LOW-RANK ESTIMATION BASED EVOLUTIONARY CLUSTERING FOR COMMUNITY DETECTION IN TEMPORAL NETWORKS

Esraa Al-Sharoa[1], Mahmood Al-khassaweneh[2] and Selin Aviyente[1]

[1] Michigan state university, USA. [2] University of Detroit Mercy, USA

ABSTRACT

Many real-world systems can be represented by networks. One common approach to characterizing the organization of networks is community detection. A lot of work has been conducted in community detection of static networks. However, most real systems are time-dependent and modeled by temporal networks with a structure that evolves across time. In this paper, a low-rank approximation based evolutionary clustering approach is introduced to detect and track the community structure of temporal networks. The proposed approach provides robustness to outliers and results in smoothly evolving cluster assignments through joint low-rank approximation and subspace learning. Moreover, a cost function is introduced to track changes in the community structure across time. The performance of the proposed approach is validated on both simulated and real temporal networks.

***Index Terms*—** Temporal networks, community detection, evolutionary spectral clustering, low-rank approximation.

1. INTRODUCTION

Network science has contributed greatly to the modeling of complex systems where the objects in the system and the interactions between them can be presented by nodes and edges in a graph, respectively. Graph-based community detection is very effective at summarizing networks into a set of communities. The majority of the prior work in community detection is limited to static networks [1] [2]. However, many real systems change over time and can be modeled by temporal networks. Consequently, it is very important to develop algorithms to detect the community structure in temporal networks and track its evolution across time.

Over the past decade, different algorithms have been developed to detect the community structure in temporal networks. In particular, evolutionary clustering approaches that preserve the community structure at the current time point and allow the structure to evolve smoothly across time have been proposed. Some examples include simple extensions of static clustering methods such as preserving cluster quality (PCQ) and preserving cluster membership (PCM) [3]. In both frameworks, a cost function that consists of temporal and snapshot terms is used to guarantee temporal smoothness. However, this cost function requires a *priori* knowledge about the community structure of the underlying network. Statistical model based clustering algorithms such as adaptive forgetting factor for evolutionary clustering and tracking (AFFECT) [4] smooths the proximity between objects over time followed by a static clustering method. However, AFFECT assumes a stochastic block model for the adjacency matrix and has high computational complexity. Tensor based community detection methods [5] use non-negative PARAFAC to detect the community structure in binary temporal networks. However, representing the temporal network by a single tensor may not track the community structure over time. Under modularity based methods, a generalized Louvain for multi-layer modularity maximization (GenLov) approach is introduced in [6]. This approach has some shortcomings as discussed in [7]. Most recently, a multi-objective optimization genetic algorithm (DYNMOGA) that consists of snapshot and temporal costs is proposed in [8]. DYNMOGA has a high computational cost that increases as the number of generations increases.

In this paper, an evolutionary spectral clustering approach is introduced to detect and track the community structure in weighted and binary temporal networks. In particular, a low-rank approximation of the observed adjacency matrix is first obtained to reduce the effect of noise and outliers. Next, the corresponding subspace is extracted with the constraint that the network evolves smoothly over time. The current paper differs from existing work in some key ways. First, the proposed method is effective for both binary and weighted networks unlike some existing methods that are limited to binary networks [5]. Second, the low-rank approximation of the adjacency matrix removes noise and results in a more robust presentation of the network's community structure. Third, a smooth clustering assignment is attained over time by minimizing the distance between the projection matrices that span the subspaces of the consecutive time points. Finally, a cost function that tracks changes in the community structure across time is introduced.

2. BACKGROUND: SPECTRAL CLUSTERING

Let $\mathcal{G} = \{V, E, \mathbf{A}\}$ be an undirected weighted or binary graph with a set of nodes $V = \{v_1, \dots, v_n\}$, and a set of edges, E, that represent the objects in the network and pairwise similarities between them, respectively. The adjacency matrix, $\mathbf{A} \in \mathbb{R}^{n\times n}$, is symmetric with $A_{ij} \in [0, 1]$ for weighted networks and $A_{ij} \in \{0, 1\}$ for binary networks. The degree of a node is defined as $d_i = \sum\limits_{j=1}^{n} A_{ij}$ and the degree matrix, $\mathbf{D}$, is defined as the diagonal matrix with $\{d_1, \dots, d_n\}$ on its diagonal [1]. For a temporal network with T time points, a set of undirected weighted or binary graphs is constructed to represent the network over time as $\{\mathcal{G}^{(t)}\}$, where $t \in \{1, 2, \dots, T\}$. The graph at each time point is represented by the adjacency matrix $\mathbf{A}^{(t)} \in \mathbb{R}^{n\times n}$.

