TEMPO

Management and Processing of Temporal Networks

H.F.R.I. Project No. 03480

D1.3: First Year Project Report

Computer Engineering & Informatics Department
University of Patras
Greece
30/04/2023

First Year Project Report of TEMPO

Kostas Tsichlas

April 30, 2023

Abstract

This document discusses the progress during the first year of the project. First, we report on the deliverables, milestones as well as on the risks and deviations from the initial plan. Then, we discuss the work performed within each WorkPackage (WP) and finally we report on dissemination activities.

1 Introduction

In the first year of the project, significant progress has been made towards the goals of the project. The project officially began on 1st April 2022. At that time, there were only two main researchers, as named in the proposal:

1. Konstantinos Tsichlas: Assistant Professor at the University of Patras (Scientific Coordinator).

2. Spyros Sioutas: Professor at the University of Patras.

The following two PhD students were recruited by mid of July 2022 (M3-M4 as planned) following an open-call procedure in M2 (May 2022):

1. Konstantinos Christopoulos: PhD student that focuses on community detection (WP4) and outlier detection (WP5).

2. Alexandros Spitalas: PhD student that focuses on the implementation of the historical graph management system (WP2) as well as on the implementation of the query engine for transactional and analytic queries (WP3).

Both students started officially working on the project on the 14th of July 2022 and their contract ends on the 31st of March 2025. In addition, two external collaborators provide us with valuable knowledge concerning such systems:

1. Anastasion Gounaris: Associate Professor in the Informatics department of Aristotle University of Thessaloniki. His area of expertise in on distributed systems among others.

2. Apostolos Papadopoulos: Associate Professor in the Informatics department of Aristotle University of Thessaloniki. His area of expertise in on graph databases among others.

Overall, by M4, the complete research team was formed, as planned. In addition, 4 deliverables were prepared in the 1st year of the project, as will be detailed later, while special attention was given to dissemination activities. Apart from the website and the publications, the logo image was prepared as shown in Figure 1.

In the remainder of this section, we give an overview of the progress in terms of deliverable preparation and milestone achievements. We also include a discussion regarding risks and deviations from the plan. In Section 2, we provide more details about the technical work in each WP. Finally, in Section 3, we provide an overview of the dissemination activities that took place in the first year of the project.

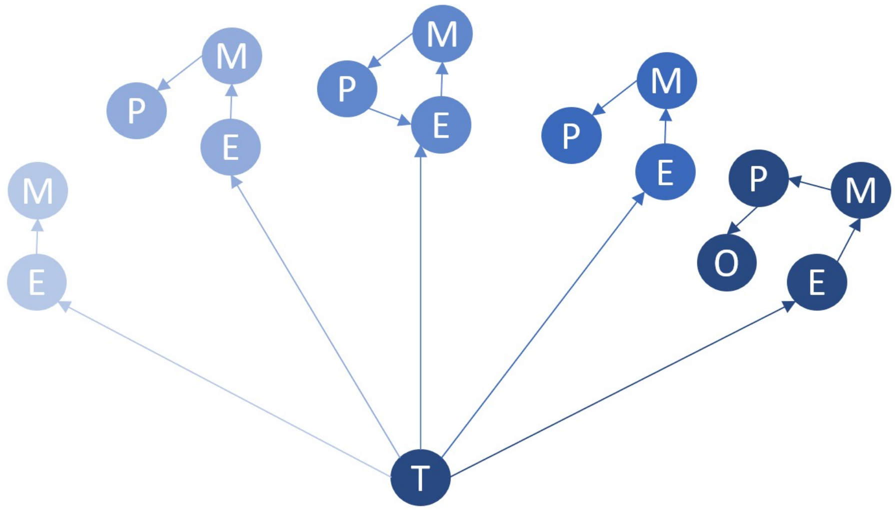

Figure 1: Logo of Project TEMPO.

1.1 Deliverables

In the first year of the project, four deliverables were planned and all of them were prepared accordingly.

D1.1 Project Website:[1] A screenshot of the project website can be seen in Figure 2. It contains a description of the project, the research team as well as external collaborators, and the dissemination material that includes public deliverables and publications to conferences and journals.

