Neighbourhood matching creates realistic surrogate temporal networks

Antonio Longa[1,2], Giulia Cencetti[1], Sune Lehmann[3,4], Andrea Passerini[2] and Bruno Lepri[1]
[1]Fondazione Bruno Kessler, Trento, Italy
[2]University of Trento, Trento, Italy
[3]Tecnical University of Denmark, Kongens Lyngby, Denmark
[4]Copenhagen Center for Social Data Science, Copenhagen, Denmark

Temporal networks are essential for modeling and understanding systems whose behavior varies in time, from social interactions to biological systems. Often, however, real-world 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. Until now, the generation of realistic surrogate temporal networks has remained an open problem, due to the difficulty of capturing both the temporal and topological properties of the input network, as well as their correlations, in a scalable model.

In this work[2], we propose a method able to generate high temporal resolution surrogate networks that are able to match real-networks in terms of a wide range of topological and dynamic measures. Our generative algorithm is based on the idea of the *egocentric temporal neighborhood*[1] $\mathcal{E}_n^{\{t-k,\dots,t\}}$ for node n at time t, including a small number k of prior time steps. Here we assume that the network is represented in discrete time with each time step corresponding to a static graph, also referred to as a 'layer' of the network. Crucially, $\mathcal{E}_n^{\{t-k,\dots,t\}}$ does not include interactions between the neighbors of n.

A major advantage of the egocentric perspective (that ignores connections among neighbors of an ego node) is that it allows us to linearize the concept of node neighborhood sidestepping the subgraph isomorphism problem, making the generation process fast and scalable both in terms of the number of nodes and the number of temporal snapshots. Speed turns out to be a fundamental feature, because the other existing methods rely on algorithms of considerably higher complexity that prevent those methods from scaling to even moderately-sized networks.

We test the method, named *Egocentric Temporal Neighborhood Generator (ETN-gen)*, on a range of different temporal networks. In our testing, we mainly use social interactions datasets, because of richness and availability of these datasets, but the method is general and can be used to generate any kind of graph. The simplicity of our algorithm makes it easily interpretable, extendable and algorithmically scalable. In the Fig. 1, the surrogate networks that we generate match original networks with a high degree of accuracy, not only in terms magnitude of interaction, but also respect the daily-night periodicity. The ability to generate surrogate temporal graphs that reproduce real behaviors allows us to obtain large as desired data, without resolution limits.

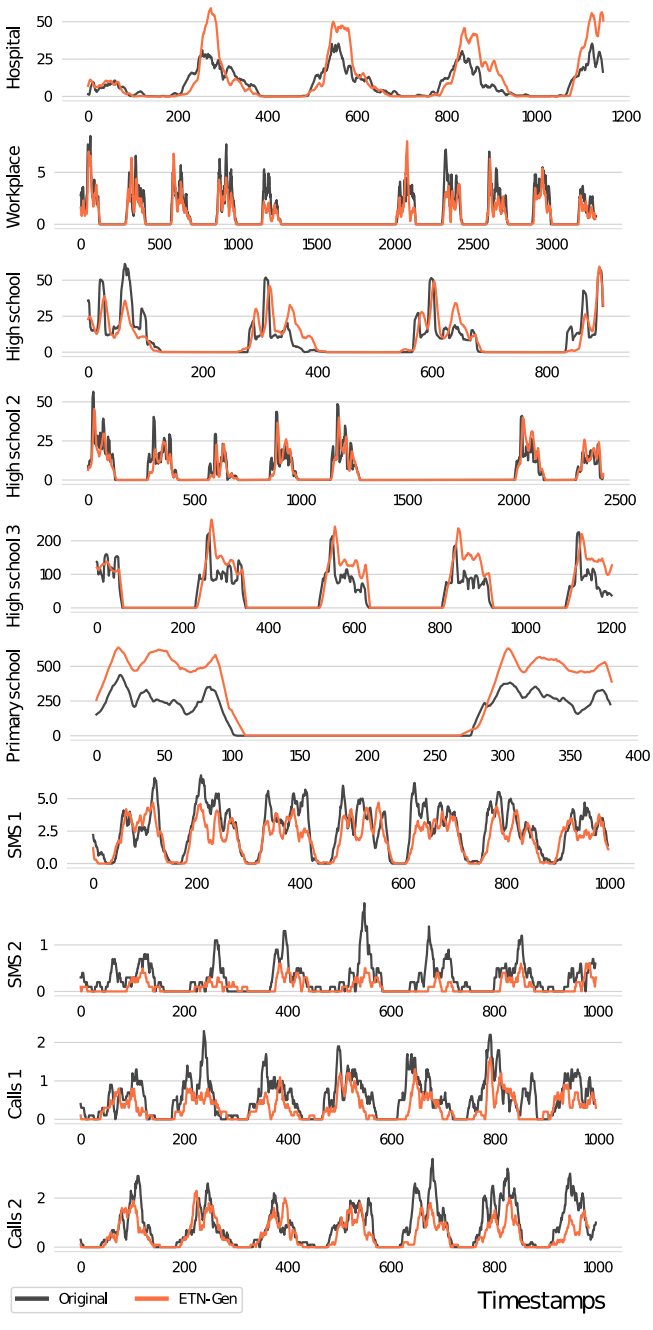

Fig. 1. **Number of interactions in the generated network for different datasets.** Each panel shows the number of interactions of the original (black curve) and *ETN-gen* (orange curve) graphs. We use a temporal gap of 5 minutes for face-to-face interactions and 10 minutes for calls and SMS (intrinsically sparser networks).

[1] A. Longa, G. Cencetti, B. Lepri and A. Passerini, *An efficient procedure for mining egocentric temporal motifs*, Data Mining and Knowledge Discovery **36**, 355-378 (2022).

[2] A. Longa, G. Cencetti, S. Lehmann, A. Passerini and B. Lepri *Neighbourhood matching creates realistic surrogate temporal networks*, ArXiv (2022).