One common community detection method for static networks is spectral clustering, which maps the relationship between the nodes in a graph to a lower dimensional subspace. Spectral clustering solves the optimization problem: $\min_{\mathbf{U}\in\mathbb{R}^{n\times r}} Tr\left(\mathbf{U}^\top \mathbf{\Phi}_\mathbf{A} \mathbf{U}\right)$, $s.t$ $\mathbf{U}^\top\mathbf{U} = \mathbf{I}_r$ [1], where $\top$ is the transpose operator and $\mathbf{\Phi}_\mathbf{A} = \mathbf{I} - \mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}}$ is the normalized Laplacian matrix with $\mathbf{D}$ as the

This work was in part supported by the NSF CCF-1615489 and Schlumberger Foundation, Faculty for the Future.

diagonal degree matrix. The solution to the optimization problem is the matrix $\mathbf{U}$ where its columns are the r eigenvectors that correspond to the smallest r eigenvalues of $\mathbf{\Phi_A}$. The community structure is then determined by applying k-means to the matrix $\mathbf{U}$ [1]. The number of eigenvectors, r, refers to the number of clusters in the network.

3. LOW-RANK ESTIMATION BASED EVOLUTIONARY SPECTRAL CLUSTERING (LR-ESC)

3.1. Problem formulation

Let $\mathbf{A}^{(t)} \in \mathbb{R}^{n\times n}$ be a noisy observed adjacency matrix of a low-rank network $\mathbf{L}^{(t)} \in \mathbb{R}^{n\times n}$. The objective is to attain a smooth community assignment from the noisy observations of the temporal network. In order to achieve this goal, we propose to estimate the low-rank adjacency matrix at each time point and extract the corresponding subspace with the constraint that the network evolves smoothly across time. This objective can be presented with the following optimization problem [1]:

$$\min_{\mathbf{L}\in\mathbb{R}^{n\times n},\mathbf{U}\in\mathbb{R}^{n\times r}} \frac{1}{2}\|\mathbf{A}-\mathbf{L}\|_F^2 + \mu\|\mathbf{L}\|_* + \lambda_1 Tr(\mathbf{U}^\top \mathbf{\Phi_L U}) \quad (1)$$
$$+ \lambda_2 \|\mathbf{U}\mathbf{U}^\top - \mathbf{U}^{(t-1)}\mathbf{U}^{(t-1)^\top}\|_F^2, \quad s.t$$
$$\mathbf{U}^\top\mathbf{U} = \mathbf{I}, \mathbf{L} = \mathbf{L}^\top,\ L_{ij} \geq 0,\ L_{ii} = 0.$$

The first two terms characterize the low-rank approximation problem [9], [10], [11] with additional constraints to impose symmetry and nonnegativity of the estimated low-rank adjacency matrix with zeros on the diagonal. The third term with the constraint $\mathbf{U}^\top\mathbf{U} = \mathbf{I}$, represents the spectral clustering problem which is included to detect the community structure of the estimated low-rank network, $\mathbf{L}$, at time point t. The normalized Laplacian matrix, $\mathbf{\Phi_L}$, is defined in terms of the low-rank approximation as $\mathbf{\Phi_L} = \mathbf{I} - \mathbf{D}^{-0.5}\mathbf{L}\mathbf{D}^{-0.5}$. The last term quantifies the distance between two projection matrices [12], $span(\mathbf{U})$ and $span(\mathbf{U}^{(t-1)})$, which is introduced to minimize the distance between the consecutive networks' subspaces at time points t and $t-1$. This term guarantees that the network structure evolves smoothly over time. Moreover, this distance measure is not affected by the variation of number of clusters across time. The parameters μ, λ_1 and $\lambda_2 > 0$ are regularization parameters and are chosen empirically.

3.2. Problem solution

In order to solve the problem in Eq. (1), alternating direction method of multipliers (ADMM) is used [13]. The convergence of this method is well-studied in literature [13] [14]. Moreover, proximal algorithms [15] [16] are used in combination with ADMM since the nuclear norm is non-differentiable. The proximity operator of a convex function f is defined as $prox_f(\mathbf{Y}) = \text{argmin}_{\mathbf{L}\in\mathbb{R}^{n\times n}} f(\mathbf{L}) + \frac{1}{2}\parallel \mathbf{L}-\mathbf{Y}\parallel_F^2$. Our proposed solution of the problem starts by introducing an auxiliary variable $\mathbf{W}$ to split the problem as follows:

$$\min_{\substack{\mathbf{L}\in\mathbb{R}^{n\times n},\mathbf{W}\in\mathbb{R}^{n\times n}\\ \mathbf{U}\in\mathbb{R}^{n\times r}}} \frac{1}{2}\|\mathbf{A}-\mathbf{L}\|_F^2 + \mu\|\mathbf{L}\|_* + \lambda_1 Tr(\mathbf{U}^\top\mathbf{\Phi_W U})$$
$$+ \lambda_2\|\mathbf{U}\mathbf{U}^\top - \mathbf{U}^{(t-1)}\mathbf{U}^{(t-1)^\top}\|_F^2, s.t\ \mathbf{W} = \mathbf{L}, \mathbf{U}^\top\mathbf{U} = \mathbf{I},$$
$$\mathbf{W} = \mathbf{W}^\top,\ W_{ij} \geq 0,\ W_{ii} = 0, \quad (2)$$

where $\mathbf{\Phi_W} = \mathbf{I} - \mathbf{D}^{-0.5}\mathbf{W}\mathbf{D}^{-0.5}$.

Next, the augmented Lagrange multiplier (ALM) function can be reformulated by adding linear and quadratic terms to relax the first equality constraint as follows:

$$\min_{\substack{\mathbf{L}\in\mathbb{R}^{n\times n},\mathbf{W}\in\mathbb{R}^{n\times n}\\ \mathbf{U}\in\mathbb{R}^{n\times r}}} \frac{1}{2}\|\mathbf{A}-\mathbf{L}\|_F^2 + \mu\|\mathbf{L}\|_* \quad (3)$$
$$+ \lambda_1 Tr(\mathbf{U}^\top(\mathbf{I}-\mathbf{D}^{-0.5}\mathbf{W}\mathbf{D}^{-0.5})\mathbf{U})$$
$$+ \lambda_2\|\mathbf{U}\mathbf{U}^\top - \mathbf{U}^{(t-1)}\mathbf{U}^{(t-1)^\top}\|_F^2 + \langle\mathbf{X}^k, \mathbf{W}-\mathbf{L}\rangle$$
$$+ \frac{\gamma}{2}\|\mathbf{W}-\mathbf{L}\|_F^2,\ s.t\ \mathbf{U}^\top\mathbf{U} = \mathbf{I},\ \mathbf{W} = \mathbf{W}^\top,$$
$$W_{ij} \geq 0,\ W_{ii} = 0,$$

with Lagrange multiplier at the k^{th} iteration defined as,

$$\mathbf{X}^{k+1} = \mathbf{X}^k + \gamma(\mathbf{W}^{k+1} - \mathbf{L}^{k+1}). \quad (4)$$

The optimization problem in Eq. (3) can be solved at each time point, t, by following the steps in Algorithm 1 where the primal variables $\mathbf{L}, \mathbf{W}$ and $\mathbf{U}$ are updated iteratively. First, $\mathbf{L}^{k+1}$ is updated by keeping only the terms with $\mathbf{L}$, then combining the inner product and Frobenius norm terms as:

$$\underset{\mathbf{L}\ \in\mathbb{R}^{n\times n}}{\text{argmin}}\ \mu\|\mathbf{L}\|_* + \frac{1+\gamma}{2}\|\mathbf{L} - \frac{\mathbf{A}+\gamma\mathbf{H}^k}{1+\gamma}\|_F^2, \quad (5)$$

with $\mathbf{H}^k = \mathbf{W}^k + \frac{\mathbf{X}^k}{\gamma}$. The update of $\mathbf{L}^{k+1}$ can be further simplified using the proximal operator as $prox_{\frac{\mu}{\gamma+1}\|\mathbf{L}\|_*}(\frac{\mathbf{A}+\gamma\mathbf{H}^k}{1+\gamma})$ with $\gamma_1 = \frac{1+\gamma}{2}$ and $f(\mathbf{L}) = \|\mathbf{L}\|_*$. Let $\mathbf{A}_n = \frac{\mathbf{A}+\gamma\mathbf{H}^k}{1+\gamma}$ and $\mathbf{A}_n = \mathbf{Q}_{\mathbf{A}_n}\mathbf{\Sigma}_{\mathbf{A}_n}\mathbf{V}_{\mathbf{A}_n}^\top$ be the SVD of the matrix $\mathbf{A}_n$, singular value soft thresholding is then used to update $\mathbf{L}^{k+1}$ as:

$$\mathbf{L}^{k+1} = \mathbf{Q}_{\mathbf{A}_n}\Omega_{\frac{\mu}{\gamma_1}}(\mathbf{\Sigma}_{\mathbf{A}_n})\mathbf{V}_{\mathbf{A}_n}^\top, \quad (6)$$

where Ω_τ is the element-wise thresholding operator defined as $\Omega_\tau(a) = sgn(a)max(|a|-\tau, 0)$.