D1.2 Project Scientific/Technical Plan: This deliverable contains an extensive literature survey of temporal graph management and processing systems of various flavors and different approaches related mainly to WP2 and WP3. Based on this extensive survey, we designed the architecture of the historical graph management system based on a vertex-centric approach. In addition, a timeplan was laid out as to how to proceed with respect to the implementation of the system as well as the research on community and outlier detection.

D4.1 State of the Art in Temporal Community Detection: An extensive literature survey of community detection approaches in temporal graphs is provided.

D5.1 State of the Art in Outlier Detection on Time-Evolving Graphs: An extensive literature survey of outlier detection approaches is provided. We mainly focused on community-based outlier detection approaches for time-evolving graphs. Other approaches were discussed to a lesser extent since they were not the focus of this project.

1.2 Milestones

There was one milestone for the first year, namely MS1, which refers to requirements gathering and concrete scientific and technical plan preparation to be verified through deliverable D1.2. This milestone has been achieved by M12, which means that it was slightly delayed since we expected this to happen by M8. Based on these results, we have laid the foundations for achieving MS2 and MS3, although the former is expected to be delayed by 6 months (at M24 instead of M18 for MS2). In 1.3, a thorough justification of these delays is provided.

[1] https://www.ceid.upatras.gr/webpages/faculty/ktsichlas/Tempo/tempo_web.html. As soon as the website of the laboratory ML@CLOUD is setup, to which the scientific coordinator belongs, it will be moved there.

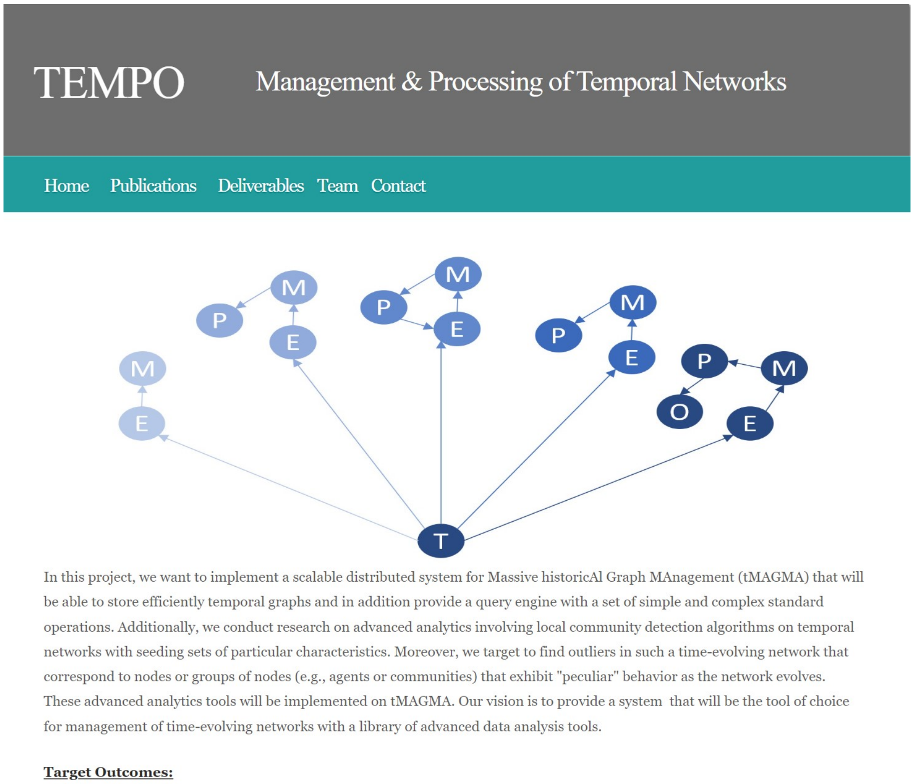

Figure 2: Screenshot from the Project Website.

1.3 Risks and Deviation from the Plan

In the first year of the project all WPs have begun. Our main deviation from the plan concerns the completion of the technical/scientific plan (D1.2), which instead of finishing on M6 it finished on M12. The main reason for this deviation was our thorough literature survey of the rather many historical graph management systems that were published during the period of 2020 to 2023, since the proposal was written during the period of 2019 to 2020 (the deadline for the submission of the call was in June 2020).

