# Temporal Subgraph Isomorphism

Ursula Redmond*, Pádraig Cunningham†
School of Computer Science and Informatics
University College Dublin, Ireland
*Email: ursula.redmond@ucdconnect.ie
†Email: padraig.cunningham@ucd.ie

***Abstract*—Temporal information is increasingly available with network data sets. This information can expose underlying processes in the data via sequences of link activations. Examples range from the propagation of ideas through a scientific collaboration network, to the spread of disease via contacts between infected and susceptible individuals. We focus on the flow of funds through an online financial transaction network, in which given patterns might signify suspicious behaviour. The search for these patterns may be formulated as a temporally constrained subgraph isomorphism problem. We compare two algorithms which use temporal data at different stages during the search, and empirically demonstrate one to be significantly more efficient.**

## I. Introduction

Increasingly, network data sets include information describing the times at which interactions occur. For example, in a communication network, the timing of contacts can illustrate the flow of a piece of information. In a network of micro-blogs, the time at which links occur can reflect the diffusion of sentiment. Infection spreads among individuals through their physical contacts, and depends on features such as infection time and duration.

The focus of our current work is on the flow of money through an online peer-to-peer lending network, in which users lend to and borrow from each other directly. In this setting, a temporally increasing path of transactions may warrant further investigation. If there are defaults along the path, that could point towards money laundering. If the interest rates of loans on a path are increasing, there could be arbitrageurs [1].

Given a particular subgraph pattern of interest, our task is to find all of its embeddings in a given network graph, such that the embeddings are time-respecting. We approach this as a subgraph isomorphism problem, known to be NP-complete [2], and use temporal information to prune the search space. We present two methods, which use temporal and topological information at different stages. We compare the running time of the implementations, to empirically demonstrate the benefits associated with active use of temporal information.

## II. Related Work

The subgraph isomorphism problem determines whether a given graph contains a subgraph which has the same topological structure as another given graph. Some algorithms that are widely used to tackle this problem use techniques such as backtracking [3] and recursion [4].

Holme *et al.* [5] present a comprehensive review of temporal networks. Kempe *et al.* [6] define a *time-respecting* path as a sequence of contacts with non-decreasing times. The *relay time* of an edge captures the time taken for a newly infected node to spread the infection further [7]. The spread of information through a network can be modeled by a cascade. The importance of *time-constrained* cascades is emphasized for understanding contagion [8].

## III. Methods

Some necessary concepts follow. The subgraphs we seek are time-respecting, made of paths that are activated in non-decreasing sequences [9]. We employ a variant of the subgraph isomorphism problem for our task.

*Definition 1:* A graph $G2$ is isomorphic to a subgraph of a graph $G1$ if and only if there is a one-to-one correspondence between the node sets of this subgraph and of $G2$ that preserves adjacency [3].

*Definition 2:* A directed temporal graph $G$ consists of a set $V$ of nodes and a set $E$ of ordered pairs of nodes representing interactions. An interaction $e_i$ in $E$ is represented by a three-tuple $e_i = (u_i, v_i, t_i)$, in which $u_i$ is the source node, $v_i$ is the target node and $t_i$ is the initiation time of the interaction.

*Definition 3:* Let $e_i$ and $e_j$ be edges in a temporal graph. The edges are time-respecting if $v_i = u_j$ and $0 \leq t_j - t_i \leq d$, for some threshold $d$.

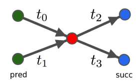
*Definition 4:* A time-respecting subgraph $S = (V', E')$ of a temporal graph $G = (V, E)$ is composed of a node set $V' \subseteq V$, and a set of interactions $E' \subseteq E$ such that every edge pair ($e_i$, $e_j$) in which $v_i = u_j$ is time-respecting.

We extend an implementation of the VF2 algorithm [4] from the NetworkX Python library [10]. $G1$ represents the large graph in which a subgraph $G2$ is sought. The matching process is described by a State Space Representation. Given an intermediate state $s$, the mapping is extended by first computing candidate node pairs (one node each from $G1$ and $G2$), then testing their topological feasibility as matching nodes.

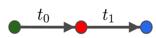
The Topology Before Time (To-Ti) solution first generates candidate subgraphs that match the topology of the query graph. The returned candidates are then evaluated, to test whether the time-respecting property holds. This evaluation applies the tests illustrated in Fig.1 to each node in all of the candidate subgraphs. Only those which are time-respecting are retained.

This work is supported by Science Foundation Ireland under Grant No. 08/SRC/I1407.

The Time and Topology Together (Ti&To) solution tests for semantic feasibility at each step of the topological test. Given an embedding of the subgraph $G2$ in the graph $G1$, we require that inclusion of a node in its subgraph will maintain the time-respecting property of that subgraph. We define $G2$ to be time-respecting by construction. Thus, only the semantic feasibility of $G1$ is checked, using the tests shown in Fig.1.




(a) The greatest time between interactions must not exceed $d$, i.e. $t_3 - t_0 \leq d$




(b) All incoming edges must precede all outgoing edges, i.e. $t_0 \leq t_1$

Fig. 1: Semantic feasibility tests, for all nodes in the mapping.

