



HAL
open science

Recovering Communities in Temporal Networks Using Persistent Edges

Konstantin Avrachenkov, Maximilien Drevetton, Lasse Leskelä

► **To cite this version:**

Konstantin Avrachenkov, Maximilien Drevetton, Lasse Leskelä. Recovering Communities in Temporal Networks Using Persistent Edges. CSoNet 2021 - International Conference on Computational Data and Social Networks, Nov 2021, Montreal, Canada. pp.243-254, 10.1007/978-3-030-91434-9_22 . hal-03472963

HAL Id: hal-03472963

<https://inria.hal.science/hal-03472963>

Submitted on 9 Dec 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Recovering communities in temporal networks using persistent edges^{*}

Konstantin Avrachenkov¹, Maximilien Drevetton¹, and Lasse Leskelä²

¹ Inria Sophia Antipolis, 2004 Route des Lucioles, 06902 Valbonne, France
{k.avrachenkov,maximilien.drevetton}@inria.fr

² Aalto University, Department of Mathematics and Systems Analysis,
Otakaari 1, 02150 Espoo, Finland
lasse.leskela@aalto.fi

Abstract. This article studies the recovery of static communities in a temporal network. We introduce a temporal stochastic block model where dynamic interaction patterns between node pairs follow a Markov chain. We render this model versatile by adding degree correction parameters, describing the tendency of each node to start new interactions. We show that in some cases the likelihood of this model is approximated by the regularized modularity of a time-aggregated graph. This time-aggregated graph involves a trade-off between new edges and persistent edges. A continuous relaxation reduces the regularized modularity maximization to a normalized spectral clustering. We illustrate by numerical experiments the importance of edge persistence, both on simulated and real data sets.

Keywords: graph clustering · temporal networks · spectral methods · stochastic block model

1 Introduction

Complex networks are commonly used to describe and analyze interactions between entities. A natural problem arising consists in identifying meaningful structures within the complex system. Community recovery, i.e., partitioning the set of nodes of a network into *communities* based on some common properties of the vertices, is now a well established area [6].

In many situations, interactions between node pairs vary over time, and classical graph-based models are replaced by temporal networks models [10]. Examples include communication, interaction, and transportation networks. The longitudinal dimension of data raises new challenges to traditional clustering

^{*} This work has been done within the project of Inria - Nokia Bell Labs “Distributed Learning and Control for Network Analysis” and was partly supported by COST-NET EU Cost Action CA15109. This is the author version of the paper accepted at the 10th International Conference on Computational Data and Social Networks (CSoNet 2021), Springer LNCS v.13116, pp. 243-254.

algorithms. Previous research has focused on evolving communities [22], for example by generalizing belief-propagation methods [8], developing variational EM algorithms [15] and introducing new spectral methods [5,13] or modularity-based methods [16,20]. Nevertheless, all of the aforementioned works focus on evolving communities for which the interactions between nodes are re-sampled at every time step. One can then treat each layer independently by applying static community detection and smoothing the community predictions. When the communities are non-evolving, the extra longitudinal dimension brings new information, and each additional snapshot makes the clustering easier. The recovery bounds established in [1] highlight this fact. Nonetheless, simple temporal aggregation of the data might lose important features such as temporal patterns. As such, [1] proposes an online algorithm tailored for Markov edge evolution dynamics, while [2] studies a spectral algorithm using the squared adjacency matrix.

In this work, we introduce a temporal extension of the degree-corrected stochastic block model [9,11], in which the community labeling is fixed and the interactions between node pairs follow a Markov evolution which only depends on the community labeling and on the degree correction parameters. To the best of our knowledge, we are the first to introduce degree-corrected parameters into temporal network models with edge persistency. We show that the maximum likelihood inference reduces to the maximization of the regularized modularity of a time-aggregated graph, in the limit of a large number of snapshots and sparse interactions. This graph is not simply the sum of the adjacency matrices over all snapshots. Instead, it involves a trade-off between the newly formed edges and the persistent ones, and this trade-off depends on the difference between the edge-persistence between intra-community and inter-community node pairs. A continuous relaxation then leads to a normalized spectral clustering algorithm. Finally, we validate the importance of taking into account the persistent edges in simulated and real data sets.

Notations Throughout this article, matrices are represented by capital letters (A , W , etc.), and the corresponding matrix elements by A_{ij} , W_{ij} , etc. $\text{Tr } A$ denotes the trace of a square matrix A , and $A \odot B$ denotes the entrywise product of two matrices. Finally, $\mathbf{1}_n$ is the n -by-1 vector of all ones, and the indicator of an event B is denoted by $\mathbf{1}(B)$. The Kronecker delta is denoted by $\delta(x, y)$, so that $\delta(x, y) = \mathbf{1}(x = y)$.

By a slight abuse of notations, graphs are represented by their adjacency matrix A . We will assume that graphs are undirected but potentially weighted, hence A is symmetric, with non-negative entries. Clustering a graph with N nodes into K clusters accounts to assigning to each node i a label $Z_i \in [K]$.