Next, $\mathbf{W}^{k+1}$ is updated by keeping the terms with $\mathbf{W}$ and combining the inner product and Frobenius norm as:

$$\underset{\mathbf{W}}{\text{argmin}}\, f(\mathbf{W})\ :\ \underset{\mathbf{W}\in\mathbb{R}^{n\times n}}{\text{argmin}}\ -\lambda_1 Tr(\mathbf{U}^{k^\top}\mathbf{D}^{k^{-0.5}}\mathbf{W}\mathbf{D}^{k^{-0.5}}\mathbf{U}^k)$$
$$+\frac{\gamma}{2}\|\mathbf{W} - (\mathbf{L}^{k+1} - \frac{\mathbf{X}^k}{\gamma})\|_F^2\ s.t\ \mathbf{W} = \mathbf{W}^\top, W_{ij} \geq 0, W_{ii} = 0. \quad (7)$$

A closed form solution can be computed by finding the gradient of $f(\mathbf{W})$ as $\gamma(\mathbf{W} - \mathbf{L}^{k+1} + \frac{\mathbf{X}^k}{\gamma}) - \lambda_1\mathbf{D}^{k^{-0.5}}\mathbf{U}^k\mathbf{U}^{k^\top}\mathbf{D}^{k^{-0.5}}$. $\mathbf{W}^{k+1}$ is then calculated as:

$$\mathbf{W}^{k+1} = \mathbf{L}^{k+1} - \frac{\mathbf{X}^k}{\gamma} + \frac{\lambda_1}{\gamma}(\mathbf{D}^k)^{-0.5}\mathbf{U}^k\mathbf{U}^{k^\top}(\mathbf{D}^k)^{-0.5}. \quad (8)$$

Finally, spectral projected gradient (SPG) [17] method is used to update $\mathbf{W}$ to ensure the satisfaction of all constraints with the projection operator, $\mathcal{P}(\mathbf{W})$, defined as:

$$\mathcal{P}(\mathbf{W}) = \begin{cases} W_{ij}, & \text{if } i \neq j \text{ and } W_{ij} \geq 0, \\ 0, & \text{if } i = j \text{ or } W_{ij} < 0, \end{cases}$$

Finally, $\mathbf{U}^{k+1}$ is updated by keeping only the terms with $\mathbf{U}$ as:

$$\underset{\mathbf{U}\ \in\mathbb{R}^{n\times r}}{\text{argmin}}\ \lambda_1 Tr(\mathbf{U}^\top\mathbf{\Phi_W}^{k+1}\mathbf{U}) \quad (9)$$
$$+\lambda_2\|\mathbf{U}\mathbf{U}^\top - \mathbf{U}^{(t-1)}\mathbf{U}^{(t-1)^\top}\|_F^2,\ s.t\ \mathbf{U}^\top\mathbf{U} = \mathbf{I}.$$

[1] For the rest of the paper, the superscript, t, which denotes the matrices at the current time point t will be taken out, e.g. $\mathbf{A}^{(t)}$ will be $\mathbf{A}$ and so on.

Using the cyclic property of the trace operator and expanding the Frobenious norm term, the update of $\mathbf{U}^{k+1}$ simplifies to the modified spectral clustering problem:

$$\min_{\mathbf{U}^{(t)} \in \mathbb{R}^{n \times r}} \lambda_1 Tr(\mathbf{U}^\top \mathbf{\Phi_W}{}_{mod}^{k+1} \mathbf{U}), \quad s.t \;\; \mathbf{U}^\top \mathbf{U} = \mathbf{I}, \tag{10}$$

where $\mathbf{\Phi_W}{}_{mod}^{k+1} = \mathbf{\Phi}_{\mathbf{W}}^{k+1} - \frac{\lambda_2}{\lambda_1} \mathbf{U}^{(t-1)} \mathbf{U}^{(t-1)^\top}$ and $\mathbf{\Phi_W}{}_{mod}^{k+1} \in \mathbb{S}_n^+$ where $\mathbb{S}_n^+$ is the set of positive semidefinite matrices.

