Community-Aware Temporal Network Generation

Nicolò Alessandro Girardin[1], Antonio Longa[2], Gaia Trebucchi[*], Giulia Cencetti[3], Andrea Passerini[2], and Bruno Lepri[1]

[1]*Fondazione Bruno Kessler, Trento, Italy*
[2]*University of Trento, Trento, Italy*
[3]*Aix-Marseille Univ, Université de Toulon, CNRS, CPT, Marseille, France*
[*]*Work done at the University of Trento*

Keywords: face-to-face interactions, temporal networks, surrogate networks, communities

Extended Abstract

The advantages of temporal networks in capturing complex dynamics, such as diffusion and contagion, has led to breakthroughs in real world systems across biology, communications, social interactions, and mobility. In the case of human behaviour, face-to-face interaction networks enable us to understand the dynamics of how communities emerge and evolve in time through the interactions, which is crucial in fields like epidemics, sociological studies and urban science. However, state-of-the-art datasets in this field suffer from a number of drawbacks. The data collection is often short in time and involves a small amount of participants. Moreover, concerns arise for the participants' privacy and the data collection costs. Thus, it is common practice to extend them with repeating patterns of the original dataset, often leading to biases.

Therefore, the generation of *surrogate temporal networks* that can mimic real-world interactions, also in scale, is a crucial issue. Over the past years, a large number of successful algorithms for static networks generation have been proposed. However, extending these models to the dynamic regime has proven prohibitively difficult, due to the increased complexity introduced by the temporal dimension. Moreover, these methods mostly ignore the social structure of human interactions. In fact, human tend to interact in communities, such as in schools, where the students mostly interact with their class.

In our work, we tackle this problem by assigning labels to the nodes of the network, which correspond to the community they belong to. This leads to the ability of capturing how interactions evolve between different communities: specifically, we test the method in several settings in which individuals interact both inside well-defined communities and also across them. Our results show that our method well captures the evolution of these interactions over time as well as the modularity of the networks.

Methods Here we represent temporal networks in discrete time with each time step corresponding to a static graph, also referred to as a 'layer' of the network. We also consider the network from a egocentric perspective, following [2].

In this perspective, we consider a (short) window of k consecutive snapshot to capture the interactions each ego node has over time, calling this the *labelled egocentric temporal neighbourhood (LETN)* of the node. From this, as in Panel **A** in Figure 1, we introduce a novel mapping of the label of each node as a binary representation, and construct a string for each of the neighbours. For each time step, we add to this string the neighbour's binary label if the connection with the ego and the node is present, otherwise with a 0. The final vector of the neighbourhood (i.e., *LETN signature, LETNS*) will be the concatenation of the encoding of the individual neighbours, organized following the lexicographic ordering.

Following this, we compute local probability distributions of the LETNS. Then, we mask the last layer of the LETNS, keeping the $k-1$ mask representation as a key for the dictionary, and compute the distribution of the candidate extensions, as in Panel **B** of Figure 1.

To capture the periodicity of interactions of the original network, we split the original layers in a *local split* and a *global split*. While the global split represents a cycle of longer time (e.g., a day), the local split represents fine-grained cycles that can repeat inside of a global one (e.g., the hours of a day). To do so, the probability dictionaries are computed by aggregating the networks at a local split level. This means that, in the case of hours as local split and days as global split, we have a dictionary for each hour of the day, which counts the LETNS present in the same hour-window over all the days present in the network.

The generation method, which we call *LETN-gen*, starts from a seed of $k-1$ original layers and elects the following layer sampling from the computed probability dictionaries for each ego node, thus creating a *provisional layer*, as in Panel **C** in Figure 1. In this provisional layer, each node has a preference for attachment to nodes with a specific label. Therefore, we validate bidirectional edges in the first place. Then, we randomly validate half of the unidirectional edges. Finally, we also connect the stubs, maximizing consensus for their preferred attachment (e.g., node with label 01 wants to interact with label 10 and vice-versa). The validation process is summarized in Panel **D** of Figure 1. The process will procede iteratively until the desired number of layers is obtained.

When the labels of the nodes are not present, we run the Louvain algorithm for community detection [1], which maximizes modularity, on the aggregated network. Following this, we assign a labels to nodes depending on their resulting partition, and procede as above. We call this extension *Community LETN-gen (CLETN-gen)*.