2 Degree-corrected temporal network model with Markov edge dynamics

Consider a population of N nodes partitioned into K static communities such that node i belongs to community $Z_i \in [K]$. We write $A_{ij}^t = 1$ if nodes i and j

interact at time t , and $A_{ij}^t = 0$ otherwise. We investigate methods of recovering the community structure $Z = (Z_1, \dots, Z_N)$ from an observed adjacency tensor $A = (A_{ij}^t)$. The following section describes a versatile statistical model for this setting.

2.1 Model description

A degree-corrected temporal stochastic block model with N nodes, K blocks and T snapshots is a probability distribution

$$\mathbb{P}(A | Z, F, \theta) = \prod_{1 \leq i < j \leq N} F_{Z_i Z_j}^{\theta_i \theta_j} (A_{ij}^1, \dots, A_{ij}^T) \quad (1)$$

of a symmetric adjacency tensor $A \in \{0, 1\}^{N \times N \times T}$ with zero diagonal entries, where $Z = (Z_1, \dots, Z_N)$ is a community assignment with $Z_i \in \{1, \dots, K\}$ indicating the community of node i , $F = (F_{k\ell}^{xy})$ is a collection of probability distributions over $\{0, 1\}^T$, and $\theta = (\theta_1, \dots, \theta_N)$ is a vector of node-specific degree correction parameters, with $0 \leq \theta_i < \infty$.

In the following, we will restrict ourselves to homogeneous inter-block interactions with Markov edge dynamics, for which the nodes' static community labellings are sampled uniformly at random from the set $[K]$ of all node labellings, and

$$F_{Z_i Z_j}^{\theta_i \theta_j} (x) = \begin{cases} \mu_{x_1}^{\theta_i \theta_j} \prod_{t=2}^T P_{x_{t-1}, x_t}^{\theta_i \theta_j} & \text{if } Z_i = Z_j, \\ \nu_{x_1}^{\theta_i \theta_j} \prod_{t=2}^T Q_{x_{t-1}, x_t}^{\theta_i \theta_j} & \text{otherwise,} \end{cases} \quad (2)$$

with initial distributions

$$\mu^{\theta_i \theta_j} = \begin{pmatrix} 1 - \theta_i \theta_j \mu_1 \\ \theta_i \theta_j \mu_1 \end{pmatrix} \quad \text{and} \quad \nu^{\theta_i \theta_j} = \begin{pmatrix} 1 - \theta_i \theta_j \nu_1 \\ \theta_i \theta_j \nu_1 \end{pmatrix}, \quad (3)$$

and transition probability matrices

$$P^{\theta_i \theta_j} = \begin{pmatrix} 1 - \theta_i \theta_j P_{01} & \theta_i \theta_j P_{01} \\ 1 - P_{11} & P_{11} \end{pmatrix} \quad \text{and} \quad Q^{\theta_i \theta_j} = \begin{pmatrix} 1 - \theta_i \theta_j Q_{01} & \theta_i \theta_j Q_{01} \\ 1 - Q_{11} & Q_{11} \end{pmatrix}. \quad (4)$$

The parameters θ_i account for the fact that some nodes might be more inclined than others to start new connections, similarly to the degree-corrected block model of [11]. To keep the model simple, we do not add degree correction parameters in front of P_{11} ; hence once a connection started, the probability to keep it active is simply P_{11} or Q_{11} . Moreover, we assume that $\min_{i,j} \{\theta_i \theta_j \delta\} \leq 1$, where $\delta = \max\{\mu_1, \nu_1, P_{01}, Q_{01}\}$. Finally, we normalise the degree correction parameters so that $\sum_i 1(Z_i = k) \theta_i = \sum_i 1(Z_i = k)$ for all k .

2.2 Maximum likelihood estimator

Proposition 1. *A maximum likelihood estimator for the Markov block model defined by (1)–(2) is any community assignment $Z \in [K]^N$ that maximizes*

$$\sum_{i,j} \delta(Z_i, Z_j) \left\{ A_{ij}^1 \left(\rho_1^{\theta_i \theta_j} - \rho_0^{\theta_i \theta_j} \right) + \rho_0^{\theta_i \theta_j} + (A_{ij}^1 - A_{ij}^T) \ell_{10}^{\theta_i \theta_j} \right\} \\ + \sum_{i,j} \delta(Z_i, Z_j) \sum_{t=2}^T \left\{ \left(\ell_{01}^{\theta_i \theta_j} + \ell_{10}^{\theta_i \theta_j} \right) (A_{ij}^t - A_{ij}^{t-1} A_{ij}^t) + \ell_{11}^{\theta_i \theta_j} A_{ij}^{t-1} A_{ij}^t - \log \frac{Q_{00}^{\theta_i \theta_j}}{P_{00}^{\theta_i \theta_j}} \right\}$$

where $\rho_a^{\theta_i \theta_j} = \log \frac{\mu_a^{\theta_i \theta_j}}{\nu_a^{\theta_i \theta_j}}$ and $\ell_{ab}^{\theta_i \theta_j} = \log \frac{P_{ab}^{\theta_i \theta_j}}{Q_{ab}^{\theta_i \theta_j}} - \log \frac{P_{00}^{\theta_i \theta_j}}{Q_{00}^{\theta_i \theta_j}}$.