Algorithm 1 LR-ESC

Input: $\mathbf{A} \in \mathbb{R}^{n \times n}$, $\mathbf{U}^{(t-1)} \in \mathbb{R}^{n \times r_1}$ μ, λ_1, λ_2, ϵ.
Output: Clustering labels.
1: $k \leftarrow 0$
2: Initialize $\mathbf{L} \leftarrow \mathbf{A}$, $\mathbf{W} \leftarrow \mathbf{L}$.
3: $\gamma \leftarrow 1$.
4: $\mathbf{X}^k \leftarrow \mathbf{W}^k - \mathbf{L}^k$. % Define dual variable
5: $\mathbf{P}_1^k = \frac{1}{2}\|\mathbf{A} - \mathbf{L}^k\|_F^2$, $\mathbf{P}_2^k = \mu \|\mathbf{L}^k\|_*$, $\mathbf{P}_3^k = \lambda_1 Tr(\mathbf{U}^{k^\top} \mathbf{\Phi_W}^k \mathbf{U}^k)$, $\mathbf{P}_4^k = \lambda_2 \|\mathbf{U}\mathbf{U}^\top - \mathbf{U}^{(t-1)}\mathbf{U}^{(t-1)^\top}\|_F^2$. % Define primal objectives
6: **while** $\frac{\|\mathbf{P}_1^{k+1} - \mathbf{P}_1^k\|_F^2}{\|\mathbf{P}_1^k\|_F^2} > \epsilon$ and $\frac{\|\mathbf{P}_2^{k+1} - \mathbf{P}_2^k\|_F^2}{\|\mathbf{P}_2^k\|_F^2} > \epsilon$ and $\frac{\|\mathbf{P}_3^{k+1} - \mathbf{P}_3^k\|_F^2}{\|\mathbf{P}_3^k\|_F^2} > \epsilon$ and $\frac{\|\mathbf{P}_4^{k+1} - \mathbf{P}_4^k\|_F^2}{\|\mathbf{P}_4^k\|_F^2} > \epsilon$ and $\frac{\|\mathbf{X}^{k+1} - \mathbf{X}^k\|_F^2}{\|\mathbf{X}^k\|_F^2} > \epsilon$ **do**
7: Update $\mathbf{L}^{k+1}$ using Eq. (6).
8: Update $\mathbf{W}^{k+1}$ using Eq. (8) and SPG method [17].
9: Update $\mathbf{U}^{k+1}$ by solving Eq. (10).
10: **if** $k = 2$ **then**
11: Calculate the number of clusters using Eq. (11).
12: **end if**
13: Update $\mathbf{X}^{k+1}$ using Eq. (4).
14: Update $\mathbf{P}_1^{k+1}$, $\mathbf{P}_2^{k+1}$, $\mathbf{P}_3^{k+1}$ and $\mathbf{P}_4^{k+1}$ using Step 5.
15: $k \leftarrow k + 1$.
16: **end while**
17: Apply k-means on $\mathbf{U}$ to obtain clustering labels.

3.3. Determining number of clusters

In LR-ESC, the number of clusters at each time point is determined using the asymptotical surprise (AS) metric [18]. After the iterations start at Step 10 in Algorithm 1, $\mathbf{U}^{k+1}$ is calculated using a range of (2-20) eigenvectors followed by k-means clustering. The correct number of clusters at time point t is then determined as:

$$r : \underset{r}{\operatorname{argmax}} \;\; AS(\mathbf{A}, Clustering\ labels), \tag{11}$$

where $Clustering\,labels : \{C_1, C_2, \ldots, C_r\}$. The number of clusters is then fixed for time point t for the remaining iterations until the algorithm converges and the number of clusters is updated at each time point.

3.4. Tracking changes in community structure across time

A cost function to track the changes in the community structure is defined through the distance between the projection matrices of the consecutive subspaces:

$$Cost = \|\mathbf{U}\mathbf{U}^\top - \mathbf{U}^{(t-1)}\mathbf{U}^{(t-1)^\top}\|_F^2. \tag{12}$$

The proposed cost function represents the minimum distance between consecutive subspaces and the changes in the cost function reflect the changes in the community structure.

