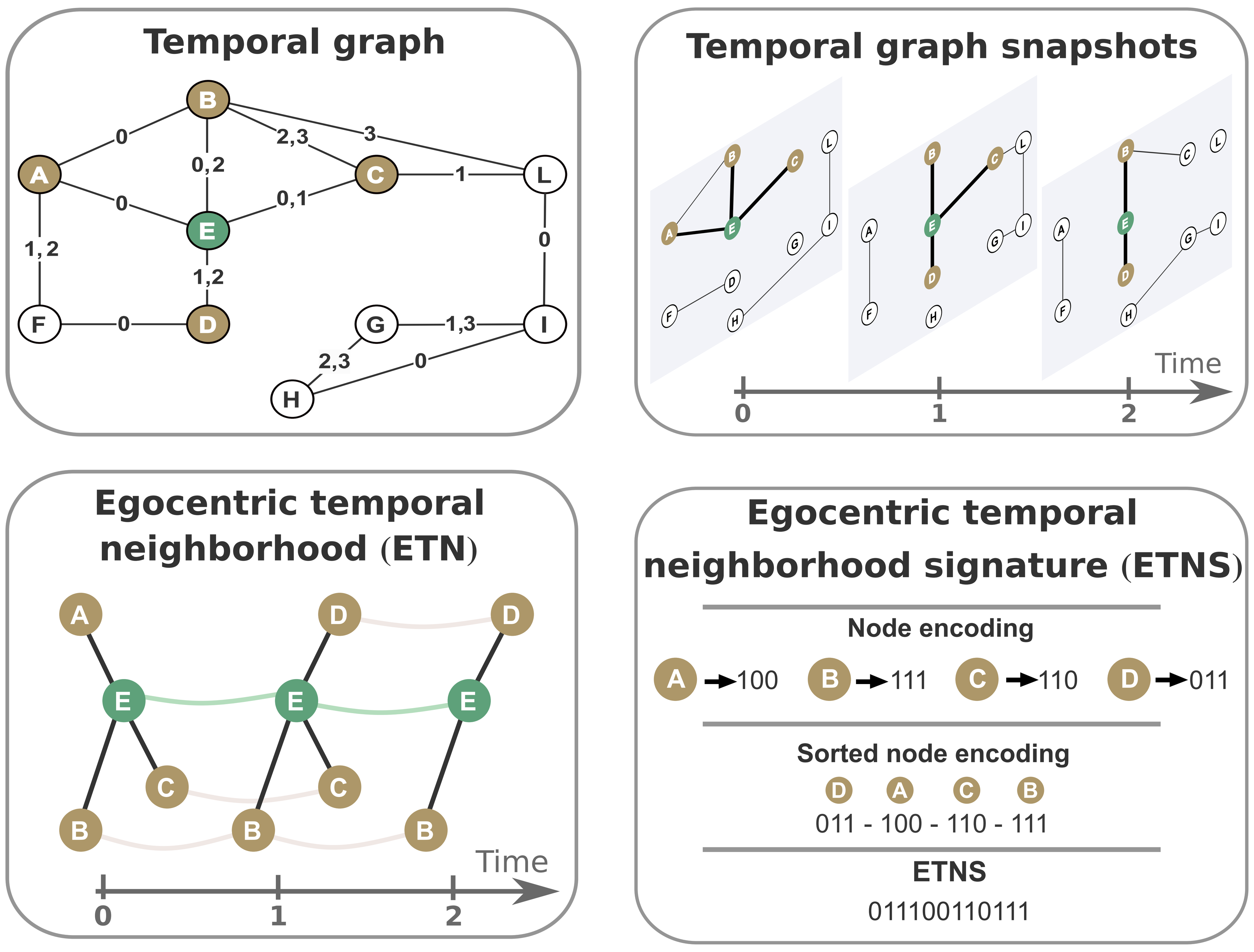
Egocentric Temporal Motifs and Temporal Graph Generation

A. Longa, G. Cencetti, A. Passerini and B.Lepri
a.longa@fbk.eu https://antoniolonga.github.io/

Introduction

A decade ago, researchers 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 idea of our Egocentric Temporal Motifs Miner is to jump inside the network and follow the path of a specific node, finding node-dependent spatio-temporal patterns.

Extract Egocentic Temporal Neighborhood



Given a temporal graph G, and a temporal gap delta T, we represent the graph as an ordered sequence of temporal aggregations.