The proof of Proposition 1 is presented in Appendix A.1. The MLE derived in Proposition 1 is more complex than summing all snapshots independently. In particular, the terms $A_{ij}^{t-1} A_{ij}^t$ account for *persistent edges* over two consecutive snapshots. Denote by $A_{\text{pers}}^t = A^{t-1} \odot A^t$ the entrywise product of adjacency matrices A^{t-1} and A^t . Then A_{pers}^t is the adjacency matrix of the graph containing the persistent edges between $t-1$ and t , and $A_{\text{new}}^t = A^t - A_{\text{pers}}^t$ corresponds to the graph containing the edges freshly appearing at time t .

Assuming that the number of snapshots T is large, we can ignore the boundary terms, and the MLE expressed in Proposition 1 reduces to maximizing

$$\sum_{t=2}^T \sum_{ij: Z_i=Z_j} \left(\left(\ell_{01}^{\theta_i \theta_j} + \ell_{10}^{\theta_i \theta_j} \right) (A_{ij}^t - A_{ij}^{t-1} A_{ij}^t) + \ell_{11}^{\theta_i \theta_j} A_{ij}^{t-1} A_{ij}^t - \log \frac{Q_{00}^{\theta_i \theta_j}}{P_{00}^{\theta_i \theta_j}} \right).$$

By utilising (3)–(4), we can further simplify it to express this as a modularity. Recall given a weighted graph W , a partition Z and a resolution parameter γ , the regularized modularity is defined as [19,21]

$$\mathcal{M}(W, Z, \gamma) = \sum_{i,j} \delta(Z_i, Z_j) \left(W_{ij} - \gamma \frac{d_i d_j}{2m} \right)$$

where $d_i = \sum_j W_{ij}$ and $m = \sum_i d_i$. Hence, suppose that $P^{\theta_i \theta_j}$ and $Q^{\theta_i \theta_j}$ are nondegenerate, and $\mu^{\theta_i \theta_j}$ (resp. $\nu^{\theta_i \theta_j}$) is the stationary distribution of $P^{\theta_i \theta_j}$ (resp. $Q^{\theta_i \theta_j}$). In a sparse setting, P_{01} and Q_{01} are small, and after a Taylor expansion (see Appendix A.2 for the full derivations) the previous expression is approximately equal to $\mathcal{M}(W, Z, \gamma)$, where W is defined by

$$W = \sum_{t=2}^T (\alpha A_{\text{new}}^t + \beta A_{\text{pers}}^t) \quad (5)$$

with

$$\alpha = \log \frac{P_{01}}{Q_{01}} + \log \frac{1 - P_{11}}{1 - Q_{11}} \quad \text{and} \quad \beta = \log \frac{P_{11}}{Q_{11}}, \quad (6)$$

and $\gamma = (P_{01} - Q_{01}) \frac{\alpha(\mu_1 + (K-1)\nu_1) + (\beta - \alpha)(\mu_1 P_{11} + (K-1)\nu_1 Q_{11})}{K}$.

Comparison with previous work Correspondence between maximum likelihood estimator and modularity maximization are long known in static block models [18]. Analogously to the single-layer case, the modularity of a temporal network, with possibly time-dependent community structure, was previously defined in [16,20] by

$$\sum_{t=1}^T \mathcal{M}(A^t, Z^t, \gamma_t) + \sum_{t=1}^T \sum_{s \neq t} \sum_i \omega_i^{st} \delta(Z_i^s, Z_i^t) \quad (7)$$

where γ_t is the resolution parameter for layer t , Z_i^t is the community membership of node i at time step t , and ω_i^{st} denotes a coupling between time instants s and t . For a static community structure, the second term in (7) is irrelevant. When the resolution is constant over time, the relevant term in (7) can be written as

$$\sum_{t=1}^T \mathcal{M}(A^t, Z, \gamma) = \mathcal{M}(A^{\text{agg}}, Z, \gamma),$$

where $A^{\text{agg}} = \sum_{t=1}^T A^t$ is the weighted adjacency matrix of the time-aggregated data. In contrast, the matrix W in (5) involves a trade-off between new edges and persistent edges. We notice that $W = A^{\text{agg}}$ only if $\alpha = \beta = 1$.

2.3 Temporal spectral clustering combining new and persistent edges

Following our analysis in Section 2.2, the community prediction should verify

$$\hat{Z} = \arg \max_{Z \in [K]^N} \mathcal{M}(W, Z, \gamma)$$

where W is defined in Equation (5) and γ is a proper resolution parameter. This optimisation problem is NP-complete in general [4], but can be approximately solved by continuous relaxation. We can choose the relaxation so that the optimization problem reduces to normalized spectral clustering algorithm on the weighted graph W (we refer to [17] and to the Appendix for the full computations). We note that in order to compute the normalized Laplacian of W , we should restrict $\alpha, \beta \geq 0$, which is not necessarily guaranteed by Formula (6). We summarize this in Algorithm 1.

3 Numerical experiments

The Python source code for reproducing our results is available online³.