Data Sets	Method	VI	DNOC
$AD = 32$, $z = 2$, $nc\% = 10\%$, $\mu = 0.5$, $\lambda_1 = 0.1$, $\lambda_2/\lambda_1 = 0.1$	LR-ESC	$\mathbf{0 \pm 0}$	$\mathbf{4 \pm 0}$
	AFFECT	0.2 ± 0.047	$\mathbf{4 \pm 0}$
	GenLov	0.119 ± 0.043	$\mathbf{4 \pm 0}$
	DYNMOGA	$7.57 \times 10^{-5} \pm 0.001$	$\mathbf{4 \pm 0}$
$AD = 32$, $z = 5$, $nc\% = 30\%$, $\mu = 0.5$, $\lambda_1 = 0.1$, $\lambda_2/\lambda_1 = 0.1$	LR-ESC	$\mathbf{0 \pm 0}$	$\mathbf{4 \pm 0}$
	AFFECT	0.471 ± 0.100	4.79 ± 1.438
	GenLov	0.259 ± 0.076	$\mathbf{4 \pm 0}$
	DYNMOGA	0.087 ± 0.032	4.613 ± 0.766
$AD = 20$, $z = 3$, $nc\% = 20\%$, $\mu = 0.1$, $\lambda_1 = 0.1$, $\lambda_2/\lambda_1 = 0.1$	LR-ESC	$\mathbf{0.002 \pm 0.004}$	$\mathbf{4 \pm 0}$
	AFFECT	0.408 ± 0.091	4.05 ± 0.22
	GenLov	0.2 ± 0.042	$\mathbf{4 \pm 0}$
	DYNMOGA	0.094 ± 0.035	5.06 ± 1.05

Table 1. Comparison of the performance of the LR-ESC and state-of-the-art algorithms in detecting the community structure of Newman-Girvan benchmark networks in terms of VI and DNOC averaged over time and 100 simulations.

4. RESULTS

4.1. Simulated networks

In order to evaluate the performance of the proposed approach, a set of simulated binary and weighted temporal networks is generated. LR-ESC is compared to the state-of-the-art algorithms including AFFECT [4], DYNMOGA [8] and GenLov [6] using the variation of information (VI) metric [19] and detected number of clusters (DNOC). A low score of VI reflects a better performance.

4.1.1. Newman-Girvan benchmark

The data set described by Girvan and Newman in [5] is used to validate the performance of LR-ESC in detecting and tracking the community structure in binary temporal networks. The temporal network consists of 20 time points with 128 nodes divided into 4 clusters. Each node has a fixed average degree (AD) and shares a number (z) of edges with the nodes from other communities. As AD increases, the network becomes more dense and as z decreases, the clusters become more distinct. In order to introduce temporal changes to the community structure over time, a percentage of nodes ($nc\%$) is moved randomly from their communities and assigned to other communities at each time step. A set of temporal networks with different community structures using different parameters are generated as presented in Table 1. The simulations are repeated 100 times and the average VI for the different methods and networks is reported in Table 1. The selected values of the regularization parameters are also given in Table 1. As it can be seen from the table, LR-ESC algorithm performs better than the other methods in detecting the community structure and the correct number of clusters over time for the different networks. Moreover, LR-ESC is robust and can detect the correct community structure even when AD is low or z and $nc\%$ are high as in data sets 2 and 3.

4.1.2. *Simulated weighted temporal networks*

A set of weighted dynamic networks are generated for 60 time points with 100 nodes. Intra and inter-cluster edges are selected from a truncated Gaussian distribution, $\mathcal{N}(\mu_{intra}, \sigma_{intra}, \mu_{inter}, \sigma_{inter})$, in the range of $[0, 1]$. Sparse noise ($SN\%$) is added to the inter-cluster edges with different percentages. For illustration purposes, the results for two data sets are presented in Fig. 1. Data set 1 is generated with $SN\% = 20\%$ and the edges during time intervals $(1-20,\ 41-60)$ and $(21-40)$ are selected from $\mathcal{N}(0.6, 0.3, 0.2, 0.1)$ and $\mathcal{N}(0.8, 0.2, 0.3, 0.2)$, respectively. The first 20 networks consist of 4 clusters and the nodes in each cluster are: $C_1(1-30)$, $C_2(31-60)$, $C_3(61-80)$, $C_4(81-100)$. The next 20 networks consist of 5 clusters and the nodes in each cluster are: $C_1(1-30)$, $C_2(31-45)$, $C_3(46-60)$, $C_4(61-80)$, $C_5(81-100)$. The last 20 networks consist of 6 clusters and the nodes in each cluster are: $C_1(1-15)$, $C_2(16-30)$, $C_3(31-45)$, $C_4(46-60)$, $C_5(61-80)$, $C_6(81-100)$. The second data set is generated with $SN\% = 10\%$ and the edges during time intervals $(1-20,\ 41-60)$ and $(21-40)$ are selected from $\mathcal{N}(0.5, 0.1, 0.2, 0.1)$ and $\mathcal{N}(0.7, 0.1, 0.2, 0.1)$, respectively. The networks during time intervals $(1-20,\ 41-60)$ consist of 4 clusters and the nodes in each cluster are: $C_1(1-30)$, $C_2(31-60)$, $C_3(61-80)$, $C_4(81-100)$ whereas time interval $(21-40)$ consists of 8 clusters and the nodes in each cluster are: $C_1(1-15)$, $C_2(16-30)$, $C_3(31-45)$, $C_4(46-60)$, $C_5(61-70)$, $C_6(71-80)$, $C_7(81-90)$, $C_8(91-100)$.