## IV. Results

We analyzed data from the peer-to-peer lending website Prosper.com [11]. We constructed a directed, temporal network made up of loans between Prosper members. Each edge is time-stamped with the origination date of the loan. The period we study covers November and December 2006. There are 72 215 edges and 8 690 nodes. We set the waiting time between interactions to $d$ = 5 days [1]. The query subgraphs include paths, cliques, fan-in-fan-out motifs and combinations of these.

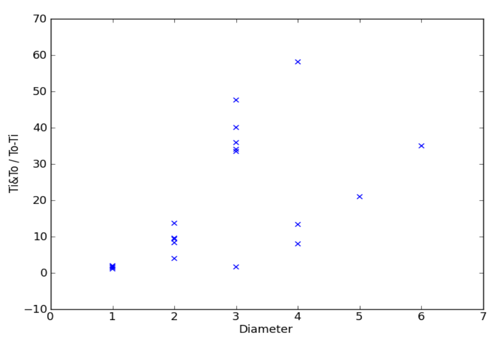



Fig. 2: The ratio of *Ti&To* / *To-Ti* compared with the diameter of each subgraph (jitter applied for clarity).

The *Ti&To* algorithm was up to sixty time faster than *To-Ti*. We exclude the subgraphs for which *To-Ti* did not complete within 48 hours. Since *Ti&To* performed faster than its opponent on all query subgraphs, we sought a predictor for the types of subgraph that are found fastest. We found subgraph diameter (the length of the longest shortest path) to be the best predictor, as shown in Fig.2. With a longer path, there are more immediate opportunities for pruning based on a comparison of the times stamped on adjacent edges. Fig.3 illustrates the speedup that occurs when the number of spurious candidates increases. It can be seen that, to some extent, as the number of candidates increases, *To-Ti* takes increasingly longer than *Ti&To*.

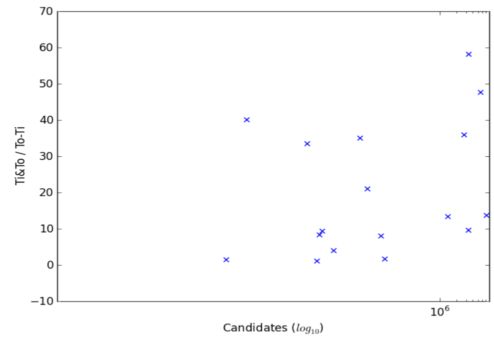



Fig. 3: The number of candidates per query subgraph compared with the ratio of *Ti&To* / *To-Ti*.

## V. Conclusion

To our knowledge, this is the first application of temporal information in the solution of the subgraph isomorphism problem. The best solution that we describe solves the problem up to sixty times faster than an approach that uses topological and temporal information in separate stages. Our approach also scales to larger target graphs than those used in previous discussions of the problem [4].

There are a number of research directions that could follow this work. Since our contributions relate to graph theoretical and application-centered problems, there is broad scope for development. Our next objective with this work is to evaluate the methods on other network datasets where the identification of time-respecting structure is of interest.

## References


[1] U. Redmond and P. Cunningham, “A temporal network analysis reveals the unprofitability of arbitrage in The Prosper Marketplace,” *Expert Systems with Applications*, vol. 40, no. 9, pp. 3715–3721, 7 2013.

[2] Garey, Michael R. and Johnson, David S., *Computers and Intractability: A Guide to the Theory of NP-Completeness*. New York, NY, USA: W. H. Freeman & Co., 1979.

[3] Ullmann, J., “An Algorithm for Subgraph Isomorphism,” *Journal of the ACM (JACM)*, vol. 23, no. 1, pp. 31–42, 1976.

[4] Cordella,L. P., Foggia,P., Sansone, C., and Vento, M., “An Improved Algorithm for Matching Large Graphs,” *$3^{rd}$ IAPR-TC15 Workshop Graph-Based Representations in Pattern Recognition*, pp. 149–159, 2001.

[5] P. Holme and J. Saramäki, “Temporal networks,” *CoRR*, vol. abs/1108.1780, 2011. [Online]. Available: http://arxiv.org/abs/1108.1780

[6] D. Kempe, J. Kleinberg, and A. Kumar, “Connectivity and inference problems for temporal networks,” *Journal of Computer and System Sciences*, vol. 76, no. 036113, 2002.

[7] M. Kivelä, R. K. Pan, K. Kaski, J. Kertész, J. Saramäki, and M. Karsai, “Multiscale analysis of spreading in a large communication network,” *CoRR*, 2011. [Online]. Available: http://arxiv.org/abs/1112.4312

[8] R. A. Baños, J. Borge-Holthoefer, and Y. Moreno, “The role of hidden influentials in the diffusion of online information cascades,” *CoRR*, 2013. [Online]. Available: http://arxiv.org/abs/1303.4629

[9] R. K. Pan and J. Saramäki, “Path lengths, correlations, and centrality in temporal networks,” *Physical Review E*, vol. 84, no. 1, 2011.

[10] NetworkX Developers, “NetworkX,” networkx.lanl.gov, 2010.

[11] Prosper Marketplace, Inc., “Personal Loans and Online Investing – Peer to Peer Lending – Prosper,” http://www.prosper.com/, 2010.