ETMM: Egocentric Temporal Motifs Miner

A. Longa, G. Cencetti, A. Passerini and B.Lepri

a.longa@fbk.eu https://antoniolonga.github.io/

Longa, Antonio, et al. "An efficient procedure for mining egocentric temporal motifs." Data Mining and Knowledge Discovery (2021): 1-24.

ECML PKDD 2022

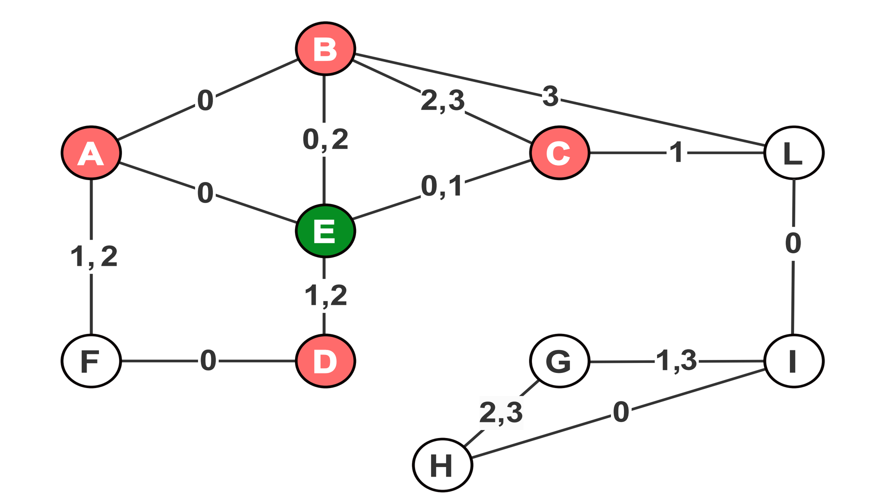
FONDAZIONE BRUNO KESSLER

Introduction

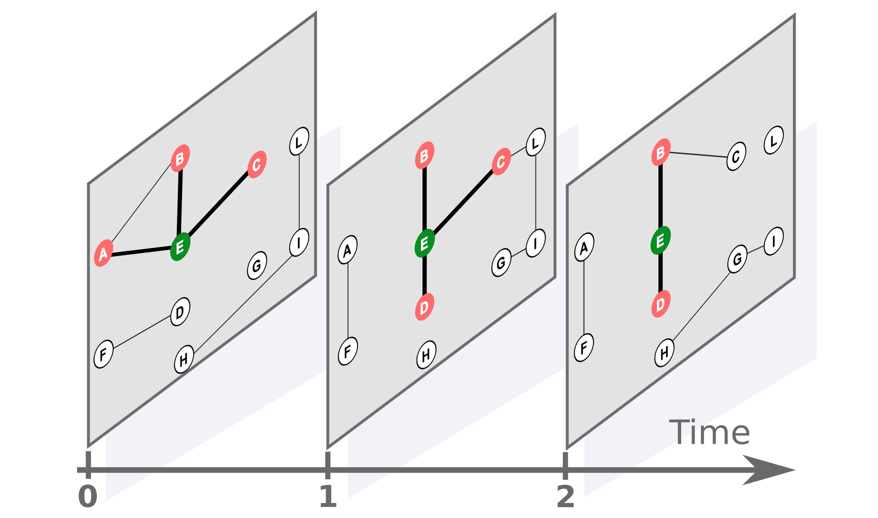
Temporal graphs are structures which model relational data between entities that change over time. Due to the complex structure of data, mining statistically significant temporal subgraphs, also known as temporal motifs, is a challenging task. In this work, we present an efficient technique for extracting temporal motifs. Our method is based on the novel notion of egocentric temporal neighborhoods, namely multi-layer structures centered on an ego node. The strength of this approach lies in the possibility of encoding these structures into a unique bit vector, thus bypassing the problem of graph isomorphism in searching for temporal motifs.