The performance of LR-ESC is compared to other methods in Fig. 1. The regularization parameters, (μ, λ_1, λ_2/λ_1) are (1, 0.1, 0.3). The simulations are repeated for 100 times and the cost function, DNOC and VI are averaged over all simulations. From Fig. 1(a) and (b), it can be seen that the cost function reflects the changes in the community structure over time, around time points 20 and 40. Moreover, Fig. 1(c)-(f) show that LR-ESC outperforms other methods in terms of detecting the correct structure and number of clusters over time. The performances of AFFECT and GenLov decline as the size of the clusters becomes smaller and $SN\%$ increases. In particular, GenLov prefers bigger clusters and tends to merge small clusters as can be seen during time interval $20-40$ in Fig. 1(f).

4.2. Real network: Contacts in a workplace temporal network

LR-ESC is evaluated on social network collected by the SocioPatterns collaboration [20] representing contacts during working hours between 92 employees in a workplace over two weeks (June, 24^{th}-July, 5^{th}) [21]. There are 5 departments with 3 scientific departments, DISQ (15 people), DSE (33 people) and DMCT (26 people); and two other departments SFLE (4 people) and SRH (13 people). 10 weighted temporal networks are constructed to represent the face-to-face interactions between the employees for 10 workdays. Each network consists of 11 time points and each time point refers to an hour during the day. LR-ESC is applied to detect and track the structure over time for each day separately. For illustration purposes, the results for workday 10 are presented in Fig. 2. The changes in the cost function presented in Fig. 2(a) reflect the times that correspond to meetings or break times in the morning, lunch time and end of the day similar to the results reported in [21]. The detected community structures for selected change points are shown in Fig. 2(b)-(d) [22]. As it can be seen from Fig. 2(b), big clusters that consist of employees from the three scientific departments are detected during meeting time. A big cluster that consists of employees from different departments is also detected during lunch time as in Fig. 2(c), whereas small clusters that correspond to the different departments are detected near the end of the day as in Fig. 2(d). These results agree with the observed contact activity pattern for the employees in the building as discussed in [21].

5. CONCLUSIONS

In this paper, a low-rank approximation based evolutionary clustering approach is introduced to detect and track the community structure of temporal networks. In addition, a cost function is defined to track changes in the structure of the temporal network. The proposed algorithm detects the community structure in both weighted and binary temporal networks. Furthermore, it provides robustness to noise and outliers and results in temporally smooth cluster assignment. The results show that the proposed approach can detect and track the community structure in temporal networks efficiently and outperforms state-of-the-art algorithms.

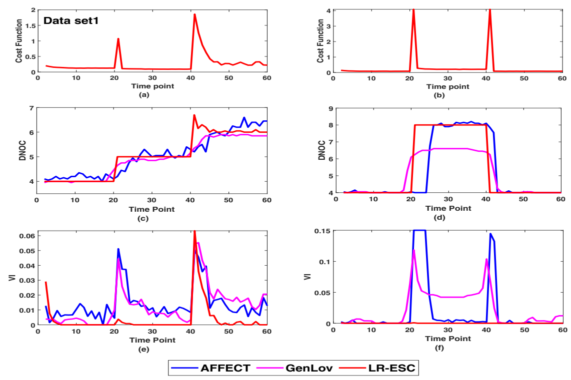

Fig. 1. Comparison between LR-ESC and state-of-the art algorithms for data sets 1 and 2: (a) and (b) The cost function calculated by LR-ESC, (c) and (d) DNOC for different algorithms, (e) and (f) Comparison of VI between LR-ESC, AFFECT and GenLov.

