1. Schematic of the research trajectory, with milestones, dependencies, and current status. The central question is tackled by three interconnected challenges: software for dynamic graph representations, principled model evaluation, and community detection in itself, addressed in both attributed and non-attributed temporal graph domains.

of the mechanisms at play. This research concerns mainly community detection in *time-varying attributed graphs* — relational systems whose connections are not fixed over time and whose elements may carry diverse (high-dimensional, real-valued) feature sets. The problem is notoriously hard: communities can grow, shrink, merge and split, or vanish and resurge [3], and a model is often asked not merely to detect them at each instant but to track coherent assignments across time, jointly reconciling structural (spatial), temporal (dynamic), and attribute (metadata) information. Compounding this, there is 'no free lunch' in what regards community detection [20]: no algorithm is universally optimal, and ground truths in real networks need not align with any single notion of community, yielding diverse and often conflicting methods across scientific disciplines [4,5].

Graph neural networks (GNNs) seem an appealing response, having reached the state of the art across graph-related tasks by jointly exploiting topology and attributes, with extensions to the temporal setting [9]. This motivates the central question of this research: *how to evaluate and improve algorithmic and neural community detection on temporal graphs*. We here define a 'neural' community as a set of nodes clustered in a real-valued vector space optimizing a graph-theoretic objective (e.g., with a GNN); while other descriptive and inferential methods, such as modularity and stochastic blockmodeling, are also employed throughout this research as baselines. Yet the answer is far from settled: neural partitions may themselves be read as posterior distributions of a block model, and their advantage over principled statistical approaches — particularly under the dual demands of *efficacy* and computational *efficiency* — remains unclear at best [21].

The question required tackling three interconnected challenges (Figure 1): new software developments for dynamic (temporal) graph representations, generative synthetic benchmarks for model evaluation under controlled experimental settings, and neural architectures for temporal community detection — the latter

currently an ongoing effort. Each yielded self-contained contributions, including NetworkX-Temporal [16], a library extending the well-known NetworkX [7] library to build, manipulate, and analyze graphs; TADC-SBM [17], a time-varying, attributed, degree-corrected stochastic block model generator for graphs with known ground truths; and LMoN [19], an end-to-end differentiable GNN optimizing a temporal variant of modularity [1]. Across these strands, efficient computation was a cross-cutting concern — from GPU-accelerated extensions of spectral clustering and greedy optimization algorithms, to neural methods that consider rich, high-dimensional node and edge metadata, applicable to real-world data at scale [18]. Together, they represent a step toward answering the central question, and more broadly, in the direction of building new knowledge and novel methods for temporal graph clustering that bridge network science and machine learning literature. This paper recounts that trajectory: the problem and its motivations, the contributions made to date, and open problems that remain.

2 Background & Related Work

Community detection seeks to find groups of nodes with shared characteristics or functional similarity. The prevailing assumption is that of assortative (homophilic) structures [10], i.e., nodes more densely connected among themselves than with the rest of the network are more likely to share properties, roles, or functions — dominating domains such as social network analysis, partly due to the way graphs are constructed from, e.g., friendship/follower interactions, while others may exhibit hierarchical (nested), disassortative (bipartite), core-periphery, or overlapping (mixed-membership) groups that call for alternative definitions [5].

A rich, interdisciplinary compendium of techniques has accrued around the problem, including divisive edge-betweenness, statistical physics models, inferential approaches, information-theoretic methods such as Infomap, label and belief propagation, non-negative matrix factorization, random-walk sampling, and ensemble combinations [4]. Among these, modularity optimization is arguably among the most widely adopted [21] — owing it to its ease of implementation and overall empirical performance, resulting in greater out-of-box software availability.

More recently, machine learning on graphs has recast detection as representation learning. Graph embeddings and neural models approximate a mapping $f(\mathcal{G}) \to \mathbb{R}^d$, with $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{X})$, placing nodes in a latent space where their relative distance encodes topological and/or attribute-based similarity [23,26]. The resulting representations are either clustered post hoc with conventional methods (e.g., k-means), or optimized end-to-end against a differentiable objective. By exploiting attributes jointly with structure, GNNs have reached the state of the art across diverse graph-related tasks — though for community detection, their advantage over principled statistical methods is not as clear-cut [9,20,21].

Real systems, however, are only seldom static. Extending these methods to the temporal domain typically relies on heuristic procedures, such as segmenting graphs into snapshots, with added inter-slice couplings among temporal nodes [8]. Compounding it, the task itself grows more complex: assuming communities

evolve — emerge, grow, merge, split, and dissolve [3] — their detection must distinguish fixed (static) from dynamic (temporal) membership assignments, ideally tracking coherent identities across the observed time span: a problem akin to that of the ship of Theseus, begging the question of whether a community that changes over time retains its identity, how so, and for how long [2]. Most methods, moreover, often rely on costly operations — e.g., repeated eigendecompositions and orthogonal matrix alignment [25] — that strain large or rapidly evolving networks under latency and tractability constraints, limiting their applicability to tasks such as mobility and trajectory analysis, financial fraud prevention, and wildlife conservation [6,22,27], which may require analyzing the spatio-temporal properties (and attributes, when available) of such events over large periods.