Extract Egocentic Temporal Neighborhood

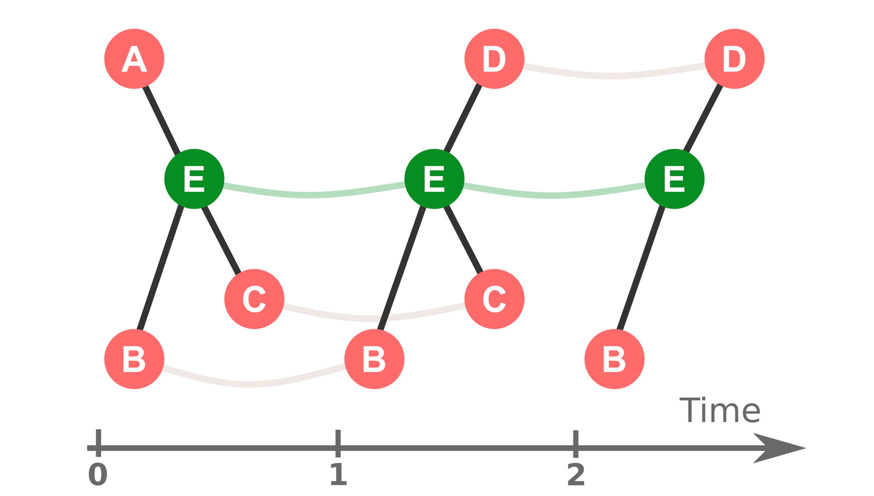
Temporal graph



Temporal graph snapshots



(ETN)



(ETNS)

Node encoding

A → 100 B → 111 C → 110 D → 011

Sorted node encoding

D A C B

011 - 100 - 110 - 111

ETNS

011100110111

From ETN to ETM and definitions

An ETN is considered Egocentric Temporal Motifs (ETM) if:

- It is over-represented with respect to a null model.
- It has a minimum deviation.
- It has a minimum frequency.

Definition 1: (ETM-based embedding) Given a temporal graph G and a list M of ETMs, we define EMBM(G) as the embedding of G in a vector of cardinality |M|, in which the i-th element of EMBM(G) represents the number of occurrences of M[i] in G.

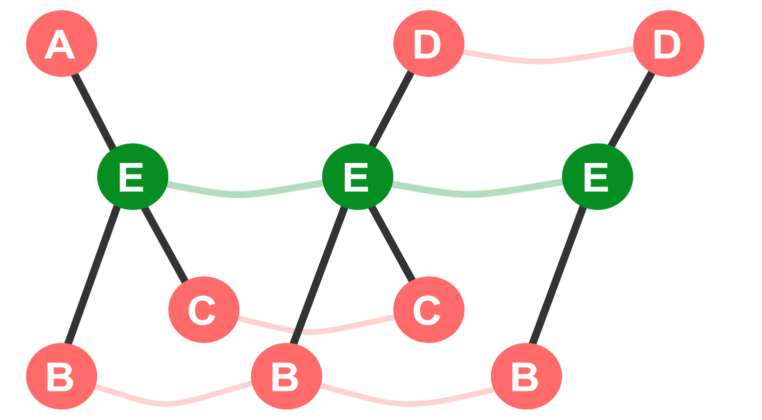
Given a list of ETM, the distance between two temporal graphs is then defined as the distance between their respective ETM-based embeddings.

Definition 2: (ETM-based distance) Given two temporal graphs G_1 , G_2 and a list M of ETMs, we define $dist_M(G_1 , G_2)$ as the cosine distance between the ETM-based embeddings of G_1 and G_2:

$$dist_M(\mathcal{G}_1, \mathcal{G}_2) = 1 - \frac{EMB_M(\mathcal{G}_1) \cdot EMB_M(\mathcal{G}_2)}{||EMB_M(\mathcal{G}_1)|| \; ||EMB_M(\mathcal{G}_2)||}$$

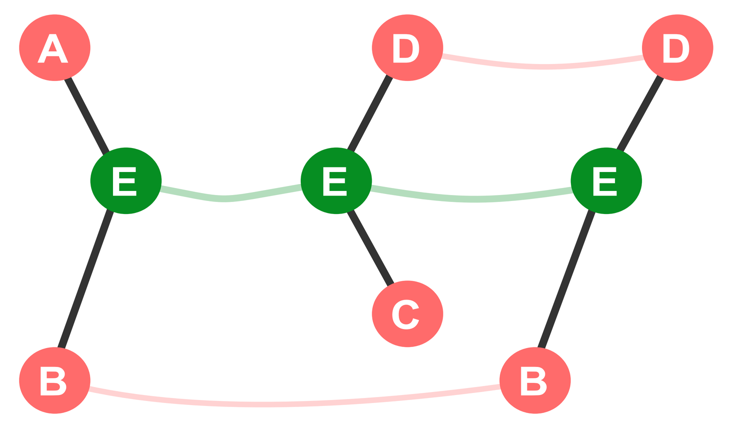
where · is the dot product and || · || is the Euclidean norm.