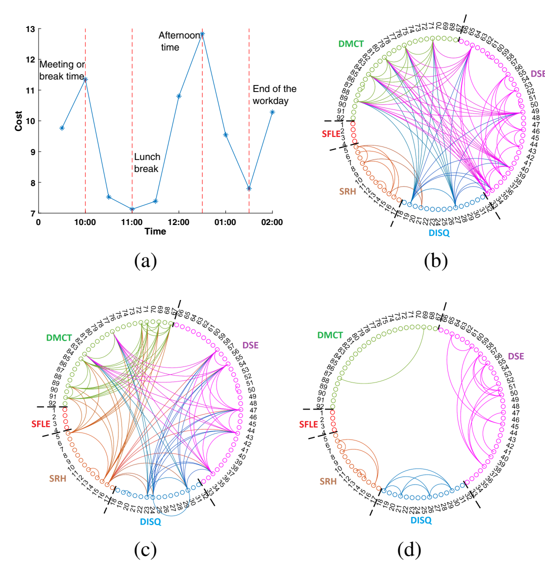

Fig. 2. Results for contacts in a workplace workday 10 with ($\mu = \lambda_1 = \lambda_2/\lambda_1 = 0.1$): (a) Cost function calculated by LR-ESC; (b) Community structure between 10-11 am; (c) Community structure between 12-1 pm; and (d) Community structure between 3-4 pm.

6. REFERENCES

[1] U. Luxburg, "A tutorial on spectral clustering," *Statistics and computing*, vol. 17, no. 4, pp. 395–416, 2007.

[2] S. Borgatti, A. Mehra, D. Brass, and G. Labianca, "Network analysis in the social sciences," *science*, vol. 323, no. 5916, pp. 892–895, 2009.

[3] Y. Chi, X. Song, D. Zhou, K. Hino, and B. Tseng, "Evolutionary spectral clustering by incorporating temporal smoothness," in *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2007, vol. II, pp. 153–162.

[4] Kevin K. Xu, M. Kliger, and A. Hero, "Adaptive evolutionary clustering," *Data Mining and Knowledge Discovery*, vol. 28, no. 2, pp. 304–336, 2014.

[5] M. Girvan and M. Newman, "Community structure in social and biological networks," *Proceedings of the national academy of sciences*, vol. 99, no. 12, pp. 7821–7826, 2002.

[6] P. Mucha, T. Richardson, K. Macon, M. Porter, and J. Onnela, "Community structure in time-dependent, multiscale, and multiplex networks," *science*, vol. 328, no. 5980, pp. 876–878, 2010.

[7] S. Fortunato and D. Hric, "Community detection in networks: A user guide," *Physics Reports*, vol. 659, pp. 1–44, 2016.

[8] F. Folino and C. Pizzuti, "Multiobjective evolutionary community detection for dynamic networks," in *Proceedings of the 12th annual conference on Genetic and evolutionary computation*. ACM, 2010, pp. 535–536.

[9] C. Eckart and G. Young, "The approximation of one matrix by another of lower rank," *Psychometrika*, vol. 1, no. 3, pp. 211–218, 1936.

[10] H. Hotelling, "Analysis of a complex of statistical variables into principal components.," *Journal of educational psychology*, vol. 24, no. 6, pp. 417, 1933.

[11] N Kishore Kumar and J Schneider, "Literature survey on low rank approximation of matrices," *Linear and Multilinear Algebra*, vol. 65, no. 11, pp. 2212–2244, 2017.

[12] G. Golub and C. Van Loan, *Matrix computations*, vol. 3, JHU Press, 2012.

[13] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, et al., "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine learning*, vol. 3, no. 1, pp. 1–122, 2011.

[14] S. Kontogiorgis and R. Meyer, "A variable-penalty alternating directions method for convex optimization," *Mathematical Programming*, vol. 83, no. 1-3, pp. 29–53, 1998.

[15] N. Parikh, S. Boyd, et al., "Proximal algorithms," *Foundations and Trends® in Optimization*, vol. 1, no. 3, pp. 127–239, 2014.

[16] P. Combettes and J. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-point algorithms for inverse problems in science and engineering*, pp. 185–212. Springer, 2011.

[17] E. Birgin, J. Martínez, M. Raydan, et al., "Spectral projected gradient methods: review and perspectives," *J. Stat. Softw*, vol. 60, no. 3, pp. 1–21, 2014.

[18] V. Traag, R. Aldecoa, and J. Delvenne, "Detecting communities using asymptotical surprise," *Physical Review E*, vol. 92, no. 2, pp. 022816, 2015.

[19] M. Meilă, "Comparing clusteringsan information based distance," *Journal of multivariate analysis*, vol. 98, no. 5, pp. 873–895, 2007.

[20] "sociopatterns," `http://www.sociopatterns.org`.

[21] M. Génois, C. Vestergaard, J. Fournet, A. Panisson, I. Bonmarin, and A. Barrat, "Data on face-to-face contacts in an office building suggest a low-cost vaccination strategy based on community linkers," *Network Science*, vol. 3, no. 3, pp. 326–347, 2015.

[22] "Circulargraph toolbox," `https://www.mathworks.com/matlabcentral/fileexchange/48576-circulargraph`.