- Given an ego node E and a temporal order k (k = 2) , we extract ETN from the input graph.
- ETN can be encoded in an Egocentric Temporal Neighborhood Signature (ETNS)

From ETN to ETM

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

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

Definitions

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

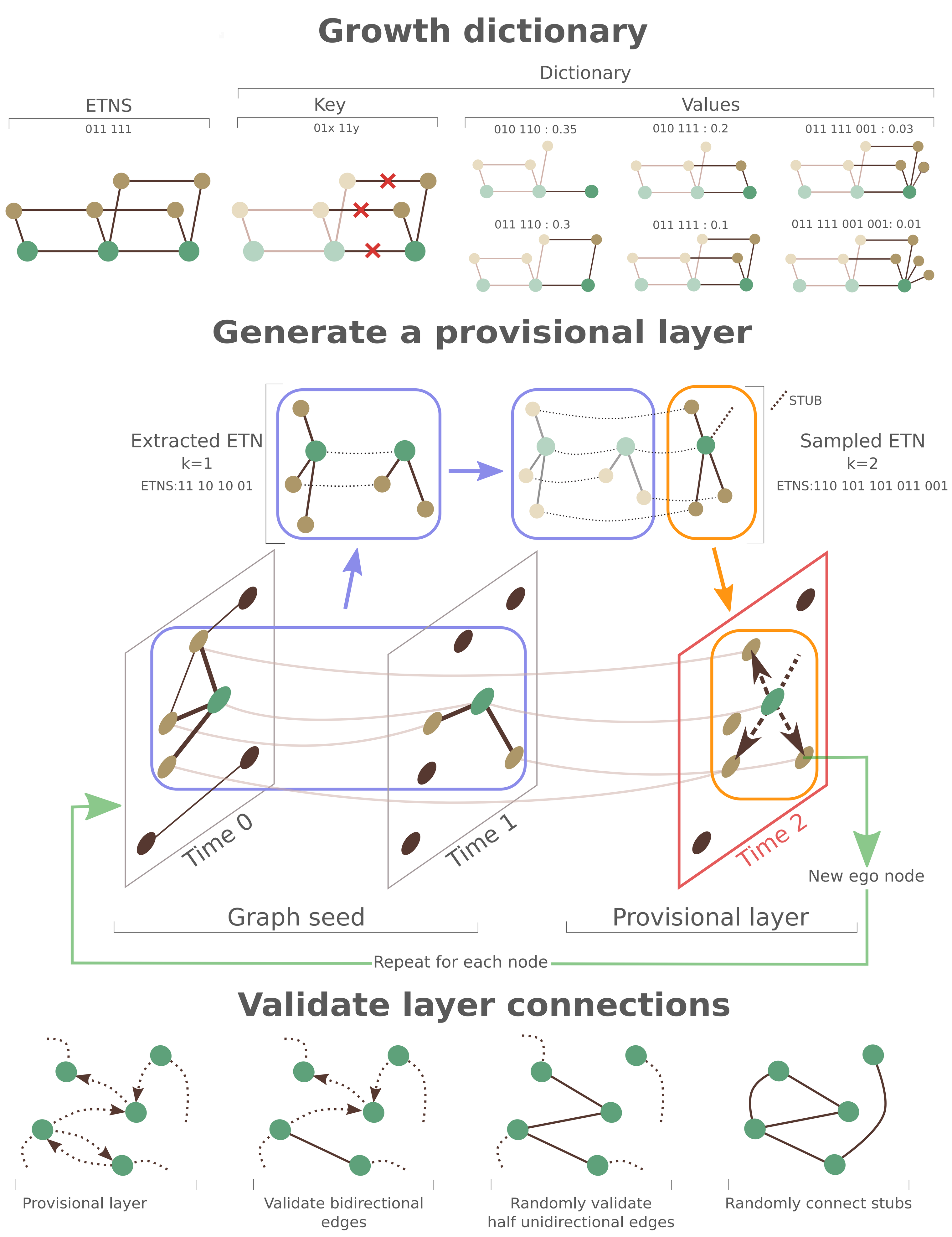
Results

	Workplace	Hospital	HS11	HS12	HS13	PS	DTU
Workplace	0.00	0.07	0.29	0.22	0.29	0.67	0.47
Hospital		0.00	0.29	0.22	0.30	0.66	0.45
High school 11			0.00	0.04	0.04	0.59	0.06
High school 12				0.00	0.02	0.61	0.13
High school 13					0.00	0.62	0.08
Primary school						0.00	0.62
DTU blue							0.00

		Calls		SMS		Email	
		DTU C	**Friend**	**Dtu S**	**Friend**	**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

		Erdos Renyi		Scale free		Small world	
		P 0.01	**P 0.001**	**g1**	**g2**	**p2 k3**	**p8 k3**
Erdos Renyi	**P 0.01**	0.00	0.03	0.07	0.11	0.37	0.18
	P 0.001		0.00	0.09	0.15	0.46	0.25
Scale free	**g1**			0.00	0.03	0.39	0.14
	g2				0.00	0.30	0.09
Small world	**p2 k3**					0.00	0.14
	p8 k3						0.00