³ <https://github.com/mdreveton/Spectral-clustering-with-persistent-edges>

Algorithm 1: Spectral clustering for temporal networks with Markov edge dynamics and static node labeling.

Input: Adjacency matrices A^1, \dots, A^T , number of clusters K , parameter α, β .

Output: Predicted membership matrix $\hat{Z} \in \mathcal{Z}_{N,K}$

1

Process:

- Let $W = \sum_{t=2}^T \alpha A_{\text{new}}^t + \beta A_{\text{pers}}^t$ where $A_{\text{new}}^t = A^t - A^{t-1} \odot A^t$ and $A_{\text{pers}}^t = A^{t-1} \odot A^t$;
- Compute $\mathcal{L} = I_n - D^{-1/2} W D^{-1/2}$ where $D = \text{diag}(W \mathbf{1}_n)$;
- Compute $\hat{X} \in \mathbb{R}^{N \times K}$ whose columns consist of the K orthonormal eigenvectors of \mathcal{L} associated to the K smallest eigenvalues.

2 **Return** $\hat{Z} \leftarrow \text{kmeans}(D^{-1/2} \hat{X}, K)$.

3.1 Synthetic data

Effect of persistent edges We first examine the effect of the choice of the parameters α and β in Algorithm 1. For this, we let $\alpha = 1$ and we plot in Figure 1 the averaged accuracy obtained on 25 realizations of stochastic block models with Markov edge dynamics for various β . While spectral clustering on the time-aggregated graph (corresponding to $\beta = 1$) works well, it is striking to notice that other values of β give better results. The choice of β depends on the probabilities of persistent interactions. For example, if $P_{11} > Q_{11}$ (Figure 1a), then $\beta > 1$ are preferred, while if $P_{11} < Q_{11}$ (Figure 1b) large choice of β are penalized. This is in accordance to the values of α, β derived in Formula (6) (albeit in Formula (6), α and β could be negative).

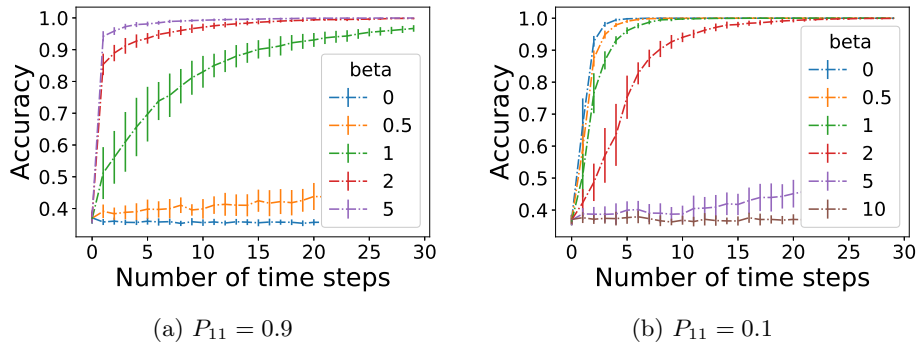


Fig. 1: Accuracy of Algorithm 1 on a SBM with 300 nodes in $K = 3$ blocks, degree correction parameters $\theta_1 = \dots = \theta_n = 1$, and a stationary Markov edge evolution $\mu_1 = 0.04$, $\nu_1 = 0.02$ and $Q_{11} = 0.3$. The results are averaged over 25 synthetic graphs, and error bars show the standard deviation.

Effect of degree correction parameters We show the robustness of Algorithm 1 on the degree correction parameters in Figure 2. More precisely:

- Figure 2a generates θ_i according to $|\mathcal{N}(0, \sigma^2)| + 1 - \sigma\sqrt{2/\pi}$ where $|\mathcal{N}(0, \sigma^2)|$ denotes the absolute value of a normal random variable with mean 0 and variance σ^2 . We choose $\sigma = 0.25$.
- Figure 2b generates the θ_i from a Pareto distribution with density function $f(x) = \frac{am}{x^{a+1}}1(x \geq m)$ with $a = 3$ and $m = 2/3$ (chosen such that $\mathbb{E}\theta_i = 1$).

Note that the sampling of the θ_i 's enforces $\mathbb{E}\theta_i = 1$ in both settings. We notice that in both cases, letting $\beta \neq 1$ improves the performance of Algorithm 1.

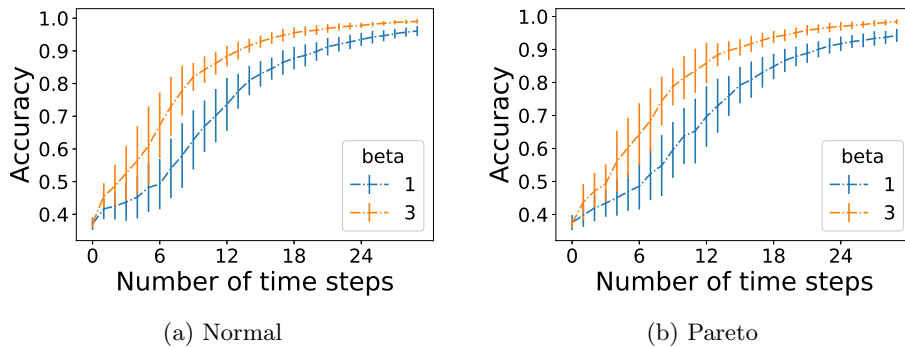


Fig. 2: Accuracy of Algorithm 1 with $\alpha = 1$ and different β , on a SBM with 300 nodes and $K = 3$ blocks (with uniform prior), and a stationary Markov edge evolution $\mu_1 = 0.06$, $\nu_1 = 0.03$, $P_{11} = 0.7$ and $Q_{11} = 0.4$, for different generation of the degree correction parameters θ . The results are averaged over 25 synthetic graphs, and error bars show the standard deviation.