Our initial approach was to implement the system from scratch. The merit of this approach was that we could make any choice related to the architecture of the system and its characteristics without been constrained by an existing system. However, we could not hope to have a completed system by the end of the project and this was reflected in the proposal since many crucial aspects of such a system (e.g., support of transactions having particular properties - or a subset - defined by ACID) were not discussed. We were aiming at implementing a prototype system that had basic functionality. However, the extensive literature survey convinced us that other approaches could work better. We analyze all these approaches in D1.2.

In particular, we decided to use an existing graph database for static graphs and extend it for time-evolving graphs without sacrificing the vertex-centric approach for storing the history of the graph. This extension does not only concern the maintenance of the history of the time-evolving graph but in addition it supports transactions at any point in its evolution. The main reasons for which we adopted this approach are the following:

1. The functionality of these graph databases allows our system to have an extended set of properties (e.g., OLTP transactions and OLAP queries) that move it closer to be a ready to use system.

2. The static graph database contains extensive additional tools that can be used either directly or with minor changes for time-evolving networks. For example, the Apache Tinkerpop graph stack can be used with some changes to accommodate time related queries.

3. The visibility and rate of adoption of such a solution based on an existing static graph databases is much larger when compared to a solution that builds from scratch.

Apparently, the choice of the static graph database was of paramount importance. As argued in D1.2, we chose Janusgraph for many reasons analyzed in this document, one of the crucial being the fact that it is open-source and no license is needed for full deployment in a distributed environment. All in all, our desire to develop a system that will evolve in a ready-to-use system for managing and processing of historical graphs led us to to adopt this approach.

2 Summary of Work per WP

2.1 WP1: Project Management - Result Dissemination

The four main activities that took place were:

1. Formation of the complete research team (as already mentioned previously).

2. Preparation of the project website (D1.1). A screenshot is shown in Figure 2.

3. Preparation of D1.2, D4.1 and D5.1. We have to highlight D1.2, in which the scientific and technical plan details are provided and several clarifications and scientific changes are made with respect to the initial proposal.

4. Three presentations in international conferences and two journals, as will be detailed in Section 3.

Regarding D1.2, some further summary details are required. Using the approved text of the proposal as a basis, a further state-of-the-art survey was conducted related to historical graph management and processing systems. Based on this effort and after extensive brainstorming within the research team and with external collaborators we decided the basic characteristics of the proposed system - always within the bounds set by the approved proposal. In summary, the changes/modifications in D1.2 compared to the initial proposal are made with a view to making the proposed system more clear and functional in the following sense: (i) support for transactional OLTP queries combined with support for OLAP queries; (ii) ready to use libraries (with or without minor modifications); (iii) high visibility in the community that already uses this system.

2.2 WP2: MAGMA System Implementation

Evaluation of NoSQL implementation alternatives (task T2.1) was already handled partially in [9]. During the writing of this report, extended experimentation is being conducted to submit these results to a journal. The research team has also been extensively involved in looking at other recently introduced systems for managing and processing of historical graphs as shown in deliverable D1.2 that was concluded by M12. Although storage and retrieval module implementation has not started yet (task T2.2), this delay was necessary to update and clarify all aspects of the system that had not been faced or they were partially faced during the writing of the proposal. All in all, although there is a minor delay in WP2, this is not considered to be a problem since the preparation during the first 12 months will speed up the implementation in the second year.

2.3 WP3: Complex Data Analytic Queries

The starting month of WP3 is M12, and thus no progress was expected to be made. However, because of the choices we made with respect to the architecture of MAGMA, the query engine is provided by Janusgraph and we need to change it in order to incorporate time as a first-class citizen and not as another property. As a result, it is expected that during the first months of the second year of the project, the research team will familiarize itself with Janusgraph and all its tools and highlight the points that need to be changed in order to handle historical graphs. In this respect, task T3.4 concerning the incorporation of the query engine in MAGMA is not the final stage of WP3, but it is continuously achieved through the changes we apply to accommodate time in Janusgraph.

2.4 WP4: Local Community Detection

Deliverable D4.1 that concerns the state of the art in temporal community detection has concluded task T4.1. Progress has been made in task T4.2 and T4.3 to a lesser extent (note that T4.3 was planned to start on M24) concerning local community detection with hints in snapshot networks and in temporal networks respectively by looking at algorithms in streaming graphs. Streaming graphs can be considered as a way to continuously generate snapshot graphs while they are also a restricted type of temporal graphs since updates take place only at the last time instance.