Results We tested our methodology, focusing on metrics relative to interactions and communities on several datasets. Table 1 shows the results for the structural metrics on tested face-to-face interaction datasets. We see that in most cases, our method is able to reproduce well the original network's characteristics, also improving on ETN-gen

Here we present in detail the results for the *primary school* dataset, in which students' classes are the labels with the addition of teachers.

Firstly, we look at the generated aggregate networks in Panel **A** of Figure 2. Our method (LETN-gen and CLETN-gen) reproduces a high number of interactions within communities and a lower number across communities, similar to the original network, while the base method (ETN-gen) is not able to capture these dynamics. This is also shown in the frequency of interactions among and across communities shown in Panel **B**, were ETN-gen has a more sparse heatmap than our method, which, in general, has more interactions inside communities.

Secondly, we see that our method is able to replicate well both the number of interactions and the periodicity of the original network, as shown in Panel **C** of Figure 2. Interestingly, the original method overestimates the number of interactions, while our method does not. Here, note that the lag we see in the

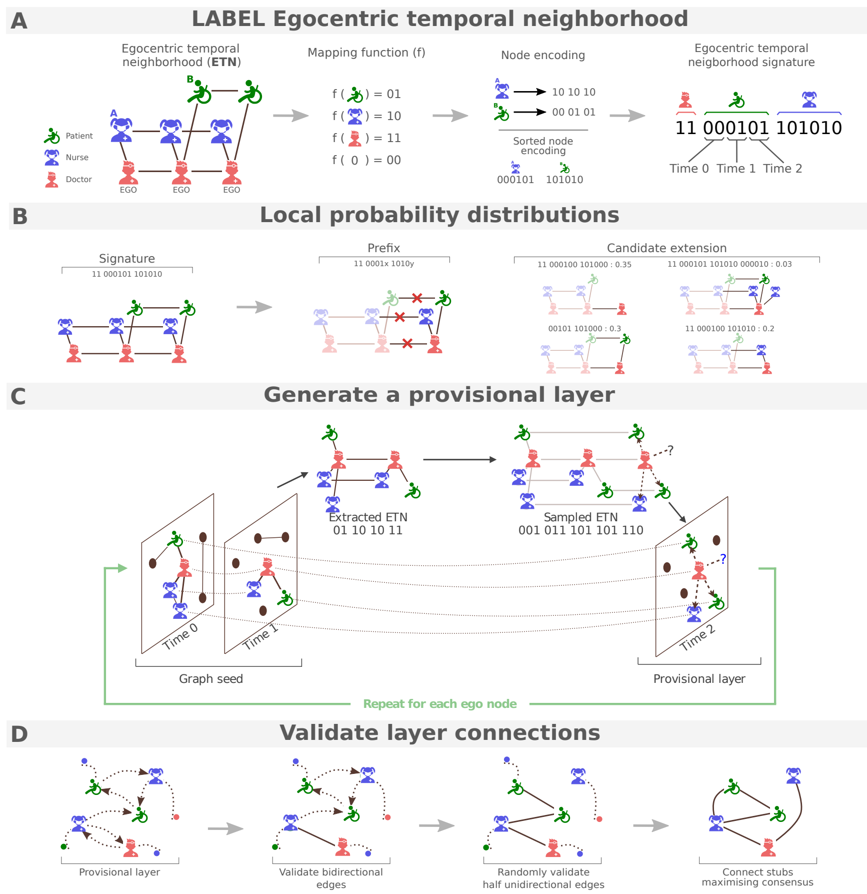


Figure 1: Algorithm for surrogate network generation. Panel **A** describes how the LETN signature is computed, with the mapping of the nodes to their label binary representation, the node encoding when the links are present and the ordered string representing the signature. Panel **B** describes how by eliminating the last snapshot of the signature we create a dictionary with the distribution of the candidate extensions. Panels **C** and **D** show the generative process, with the generation of a provisional layer, sampling from the dictionaries and the validation of the links

number of interactions is due to the local split we chose for the temporal neighbourhood.

Finally we look at the evolution of modularity in time, displayed in Panel **D** of Figure 2. We see that our method reproduces the modularity evolution over time closely to the original network, while the previous method is not able to capture this dynamic. Note that all the methods, including ETN-gen, start with a high modularity: this is because we chose the initial seeds from the original dataset.