Egocentric Temporal Neighborhood for Temporal Graph Generation

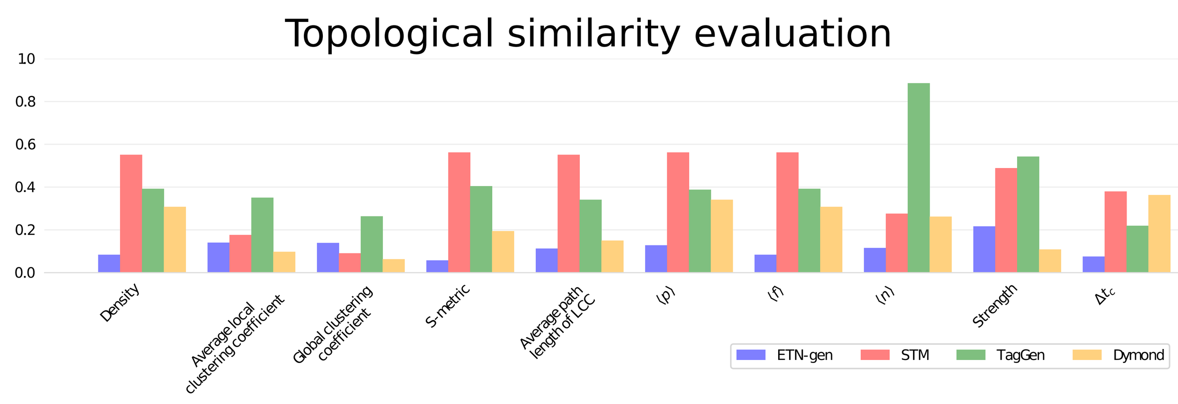


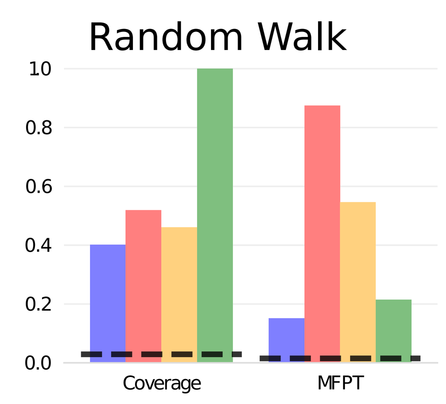
ETN-Gen

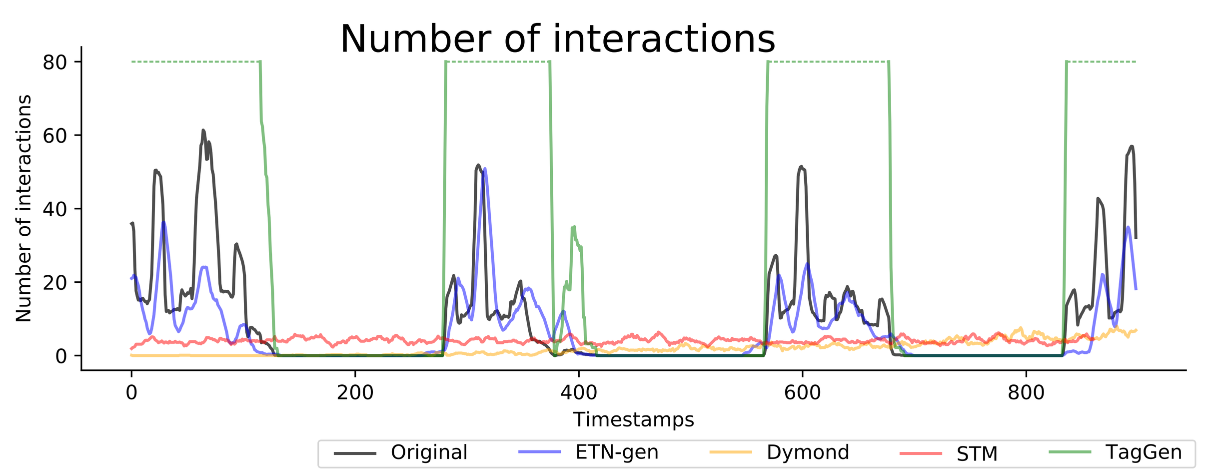
- Builty a **growth dictionary**, using as a **key ETNS with k-1** layers. The **values** are all possible **ETNS with k layers** and it's normilized frequencies.
- Starting from a graph seed, **extract ETNS** with k-1 layers, **query the dictionary**, and **generate a provisional layer**, repeat for each node.
- Finally, **validate bidirectional edges**, randomly **validate half edges**, and randomly **connect stubs.**

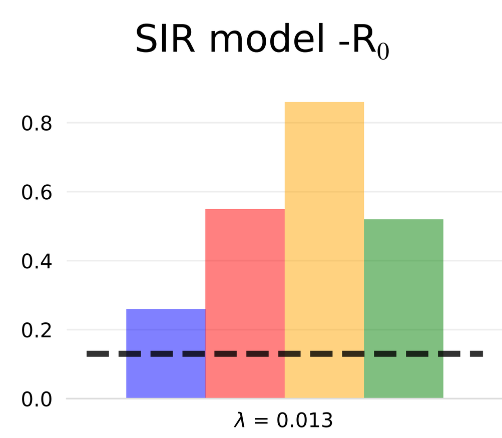
Results

- Topological and dynamical similarity evaluation









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