2.5 WP5: Outlier Detection

Deliverable D5.1 that concerns the state of the art in outlier detection in temporal graphs has concluded task T5.1. Task T5.2 that concerns design of algorithms for outlier detection in time-evolving graphs has started in M9. We have considered the definition of an outlier related to the changes in the community structure and we intend to move towards this direction effectively using the results of WP4 to design methods for this task. Since outlier detection is a problem amenable to contemporary machine learning methods, we have also considered adopting such an approach by using the definition of outlierness based on the community structure.

3 Publications and Other Dissemination Material

There are five publications already out of the work in the project:

1. **State-of-the-art in Community Detection in Temporal Networks** [5] that contains the work within deliverable D4.1.
2. **MAGMA: Proposing a Massive Historical Graph Management System** [10] that contains part of the survey within deliverable D1.2 as well as a preliminary version of the system architecture of MAGMA.
3. **Dynamic Local Community Detection with Anchors** [1] that contains a streaming algorithm for local community detection by using anchors instead of seeds.
4. **Local Community Detection: A survey** [2] that contains a survey on local community detection algorithms based partially on deliverable D4.1. Note that no such stand-alone survey existed for the problem of local community detection.
5. **Local Community Detection in Graph Streams with Anchors** [6] that contains extended research on local community detection in a streaming environment based on [1].

Note that apart from previous work of the research team that we directly build upon, such as [7] and [8], there are other three publications that where produced in the period between the proposal preparation and the project kick-off. First, both [3] and [4] present methods related to local community detection with anchors for static and dynamic graphs. Their content is straightforwardly related to WP4. In addition, [9] presents results related to how the storage model is affected for HiNode when MongoDB is used and has a straightforward contribution in WP2 and in WP3.

References

[1] **Dynamic Local Community Detection with Anchors.** G. Baltsou, K. Christopoulos and K. Tsichlas. *In Proc. of the 11th Intern. Conference on Complex Networks and their Applications*, 2022.

[2] **Local Community Detection: A survey.** G. Baltsou, K. Christopoulos and K. Tsichlas. *IEEE ACCESS*, vol. 10, pp. 110701-110726, 2022.

[3] **Local community detection with hints.** G. Baltsou, K. Tsichlas and A. Vakali. *Applied Intelligence*,doi:10.1007/s10489-021-02946-7, 2022.

[4] **Dynamic Community Detection with Anchors (Extended Abstract)** G. Baltsou and K. Tsichlas. *In Proc. of the 10th Intern. Conference on Complex Networks and their Applications*, 2021.

[5] **State-of-the-art in Community Detection in Temporal Networks.** K. Christopoulos and K. Tsichlas. *In Proc. of the 18th International Conference on Artificial Intelligence, Applications, and Innovations (AIAI) - Mining Humanistic Data Workshop (MHDW)*, 2022.

[6] **Local Community Detection in Graph Streams with Anchors.** K. Christopoulos, G. Baltsou and K. Tsichlas. *Information*, 14(6): 332, 2023.

[7] **HiNode: An Asymptotically Space-Optimal Storage Model for Historical Queries on Graphs.** A. Kosmatopoulos, K. Tsichlas, A. Gounaris, S. Sioutas and E. Pitoura. *Distributed and Parallel Databases*, 35(3-4):249–285, 2017.

[8] **Hinode: Implementing a Vertex-Centric Modelling Approach to Maintaining Historical Graph Data.** A. Kosmatopoulos, A. Gounaris and K. Tsichlas. *Computing*, pp. 1-24, 2019.

[9] **Investigation of Database Models for Evolving Graphs.** A. Spitalas, A. Gounaris, A. Kosmatopoulos, K. Tsichlas. *In proc. of the 28th International Symposium on Temporal Representation and Reasoning(TIME)*, pp. 6:1-6:13, 2021.

[10] **MAGMA: Proposing a Massive Historical Graph Management System.** A. Spitalas and K. Tsichlas. *In Proc. of the 7th International Symposium on Algorithmic Aspects of Cloud Computing ALGOCLOUD*, pp. 42-57, 2022.