ETMM: Egocentric temporal motifs miner

Antonio Longa[1,2], Giulia Cencetti[2], Andrea Passerini[1], and Bruno Lepri[2]

[1] Trento University, Trento, Italy
[2] Fondazione Bruno Kessler, Trento, Italy

1 Introduction

A decade ago, Milo *et al.* [1] had introduced the notion of *network motifs* for static graphs, that are recurrent and statistically significant sub-graphs of a network. In the beginning, *motifs* were mainly used in biology and later researchers started to use this concept in other disciplines such as applications to temporal graphs. Temporal graphs are indispensable in modelling social interactions, being a standard graph not able to capture the related temporal dynamics. The concept of *motifs* suggests that in order to characterize a graph not only the global and the local scale are important, but it focuses on the mesoscale. This dimension of exploration of a graph reveals fundamental when we deal with social interactions, where people tend to reproduce similar patterns of interactions both in time and with different contacts. There are mainly two methods in the state of the art to tackle the problem of identifying the significance of such patterns in temporal graphs. One focuses on static snapshots[2, 3] and the other one derives static *motifs* from the aggregated graph validating them only if they are time consistent (temporal links appearing in a short interval of time). An example of the latter technique is reported in Paranjape *et al.* [4](TMM). In this work the authors introduce a method for the extraction of temporal *motifs* that consists in extracting static *motifs* from the aggregated graph. Only after, they introduce the temporal dimension and identify the most significant ones. [5, 6] demonstrated that the concept of centrality and motifs in temporal networks is significantly different from those found in static networks. Both approaches assume that a sub-graph that is not statistically relevant in the static or aggregated graph will not be meaningful in the temporal one. We however think that, in order to assess if a motif is significant of not, it is convenient, in some applied settings, to consider not only the spatial component, but also the temporal one. For instance a *motif* that is statistically insignificant in the aggregated graph, may be vital in the temporal version, just due to its specific link appearance order in time. We hence look for motifs among all the spatio-temporal subnetworks, capturing both the specific shape of connections but also their temporal correlation. For instance, in networks of social interactions, the probability that two people interact at one time is higher if they already interacted in the past.

This represents the main novelty of the work, together with the fact that we are able to simplify the motif search by focusing on the egocentric perspective. This allows us to reduce computational complexity with respect to other methods. Moreover we drastically diminish the number of significant patterns, allowing us to interpret the results in the context of social interactions. In other words, the discovered motifs are easily associated with real scenarios, informing us about variability of contacts, strength of ties, conversation durations, to mention some, for each individual node. To succeed in our aims, we applied an innovative technique that combines data mining and network science to identify the backbone of the complex scenario represented by social interactions. With our approach, we ultimately disentangled the bricks that constitute the essence of human interactions to build our knowledge of them.

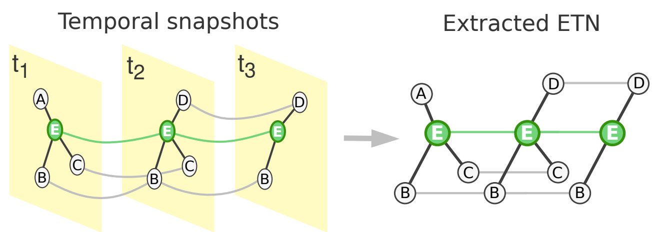


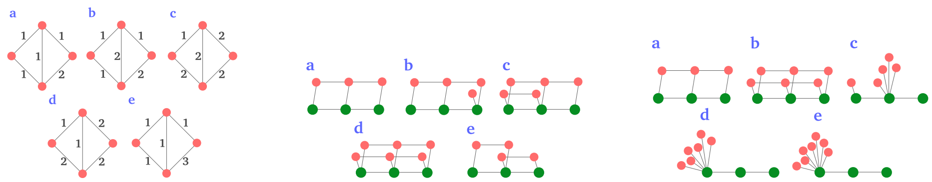
Fig. 1: On the left, three snapshots of a temporal graph focusing on the ego node E. On the right, the structure created by merging three temporal layers from the point of view of the egocentric node E

2 Results

In order to give an idea of the difference between egocentric and non-egocentric motifs and highlight the utility of the former in discovering patterns of social interaction, we report in the following the five most frequent motifs found by the ETMM and TMM. In this extended abstract we only focus on a High school temporal Network [7]. The high school data set has been collected by SocioPatterns in 2011 in Lycée Thiers, Marseilles, France. It reports the interactions among 118 students and 8 teachers in three different high school classes. Number of edges: 1709, number of nodes: 126.