Graph representation learning marks the state-of-the-art paradigm from which this research departed. Time-aware GNNs augment message passing (n-hop feature aggregation over node neighborhoods) with attention, memory gating, or decay-based mechanisms to encode evolving topology and attribute signals [9]. Yet most methods still assign nodes to a fixed number of communities with static memberships, leaving genuinely *dynamic community detection* largely unexplored, while gains over algorithmic baselines remain contested under the joint demands of efficacy and efficiency — a gap this research sought to address, and whose contributions along several strands (Fig. 1) are briefly recounted next.

3 Contributions

Addressing the central question along the research axes of Figure 1 yielded several self-contained contributions, summarized here in order of maturity and impact.

The most consequential result is algorithmic. Spectral clustering [14] and multislice modularity [11], though widely adopted, are traditionally CPU-bound and ill-suited to graphs with many snapshots, where repeated eigendecompositions and greedy optimization quickly become prohibitive. We introduce GPU-accelerated formulations of both, built on the CuPy/RAPIDS ecosystem and operating over a supra-adjacency representation of the temporal graph — to our knowledge the first such implementations — exposed as a zero-code-change backend within NetworkX-Temporal [16]. Against its CPU reference under an equal-work budget (limited to two optimization passes), the multislice Leiden backend attained consistent speedups spanning roughly two to three orders of magnitude: from 2.7× on a dense low-snapshot dataset, to ∼978× on a sparse high-snapshot instance, reducing its wall-clock time from 1397 s to 1.4 s [18]. For the largest temporal instances, this represents a change in tractability: graphs that exceeded practical time limits on CPU become routine on GPUs. This advantage, however, is naturally confined to scale — on small static graphs, CPU dominates due to launch and transfer overheads to GPUs, a regime our work does not target.

To overcome the scarcity of temporal graphs with reliable ground truths [20], TADC-SBM [17] generates time-varying, attributed, degree-corrected graphs for controlled benchmarks. Experiments on real-world and synthetic graphs generated

with it show that neural methods do not consistently outperform established baselines [15], motivating both efficiency and accuracy as research goals.

On the neural strand, LMoN [19] extends the scalable DMoN [24] architecture to the temporal setting: a spectral relaxation of longitudinal modularity [1] yields a differentiable objective that recovers static modularity in the absence of temporal coupling, while a Hawkes-Laplacian normalization scheme helps distinguish persistent structure from transient events. The operator is encoder-agnostic and, to our knowledge, the first to optimize temporal modularity end-to-end, retaining its backbone's best-in-class scalability of $\mathcal{O}(d^2n + m)$, where m is the number of edges, n the number of nodes, and d the embedding dimension. It presently recovers fixed-membership communities, while genuinely dynamic detection at scale remains an open problem that we seek to address in the near future.

4 Final Remarks

This research charts a path toward community detection on time-varying attributed graphs in high-performance settings, contributing new software and datasets for temporal graph representation learning, synthetic benchmark generators for self-consistent model evaluation, and two complementary strands: a GPU-accelerated algorithmic route that turns previously intractable instances into routine ones, and a neural route that folds spatio-temporal and feature signals into a scalable, differentiable modularity objective. Several open problems worthy of note remain, among which we highlight the design of neural architectures for truly dynamic, evolving-label community detection, and the implementation of efficient eigensolvers for asymmetric, directed matrices on GPU hardware, which would enable spectral methods to scale to large temporal graphs under optimal asymptotic conditions without resorting to normalization or costly heuristics.

References

1. Brabant, V., Asgari, Y., Borgnat, P., Bonifati, A., Cazabet, R.: Longitudinal modularity, a modularity for link streams. EPJ Data Science **14**(1) (Feb 2025)
2. Dakiche, N., Benbouzid-Si Tayeb, F., Slimani, Y., Benatchba, K.: Tracking community evolution in social networks: A survey. Information Processing & Management **56**(3), 1084–1102 (5 2019)
3. Failla, A., Cazabet, R., Rossetti, G., Citraro, S.: Describing group evolution in temporal data using multi-faceted events. Machine Learning **113**(10) (2024)
4. Fortunato, S.: Community detection in graphs. Physics Reports **486**(3-5), 75–174 (2 2010)
5. Fortunato, S., Hric, D.: Community detection in networks: A user guide. Physics Reports **659**, 1–44 (Nov 2016)
6. Gouvêa, A.M.M.M., Vega-Oliveros, D.A., Cotacallapa, M., Ferreira, L.N., Macau, E.E.N., Quiles, M.G.: Dynamic community detection into analyzing of wildfires events. In: Computational Science and Its Applications extendash ICCSA 2020, pp. 1032–1047. Springer International Publishing (2020)