Isomorphism test in constant time



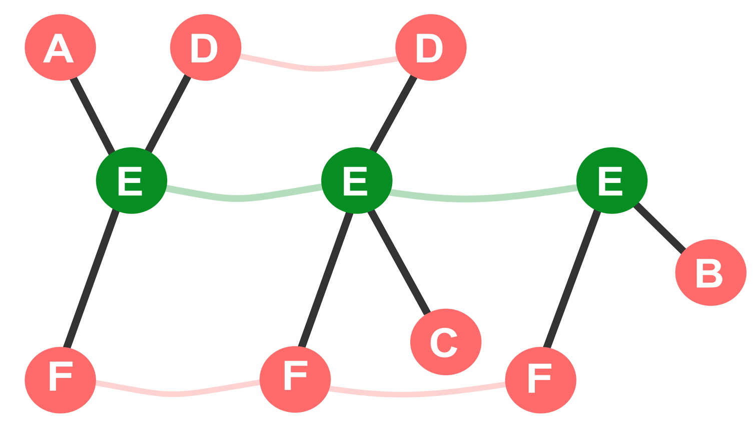
D A C B

ETNS = 011-100-110-111



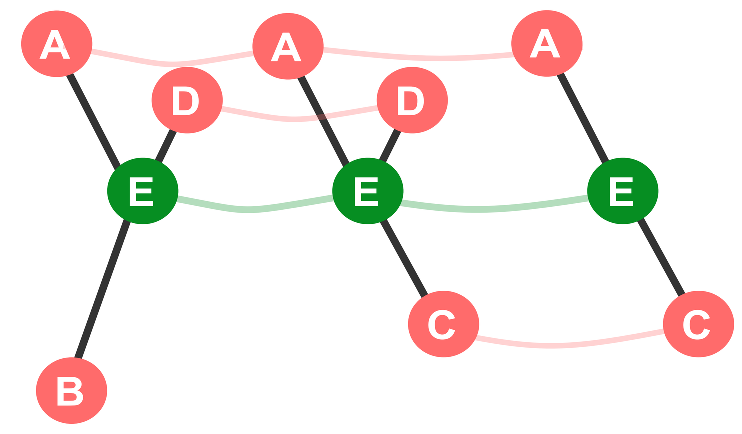
C D A B

ETNS = 010-011-100-101



B C A D F

ETNS = 001-010-100-110-111



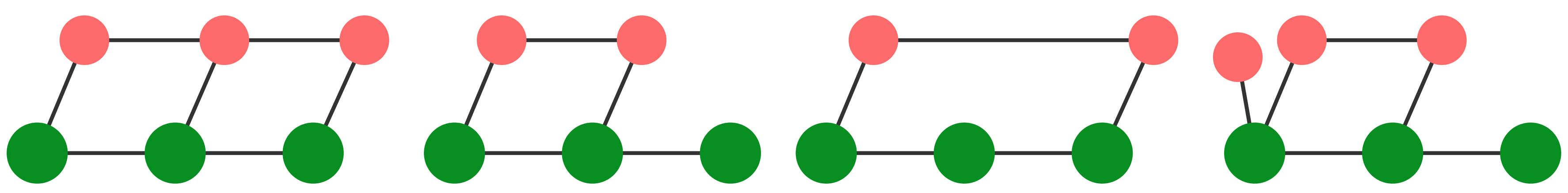
C B D A

ETNS = 011-100-110-111

Results

Egocentric temporal motifs

Workplace



Communication networks

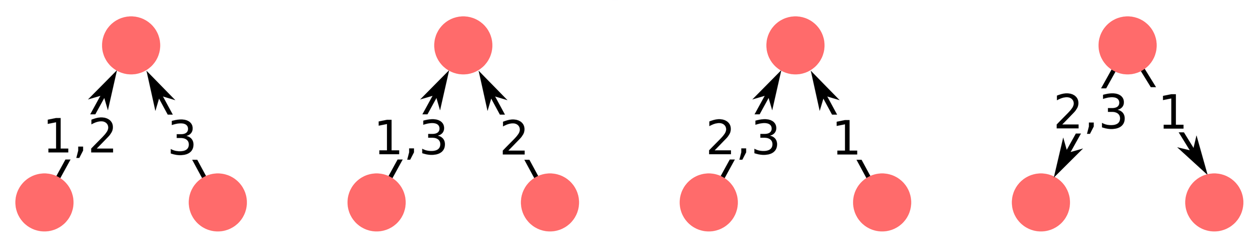
		Calls		SMS		Email	
		DTU C	Frien C	Dtu S	Frien S	Email	DNC
Calls	DTU C	0.00	0.37	0.28	0.26	0.61	0.66
	Friend C		0.00	0.34	0.33	0.59	0.58
SMS	Dtu S			0.00	0.06	0.65	0.64
	Friend S				0.00	0.65	0.64
Email	Email					0.00	0.38
	Email DNC						0.00

Face-to-face interactions

	VS13	LH10	HS11	HS12	HS13	PS	DTU
InVS13	0	0.07	0.29	0.22	0.29	0.67	0.47
LH10		0	0.29	0.22	0.30	0.66	0.45