3.2 Social networks of high school students

We investigate three data sets collected during three consecutive years from a high school Lycéé Thiers in Marseilles, France [7,14]. Nodes correspond to students, interactions to close-proximity encounters, and communities to classes, with dimensions given in Table 1.

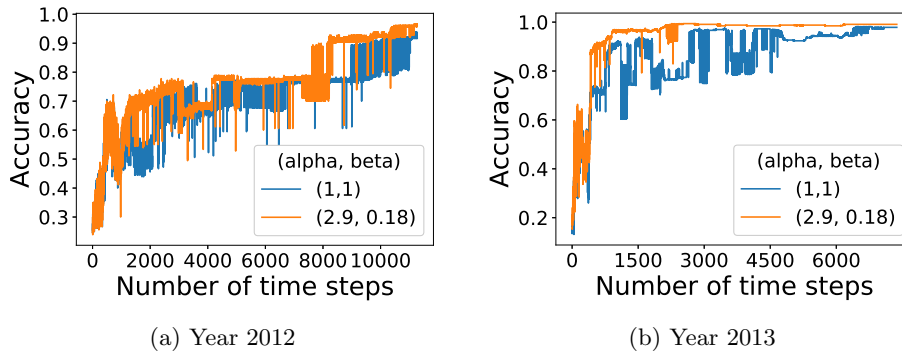
We make a hypothesis that the temporal characteristics of the interactions are similar each year. We then use the 2011 data set to estimate the transition probability matrices P and Q , and use these for clustering the 2012 and 2013 data sets. We assume that $\theta_i = 1$ (no degree correction). A standard estimator of Markov chain transition probability matrices [3] gives

$$\hat{P} = \begin{pmatrix} 0.9992 & 0.0008 \\ 0.37 & 0.63 \end{pmatrix} \quad \text{and} \quad \hat{Q} = \begin{pmatrix} 0.999967 & 3.3 \times 10^{-5} \\ 0.48 & 0.52 \end{pmatrix}.$$

| Year | N | K | T |
|------|-----|-----|-------|
| 2011 | 118 | 3 | 5609 |
| 2012 | 180 | 5 | 11273 |
| 2013 | 327 | 9 | 7375 |

Table 1: Dimensions of three data sets of interacting high school students.

Using (6), leads to $\hat{\alpha} = 2.9$ and $\hat{\beta} = 0.18$. We observe in Figure 3b that this choice of parameters gives a better accuracy on the 2013 data set than simply applying spectral clustering on the time-aggregated graph ($\alpha = \beta = 1$). For the 2012 data set (Figure 3a), this improvement is not so clearly visible.

Fig. 3: Accuracy of Algorithm 1 on the 2012 and 2013 high school datasets, using uniform $\alpha = \beta = 1$ (blue) and adjusted α, β predicted using 2011 data (orange).

To understand why Algorithm 1 performs better for 2013 than for 2012, we have listed in Table 2 temporal transition probabilities and clustering weights $\hat{\alpha}, \hat{\beta}$ estimated separately for each dataset. For year 2012, the difference between intra-community edge persistence \hat{P}_{11} and inter-community edge persistence \hat{Q}_{11} is small, implying that persistent edges do not add much extra information for distinguishing communities ($\hat{\beta} \approx 0$). For years 2011 and 2013, this difference is larger, manifesting that edge persistence contains information that can be employed to recover communities with a higher accuracy.

| Dataset | \hat{P}_{01} | \hat{Q}_{01} | \hat{P}_{11} | \hat{Q}_{11} | $\hat{\alpha}$ | $\hat{\beta}$ | $\hat{\beta}/\hat{\alpha}$ |
|---------|----------------|----------------|----------------|----------------|----------------|---------------|----------------------------|
| 2011 | 0.00080 | 0.000033 | 0.63 | 0.52 | 2.9 | 0.58 | 0.060 |
| 2012 | 0.00050 | 0.000011 | 0.57 | 0.56 | 3.8 | 0.01 | 0.003 |
| 2013 | 0.00150 | 0.000014 | 0.64 | 0.40 | 4.5 | 0.07 | 0.015 |

Table 2: Markov chain transition probabilities and adjusted clustering weights estimated separately for each dataset.