Figures 2a show the first five motifs found by TMM on the *High School*, network. These motifs do shown some dynamics in the interaction, but it is difficult to interpret them in terms of social interaction patterns or to identify some clear features.

The five most frequents motifs discovered by our method on *High school 11* are reported in Figure 2b. Note that the egocentric focus allows to generate motifs which are quite interpretable in terms of social interaction of the person under investigation, like a continuous interaction with one (a) or two (d) other persons, somebody joining (b) or leaving (c) a conversation at a certain point, or a rotation of two partners (e). Figure 2c shows the five most frequent motifs found with $\Delta t = 300$. Note that this relatively large gap (5 minutes) drives ETMM to discover as significant motifs with many more interacting neighbors than those found in the previous cases. This is possible because the egocentric focus allows the miner to consider the entire neighborhood of a node in mining patterns, something computationally infeasible for alternative non-egocentric approaches.



(a) Most frequent temporal motifs discovered by TMM on *High School 11* for $\Delta t = 60$.

(b) Most frequent egocentric temporal motifs discovered by ETMM on *High School 11* for $\Delta t = 60$.

(c) Most frequent egocentric temporal motifs discovered by ETMM on *High School 11* for $\Delta t = 300$.

Fig. 2: Most frequent temporal motifs discovered by TMM(a), ETMM(b) with $\Delta t = 60$, and those discovered by ETMM(c) with $\Delta t = 300$

In this extended abstract, we present a novel strategy to extract statistically significant sub-graphs in temporal networks by concentrating on the egocentricity of a node. We argue that by aggregating the temporal graphs, temporal-dependent information such as the length over time of the interactions, their frequency, periodicity and others are lost. Accordingly, for each node in a three contiguous temporal graph, a structure is created, where relations within the same temporal gap maintain their structure and relations among different contiguous temporal gap exist if the node connected to the ego is the same. Figure 1 shows an example of the structure created by the aggregation of the three contiguous time steps. Starting from the extracted structures, we implemented a null model in order to obtain patterns and sub-graphs that are statistically meaningful. Later, we show the five most frequent temporal motifs discovered by the our egocentric temporal motifs miner (ETMM) and temporal motifs miner (TMM) developed by [4]

References

1. R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, "Network motifs: simple building blocks of complex networks," *Science*, vol. 298, no. 5594, pp. 824–827, 2002.
2. M. Araujo, S. Papadimitriou, S. Günnemann, C. Faloutsos, P. Basu, A. Swami, E. E. Papalexakis, and D. Koutra, "Com2: fast automatic discovery of temporal ('comet') communities," in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pp. 271–283, Springer, 2014.
3. D. M. Dunlavy, T. G. Kolda, and E. Acar, "Temporal link prediction using matrix and tensor factorizations," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 5, no. 2, pp. 1–27, 2011.
4. A. Paranjape, A. R. Benson, and J. Leskovec, "Motifs in temporal networks," in *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, pp. 601–610, 2017.
5. D. Braha and Y. Bar-Yam, "Time-dependent complex networks: Dynamic centrality, dynamic motifs, and cycles of social interactions," in *Adaptive Networks*, pp. 39–50, Springer, 2009.
6. D. Braha and Y. Bar-Yam, "From centrality to temporary fame: Dynamic centrality in complex networks," *Complexity*, vol. 12, no. 2, pp. 59–63, 2006.
7. J. Fournet and A. Barrat, "Contact patterns among high school students," *PLoS ONE*, vol. 9, p. e107878, 09 2014.