HighSchool11			0	0.04	0.04	0.59	0.06
HighSchool12				0	0.02	0.61	0.13
HighSchool13					0	0.62	0.08
primary school						0	0.62
DTU blue							0

Temporal motifs

Workplace



Communication networks

		Calls		SMS		Email	
		DTU C	Frien C	Dtu S	Frien S	Email	DNC
Calls	DTU C	0.00	0.96	0.29	0.96	0.66	0.73
	Friend C		0.00	0.98	0.02	0.97	0.89
SMS	Dtu S			0.00	0.98	0.81	0.87
	Friend S				0.00	0.97	0.89
Email	Email					0.00	0.07
	Email DNC						0.00

Face-to-face interactions

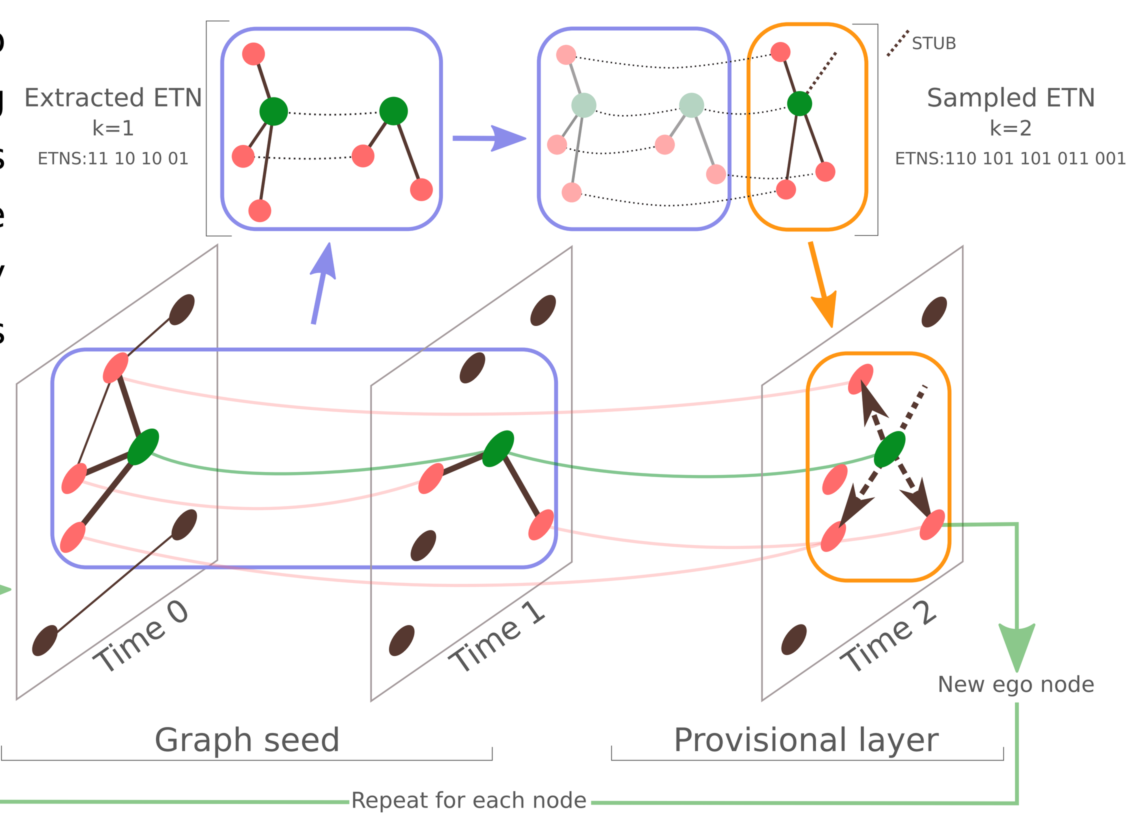
	VS13	LH10	HS11	HS12	HS13	PS	DTU
InVS13	0	0.018	0.02	0.053	0.016	0.04	0.744
LH10		0	0.012	0.014	0.001	0.005	0.707
HighSchool11			0	0.049	0.017	0.019	0.678
HighSchool12				0	0.012	0.1	0.696
HighSchool13					0	0.007	0.695
primary school						0	0.651
DTU blue							0