A Proofs of main statements

A.1 Maximum likelihood computations (Proposition 1)

Proof (Proof of Proposition 1). By the temporal Markov property, the log-likelihood of the model can be written as $\log \mathbb{P}(A | Z, \theta) = \log \mathbb{P}(A^1 | Z, \theta) + \sum_{t=2}^T \mathbb{P}(A^t | A^{t-1}, Z, \theta)$. By denoting $\rho_a^{\theta_i \theta_j} = \log \frac{\mu_a^{\theta_i \theta_j}}{\nu_a^{\theta_i \theta_j}}$, we find that

$$\begin{aligned} \log \mathbb{P}(A^1 | Z, \theta) &= \frac{1}{2} \sum_{i,j} \sum_a \delta(A_{ij}^1, a) \left(\delta(Z_i, Z_j) \rho_a^{\theta_i \theta_j} + \log \nu_a^{\theta_i \theta_j} \right) \\ &= \frac{1}{2} \sum_{i,j} \delta(Z_i, Z_j) \sum_a \delta(A_{ij}^1, a) \rho_a^{\theta_i \theta_j} + c_1(A), \end{aligned}$$

where $c_1(A) = \frac{1}{2} \sum_{i,j} \sum_a \delta(A_{ij}^1, a) \log \nu_a^{\theta_i \theta_j}$ does not depend on the community structure. Similarly, by denoting $R_{ab}^{\theta_i \theta_j} = \log \frac{P_{ab}^{\theta_i \theta_j}}{Q_{ab}^{\theta_i \theta_j}}$ we find that

$$\begin{aligned} \log \mathbb{P}(A^t | A^{t-1}, Z, \theta) &= \frac{1}{2} \sum_{i,j} \sum_{a,b} \delta(A_{ij}^{t-1}, a) \delta(A_{ij}^t, b) \left(\delta(Z_i, Z_j) R_{ab}^{\theta_i \theta_j} + \log Q_{ab}^{\theta_i \theta_j} \right) \\ &= \frac{1}{2} \sum_{i,j} \delta(Z_i, Z_j) \sum_{a,b} \delta(A_{ij}^{t-1}, a) \delta(A_{ij}^t, b) R_{ab}^{\theta_i \theta_j} + c_t(A), \end{aligned}$$

where $c_t(A) = \frac{1}{2} \sum_{i,j} \sum_{a,b} \delta(A_{ij}^{t-1}, a) \delta(A_{ij}^t, b) \log Q_{ab}^{\theta_i \theta_j}$ does not depend on the community structure. Simple computations show that

$$\sum_a \delta(A_{ij}^1, a) \rho_a^{\theta_i \theta_j} = A_{ij}^1 (\rho_1^{\theta_i \theta_j} - \rho_0^{\theta_i \theta_j}) + \rho_0^{\theta_i \theta_j}$$

and

$$\begin{aligned} \sum_{a,b} \delta(A_{ij}^{t-1}, a) \delta(A_{ij}^t, b) R_{ab}^{\theta_i \theta_j} &= R_{00}^{\theta_i \theta_j} + A_{ij}^{t-1} (R_{10}^{\theta_i \theta_j} - R_{00}^{\theta_i \theta_j}) + A_{ij}^t (R_{01}^{\theta_i \theta_j} - R_{00}^{\theta_i \theta_j}) \\ &\quad + A_{ij}^{t-1} A_{ij}^t (R_{11}^{\theta_i \theta_j} - R_{01}^{\theta_i \theta_j} - R_{10}^{\theta_i \theta_j} + R_{00}^{\theta_i \theta_j}) \\ &= R_{00}^{\theta_i \theta_j} + A_{ij}^{t-1} \ell_{10}^{\theta_i \theta_j} + A_{ij}^t \ell_{01}^{\theta_i \theta_j} + A_{ij}^{t-1} A_{ij}^t (\ell_{11}^{\theta_i \theta_j} - \ell_{01}^{\theta_i \theta_j} - \ell_{10}^{\theta_i \theta_j}). \end{aligned}$$

By collecting the above observations, we now find that $\log \mathbb{P}(A | Z, \theta)$ equals

$$c(A) + \frac{1}{2} \sum_{i,j} \delta(Z_i, Z_j) \left\{ A_{ij}^1 (\rho_1^{\theta_i \theta_j} - \rho_0^{\theta_i \theta_j}) + \rho_0^{\theta_i \theta_j} + (A_{ij}^1 - A_{ij}^T) \ell_{10}^{\theta_i \theta_j} \right\} \\ + \frac{1}{2} \sum_{i,j} \delta(Z_i, Z_j) \sum_{t=2}^T \left\{ (\ell_{01}^{\theta_i \theta_j} + \ell_{10}^{\theta_i \theta_j}) (A_{ij}^t - A_{ij}^{t-1} A_{ij}^t) + \ell_{11}^{\theta_i \theta_j} A_{ij}^{t-1} A_{ij}^t - \log \frac{Q_{00}^{\theta_i \theta_j}}{P_{00}^{\theta_i \theta_j}} \right\},$$

where $c(A) = \sum_t c_t(A)$ does not depend on Z . Hence the claim follows. \square