Take together, these results show how the generation of surrogate temporal networks with our methodology is able to reproduce the properties of real networks. Moreover, when focusing on interactions, especially among and across communities, our method is able to capture these dynamics of the original network and maintain them in the generated ones. This method can be useful in situations where the data collection can be cumbersome, either for privacy or economical reasons. Thus, fields such as epidemics, human dynamics and urban science can benefit from the use of surrogate networks generated by our method, both for longer time spans and the addition of more individuals to better understand communities interactions.

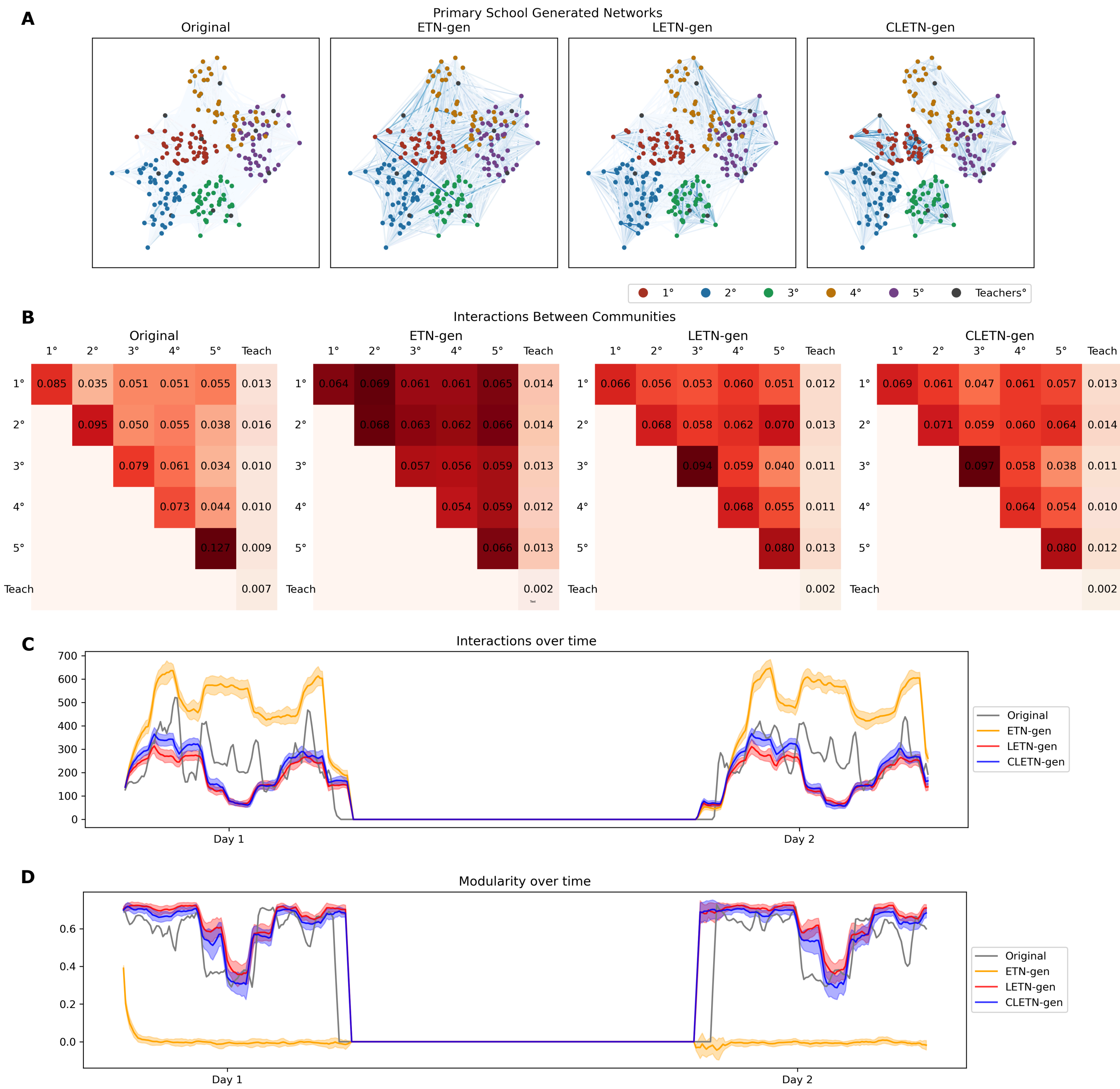
Table 1: Structural Metrics on Tested Datasets