7. Hagberg, A.A., Schult, D.A., Swart, P.J.: Exploring network structure, dynamics, and function using NetworkX. In: Proceedings of the 7th Python in Science Conference. pp. 11–15. SciPy, SciPy (2008)
8. Holme, P., Saramäki, J. (eds.): Temporal Network Theory. Springer International Publishing (2023)
9. Longa, A., Lachi, V., Santin, G., Bianchini, M., Lepri, B., Lio, P., Scarselli, F., Passerini, A.: Graph neural networks for temporal graphs: State of the art, open challenges, and opportunities. Transactions on Machine Learning Research (2023)
10. McPherson, M., Smith-Lovin, L., Cook, J.M.: Birds of a feather: Homophily in social networks. Annual Review of Sociology **27**(1), 415–444 (8 2001)
11. Mucha, P.J., Richardson, T., Macon, K., Porter, M.A., Onnela, J.P.: Community structure in time-dependent, multiscale, and multiplex networks. Science **328**(5980), 876–878 (5 2010)
12. Newman, M.E.J., Girvan, M.: Finding and evaluating community structure in networks. Physical Review E **69**(2) (2 2004)
13. Newman, M.: Networks. Oxford University Press, 2 edn. (2018)
14. Ng, A., Jordan, M., Weiss, Y.: On spectral clustering: Analysis and an algorithm. In: Dietterich, T., Becker, S., Ghahramani, Z. (eds.) Advances in Neural Information Processing Systems. vol. 14. MIT Press (2001)
15. Passos, N.A.R.A., Carlini, E., Trani, S.: Deep community detection in attributed temporal graphs: Experimental evaluation of current approaches. In: Proceedings of the 3rd Graph Neural Networking Workshop 2024. pp. 1–6. GNNet '24, Association for Computing Machinery, New York, NY, USA (2024)
16. Passos, N.A.R.A., Carlini, E., Trani, S.: NetworkX-Temporal: Building, manipulating, and analyzing dynamic graph structures. SoftwareX **31**, 102277 (2025)
17. Passos, N.A.R.A., Carlini, E., Trani, S.: TADC-SBM: a time-varying, attributed, degree-corrected stochastic block model. In: 2025 IEEE Symposium on Computers and Communications (ISCC). pp. 1–6 (2025)
18. Passos, N.A.R.A., Carlini, E., Trani, S.: Accelerating dynamic graph clustering on GPU architectures with cuGraph (2026), accepted to: 6th workshop on Flexible Resource and Application Management on the Edge (FRAME @ Euro-Par '26)
19. Passos, N.A.R.A., Carlini, E., Trani, S.: Dynamic graph clustering with graph neural networks (2026), poster: Learning on Graphs (Pisa Meetup, June 2026)
20. Peel, L., Larremore, D.B., Clauset, A.: The ground truth about metadata and community detection in networks. Science Advances **3**(5) (May 2017)
21. Peixoto, T.P.: Descriptive vs. Inferential Community Detection in Networks: Pitfalls, Myths and Half-Truths. Elements in the Structure and Dynamics of Complex Networks, Cambridge University Press (2023)
22. Sarma, D., Alam, W., Saha, I., Alam, M.N., Alam, M.J., Hossain, S.: Bank fraud detection using community detection algorithm. In: ICIRCA. pp. 642–646 (2020)
23. Su, X., Xue, S., Liu, F., Wu, J., Yang, J., Zhou, C., Hu, W., Paris, C., Nepal, S., Jin, D., Sheng, Q.Z., Yu, P.S.: A comprehensive survey on community detection with deep learning. IEEE T. on N.Networks and Learning Systems pp. 1–21 (2022)
24. Tsitsulin, A., Palowitch, J., Perozzi, B., Müller, E.: Graph clustering with graph neural networks. J. Mach. Learn. Res. **24**(1) (mar 2024)
25. Xue, G., Zhong, M., Li, J., Chen, J., Zhai, C., Kong, R.: Dynamic network embedding survey. Neurocomputing **472**, 212–223 (Feb 2022)
26. Zhang, D., Yin, J., Zhu, X., Zhang, C.: Network representation learning: A survey. IEEE Transactions on Big Data **6**(1), 3–28 (2020)
27. Zhao, Y., Chen, B.Y., Gao, F., Zhu, X.: Dynamic community detection considering daily rhythms of human mobility. Travel B. and Society **31**, 209–222 (4 2023)