A.2 Approximation of the MLE

Recall the structural assumptions (3)–(4) about the degree correction parameters. Because $P_{01}, Q_{01} = o(1)$, a first-order Taylor expansion yields

$$\log \frac{1 - \theta_i \theta_j Q_{01}}{1 - \theta_i \theta_j P_{01}} = \theta_i \theta_j (P_{01} - Q_{01}) + o(P_{01}^2 + Q_{01}^2) = \frac{\bar{d}_i \bar{d}_j}{2\bar{m}} + o(P_{01}^2 + Q_{01}^2),$$

as well as $\ell_{01}^{\theta_i \theta_j} \approx \log \frac{P_{01}}{Q_{01}}$, $\ell_{10}^{\theta_i \theta_j} \approx \log \frac{1-P_{11}}{1-Q_{11}}$ and $\ell_{11}^{\theta_i \theta_j} \approx \log \frac{P_{11}}{Q_{11}}$. Using these approximations in the MLE expression leads to the maximisation of

$$\sum_{t=2}^T \sum_{i,j: z_i=z_j} (\tilde{a}_{ij}^t - \theta_i \theta_j (P_{01} - Q_{01})).$$

where $\tilde{a}_{ij}^t = \alpha (A_{\text{new}}^t)_{ij} + \beta (A_{\text{pers}}^t)_{ij}$. Since μ and ν are stationary distributions,

$$\mathbb{E} (A_{\text{new}}^t)_{ij} = \begin{cases} \theta_i \theta_j \mu_1 (1 - P_{11}) & \text{if } Z_i = Z_j \\ \theta_i \theta_j \nu_1 (1 - Q_{11}) & \text{otherwise,} \end{cases} \\ \mathbb{E} (A_{\text{pers}}^t)_{ij} = \begin{cases} \theta_i \theta_j \mu_1 P_{11} & \text{if } Z_i = Z_j \\ \theta_i \theta_j \nu_1 Q_{11} & \text{otherwise.} \end{cases}$$

Therefore, using $W_{ij} = \sum_{t=2}^T \tilde{a}_{ij}^t$ we have

$$\mathbb{E} W_{ij} = \begin{cases} (T-1) \theta_i \theta_j \mu_1 (\alpha(1 - P_{11}) + \beta P_{11}) & \text{if } Z_i = Z_j \\ (T-1) \theta_i \theta_j \nu_1 (\alpha(1 - Q_{11}) + \beta Q_{11}) & \text{otherwise.} \end{cases}$$

Since the community labeling are sampled uniformly at random and using the normalization for the θ_i 's, we have

$$\bar{d}_i = (T-1) \theta_i N \frac{\mu_1 (\alpha(1 - P_{11}) + \beta P_{11}) + (K-1) \nu_1 (\alpha(1 - Q_{11}) + \beta Q_{11})}{K}$$

together with $\bar{m} = \frac{N^2}{2} \frac{\mu_1 (\alpha(1 - P_{11}) + \beta P_{11}) + (K-1) \nu_1 (\alpha(1 - Q_{11}) + \beta Q_{11})}{K}$. \square

A.3 Modularity and normalized spectral clustering

The regularized modularity of a partition $Z \in [K]^N$ of the graph A is defined as

$$\mathcal{M}(A, Z, \gamma) = \sum_{i,j} \delta(Z_i, Z_j) \left(A_{ij} - \gamma \frac{d_i d_j}{2m} \right)$$

where $d = A1_n$ and γ is a resolution parameter. This can be rewritten as

$$\mathcal{M}(A, Z, \gamma) = \text{Tr } \tilde{Z}^T \left(A - \gamma \frac{dd^T}{2m} \right) \tilde{Z}$$

where $\tilde{Z} \in \{0, 1\}^{N \times K}$ is the membership matrix associated to the vector Z , that is $\tilde{Z}_{ik} = 1$ for $k = Z_i$, and $\tilde{Z}_{ik} = 0$ otherwise. As maximizing the modularity over $Z \in \mathcal{Z}_{N,K}$ is in general NP-complete [4], it is convenient to perform a continuous relaxation. Following [17], we transform the problem into

$$\hat{X} = \arg \max_{\substack{X \in \mathbb{R}^{N \times K} \\ X^T D X = I_K}} \text{Tr } X^T \left(A - \gamma \frac{dd^T}{2m} \right) X. \quad (8)$$

The predicted membership matrix \hat{Z} is then recovered by performing an approximated solution to the following k -means problem (see [12])

$$\left(\hat{Z}, \hat{Y} \right) = \arg \min_{Z \in \mathcal{Z}_{N,K}, Y \in \mathbb{R}^{K \times K}} \left\| ZY - \hat{X} \right\|_F. \quad (9)$$

The Lagrangian associated to the optimization problem (8) is

$$\text{Tr } X^T \left(A - \gamma \frac{dd^T}{2m} \right) X - \text{Tr } (\Lambda^T (X^T D X - I_K))$$

where $\Lambda \in \mathbb{R}^{K \times K}$ is a symmetric matrix of Lagrangian multipliers. Up to a change of basis, we can assume that Λ is diagonal. The solution of (8) verifies

$$\left(A - \gamma \frac{dd^T}{2m} \right) X = D X \Lambda \quad \text{and} \quad X^T D X = I_K,$$

which is a generalized eigenvalue problem: the columns of X are the generalized eigenvectors, and the diagonal elements of Λ are the eigenvalues. In particular, since the constant vector 1_n verifies $(A - \gamma \frac{dd^T}{2m})1_n = (1 - \gamma)D1_n$, we conclude that the eigenvalues should be larger than $1 - \gamma$ for the partition to be meaningful.