dataset	model	density	clustering coeff.	assortativity	modularity
Primary School	original	0.29	0.53	0.12	0.59
Primary School	ETN	0.68	0.69	-0.01	0.00
Primary School	LETN	0.29	0.56	0.01	0.69
Primary School	CLETN	0.25	0.55	0.19	0.67
High School 2011	Original	0.22	0.58	0.08	0.56
High School 2011	ETN	0.23	0.24	-0.03	0.00
High School 2011	LETN	0.34	0.58	0.22	0.56
High School 2011	CLETN	0.34	0.59	0.22	0.49
High School 2012	Original	0.14	0.48	0.05	0.67
High School 2012	ETN	0.34	0.35	-0.02	-0.00
High School 2012	LETN	0.26	0.57	0.19	0.67
High School 2012	CLETN	0.27	0.54	0.20	0.49
High School 2013	Original	0.11	0.50	0.03	0.80
High School 2013	ETN	0.35	0.36	-0.01	0.01
High School 2013	LETN	0.16	0.53	0.22	0.81
High School 2013	CLETN	0.17	0.51	0.19	0.72
Hospital	Original	0.41	0.64	-0.18	0.14
Hospital	ETN	0.82	0.84	-0.05	-0.01
Hospital	LETN	0.74	0.84	-0.27	0.13
Hospital	CLETN	0.78	0.80	-0.02	0.07
Lyon School	Original	0.29	0.53	0.12	0.59
Lyon School	ETN	0.66	0.66	-0.01	-0.00
Lyon School	LETN	0.23	0.44	-0.04	0.74
Lyon School	CLETN	0.27	0.53	0.13	0.41
Workplace 2013	Original	0.18	0.43	-0.06	0.55
Workplace 2013	ETN	0.34	0.34	-0.034	-0.01
Workplace 2013	LETN	0.38	0.49	0.22	0.55
Workplace 2013	CLETN	0.36	0.52	0.03	0.43
Workplace 2015	Original	0.18	0.38	0.04	0.57
Workplace 2015	ETN	0.39	0.40	-0.01	-0.00
Workplace 2015	LETN	0.31	0.42	0.09	0.62
Workplace 2015	CLETN	0.33	0.41	0.14	0.45

References

[1] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008, 2008.

[2] A Longa, G Cencetti, S Lehmann, A Passerini, and B Lepri. Generating fine-grained surrogate temporal networks. *Communications Physics*, 7(1):22, 2024.



Original	1°	2°	3°	4°	5°	Teach
1°	0.085	0.035	0.051	0.051	0.055	0.013
2°		0.095	0.050	0.055	0.038	0.016
3°			0.079	0.061	0.034	0.010
4°				0.073	0.044	0.010
5°					0.127	0.009
Teach						0.007

ETN-gen	1°	2°	3°	4°	5°	Teach
1°	0.064	0.069	0.061	0.061	0.065	0.014
2°		0.068	0.063	0.062	0.066	0.014
3°			0.057	0.056	0.059	0.013
4°				0.054	0.059	0.012
5°					0.066	0.013
Teach						0.002

LETN-gen	1°	2°	3°	4°	5°	Teach
1°	0.066	0.056	0.053	0.060	0.051	0.012
2°		0.068	0.058	0.062	0.070	0.013
3°			0.094	0.059	0.040	0.011
4°				0.068	0.055	0.011
5°					0.080	0.013
Teach						0.002

CLETN-gen	1°	2°	3°	4°	5°	Teach
1°	0.069	0.061	0.047	0.061	0.057	0.013
2°		0.071	0.059	0.060	0.064	0.014
3°			0.097	0.058	0.038	0.011
4°				0.064	0.054	0.010
5°					0.080	0.012
Teach						0.002

Figure 2: Results for the *primary school* dataset. Panel **A** shows the aggregated networks, for which we see that both LETN-gen and CLETN-gen respect the interactions among communities. This is confirmed in Panel **B**, in which the heatmaps represent the fraction of interactions among the original communities. We see that for ETN-gen the links are similarly distributed among links inside and across communities, while our methods respect much more this distribution and are thus closer to the original network. The interactions over time are shown in Panel **C**, where ETN-gen overestimates the number of links, while out methods are close to the original (rolling average with window 2, over 10 generated networks). Finally, Panel **D** shows the modularity over time (rolling average with window 2), which is very close to the original for out merhods, while for ETN-gen it is always low.