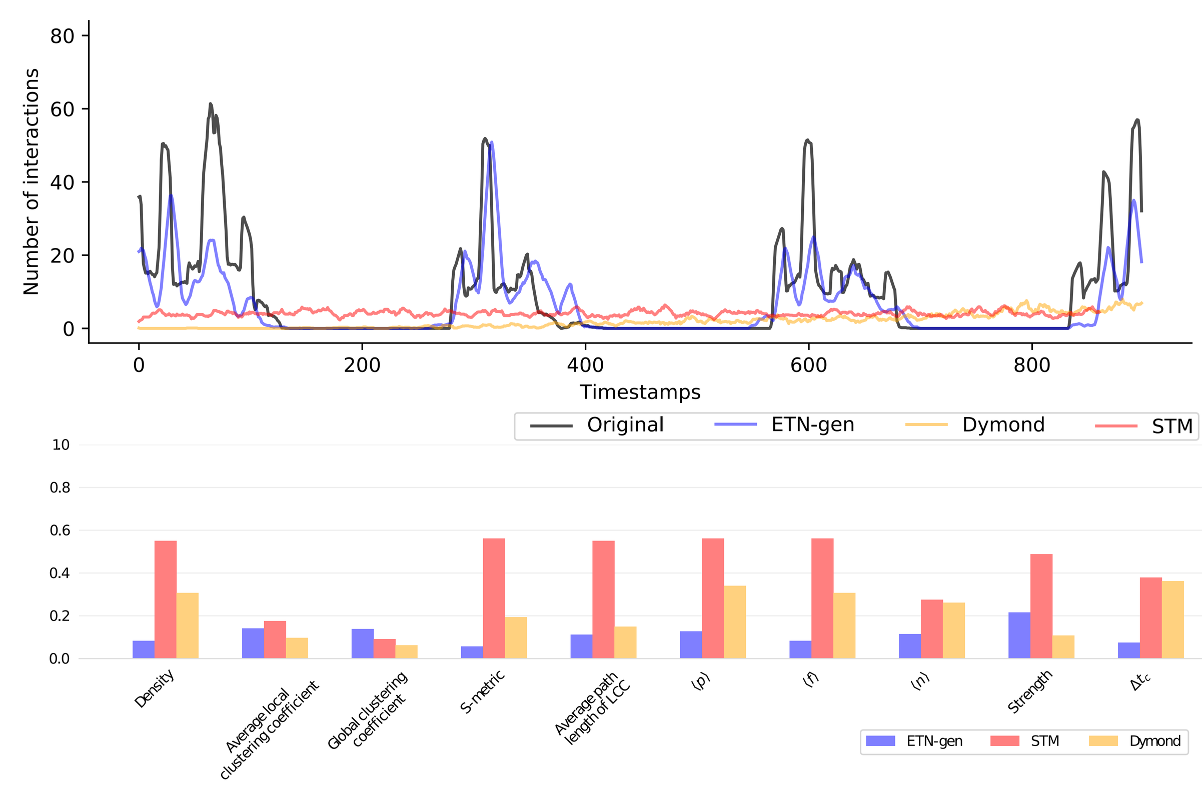
Follow up

Temporal network generation

Real-world temporal network data are prohibitively expensive to collect or unshareable due to privacy concerns. A promising solution is surrogate networks, synthetic graphs with the properties of real-world networks. Here, we propose a novel and simple method for generating surrogate temporal networks. By decomposing graphs into ETN, we can generate new networks using neighborhoods as building blocks.

Neighbourhood matching creates realistic surrogate temporal networks





PAPER

CODE