Multiplying the first equation by 1_n^T leads to $(1 - \gamma)d^T X = d^T X \Lambda$, and therefore $d^T X = 0$ (using the previous remark on Λ). The system then simplifies in

$$A X = D X \Lambda \quad \text{and} \quad X^T D X = I_K.$$

Defining a re-scaled vector $U = D^{-1/2} X$ shows that U verifies $D^{-1/2} A D^{-1/2} U = U \Lambda$ and $U^T U = I_K$. Thus, the columns of U are eigenvectors of $D^{-1/2} A D^{-1/2}$ associated to the K largest eigenvalue (or equivalently, the eigenvectors of $\mathcal{L} = I_N - D^{-1/2} A D^{-1/2}$ associated to the K smallest eigenvalues).

References

1. Avrachenkov, K., Drevetov, M., Leskelä, L.: Estimation of static community memberships from temporal network data. arXiv preprint arXiv:2008.04790 (2020)
2. Bhattacharyya, S., Chatterjee, S.: General community detection with optimal recovery conditions for multi-relational sparse networks with dependent layers. arXiv preprint arXiv:2004.03480 (2020)
3. Billingsley, P.: Statistical Methods in Markov Chains. *The Annals of Mathematical Statistics* **32**(1), 12 – 40 (1961)
4. Brandes, U., Delling, D., Gaertler, M., Görke, R., Hoefer, M., Nikoloski, Z., Wagner, D.: On finding graph clusterings with maximum modularity. In: *International Workshop on Graph-Theoretic Concepts in Computer Science*. pp. 121–132. Springer (2007)
5. Chi, Y., Song, X., Zhou, D., Hino, K., Tseng, B.L.: Evolutionary spectral clustering by incorporating temporal smoothness. In: *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. p. 153–162. KDD '07, Association for Computing Machinery, New York, NY, USA (2007)
6. Fortunato, S.: Community detection in graphs. *Physics Reports* **486**(3), 75–174 (2010)
7. Fournet, J., Barrat, A.: Contact patterns among high school students. *PLOS ONE* **9**(9), 1–17 (09 2014)
8. Ghasemian, A., Zhang, P., Clauset, A., Moore, C., Peel, L.: Detectability thresholds and optimal algorithms for community structure in dynamic networks. *Physical Review X* **6**(3), 031005 (2016)
9. Holland, P., Laskey, K.B., Leinhardt, S.: Stochastic blockmodels: First steps. *Social Networks* **5**, 109–137 (1983)
10. Holme, P., Saramäki, J.: Temporal networks. *Physics Reports* **519**(3), 97–125 (2012), *temporal Networks*
11. Karrer, B., Newman, M.E.J.: Stochastic blockmodels and community structure in networks. *Phys. Rev. E* **83**, 016107 (Jan 2011)
12. Kumar, A., Kannan, R.: Clustering with spectral norm and the k-means algorithm. In: *2010 IEEE 51st Annual Symposium on Foundations of Computer Science*. pp. 299–308. IEEE (2010)
13. Liu, F., Choi, D., Xie, L., Roeder, K.: Global spectral clustering in dynamic networks. *Proceedings of the National Academy of Sciences* **115**(5), 927–932 (2018)
14. Mastrandrea, R., Fournet, J., Barrat, A.: Contact patterns in a high school: A comparison between data collected using wearable sensors, contact diaries and friendship surveys. *PLOS ONE* **10**(9), 1–26 (09 2015)
15. Matias, C., Miele, V.: Statistical clustering of temporal networks through a dynamic stochastic block model. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **79**(4), 1119–1141 (2017)
16. Mucha, P.J., Richardson, T., Macon, K., Porter, M.A., Onnela, J.P.: Community structure in time-dependent, multiscale, and multiplex networks. *Science* **328**(5980), 876–878 (2010)
17. Newman, M.E.J.: Spectral methods for community detection and graph partitioning. *Phys. Rev. E* **88**, 042822 (Oct 2013)
18. Newman, M.E.: Equivalence between modularity optimization and maximum likelihood methods for community detection. *Physical Review E* **94**(5), 052315 (2016)
19. Newman, M.E., Girvan, M.: Finding and evaluating community structure in networks. *Physical Review E* **69**(2), 026113 (2004)

20. Pamfil, A.R., Howison, S.D., Lambiotte, R., Porter, M.A.: Relating modularity maximization and stochastic block models in multilayer networks. *SIAM Journal on Mathematics of Data Science* **1**(4), 667–698 (2019)
21. Reichardt, J., Bornholdt, S.: Statistical mechanics of community detection. *Physical Review E* **74**(1), 016110 (2006)
22. Rossetti, G., Cazabet, R.: Community discovery in dynamic networks: A survey. *ACM Comput. Surv.* **51**(2) (Feb 2018)