SPECTRAL MULTI-SCALE COMMUNITY DETECTION IN TEMPORAL NETWORKS WITH AN APPLICATION

Zhana Kuncheva[1,2] and Giovanni Montana[2,3]

[1]Clinical Development and Mathematics, C4X Discovery Ltd, M1 3LD, UK.
[2]Department of Mathematics, Imperial College London, SW7 2AZ, UK.
[3]Department of Biomedical Engineering, King's College London, SE1 7EH, UK.

1. Introduction

Complex Networks
Networks model complex relationships in the social, biological, IT and engineering sciences.

Multi-scale Communities
A community is a group of nodes that are more densely connected to each other than to the rest of the network [8]. Communities can have a hierarchical structure such as a Lancichinetti-Fortunato-Radicchi (LFR) graph or a Sales-Pardo graph.

Temporal Networks
Many systems have relationships that evolve over time. The sequence of networks describing changes occurring over time are known as *temporal networks* [4].

Importance of Multi-scale Community Detection
Extract the relevant information at each level of the network and its communities.
- In monoplex networks - indicative of the in-depth organization of the system.
- In temporal networks – changes over time taking place at one or across all scales of the community structure indicate important turns in the evolution of the system as a whole.

Separate Layer Investigation
Study each time network individually while ignoring the dependence between neighbouring time points.

Layer Aggregation Procedures
Time layers are collapsed into a single network. Afterward traditional algorithms for community detection can be used.

Modularity Maximization (MM)
Generalization of MM to temporal networks [7]

Advantages:
- Multilayer formulation [5] accounts for dependence of consecutive time points.
- Resolution λ controls the community scales [8].

Disadvantages:
- Range of λ values is manually selected.
- Effect of inter-layer weights is unaccounted for.

Contributions
A novel Temporal Multi-Scale Community Detection (TMSCD) [6] method, which
- Extends the notion of spectral graph wavelets [3] to temporal networks.
- Automatically selects the range of relevant scales within which multi-scale community partitions are sought .
- Takes advantage of the multilayer formulation [5] to introduce inter-layer weights relevant to community detection in temporal networks.
- Has competitive performance to the modularity maximization method [7], and is successfully applied to a social network.

2. Temporal Multi-Scale Community Detection

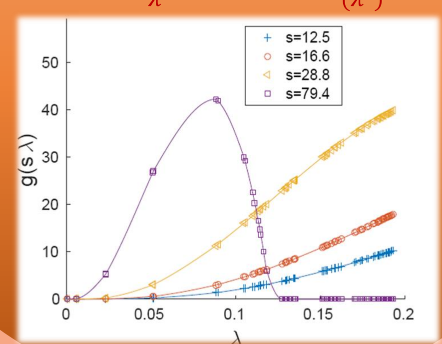
Inter-layer Weights
Let $\mathcal{N}_i^t$ be the number of neighbours for node i in layer t . Then inter-layer weight for node i between layers t and $t+1$ is
$$\omega_i^{t,t+1} := \frac{|\mathcal{N}_i^t \cap \mathcal{N}_i^{t+1}|}{2}.$$

Multilayer Framework
Network $G^t = (V, A^t)$ indicates layer $t, t = 1,2,\ldots,T$, with N nodes each. Multilayer temporal network $\mathcal{A}$ has *ordinal diagonal couplings* -inter-layer weights exist between corresponding nodes in adjacent time layers [5].

Supra-Laplacian Matrix
$$\mathcal{L} = \mathcal{D}^{-\frac{1}{2}}(\mathcal{D}-\mathcal{A})\mathcal{D}^{-\frac{1}{2}}$$
where $\mathcal{D}$ is diagonal node degree matrix [1,2], with eigenvalues $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_{NT}$ and eigenvectors $\chi = [\chi_1, \chi_2, \ldots, \chi_{NT}]$.

New 'Fiedler Vector'
The eigenvector corresponding to eigenvalue 0 of layer t is v_0^t. Eigenvalue λ^* is the smallest non-zero eigenvalue λ_i, whose eigenvector χ_i is not spanned by the set of eigenvectors v_0^t. The eigenvector corresponding to λ^* is perturbation of the separate layers' Fiedler vectors - it is carrier of the coarse description of communities within the layers.

Wavelet Filter g
Function g is the cubic B-spline range of scales s, within which communities are sought, is automatically selected as $s_{min} = \frac{\gamma_1}{\lambda^*}$ and $s_{max} = \frac{\gamma_1}{(\lambda^*)^2}$.



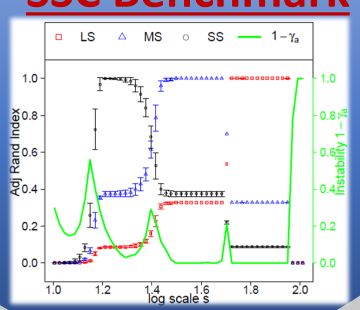
Spectral Graph Wavelets
Let the stretched wavelet filter matrix at scale s be $\mathcal{G}_s = diag(g(s\lambda_1), g(s\lambda_2), \ldots, g(s\lambda_{NT})$. Then the wavelet at scale s around node i in layer t is $\psi_{s,i}^t = \chi\mathcal{G}_s\chi^t\delta_{s,i}$ [3].

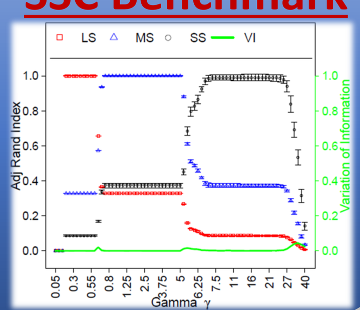
Hierarchical Clustering
At each scale s, use the correlation distance between wavelets $\psi_{s,i}^t$ to group nodes in communities using a connectivity-constrained hierarchical clustering procedure.

TMSCD Method
The TMSCD method is an extension of the multi-scale community detection method for monoplex networks using spectral graph wavelets [10]. The TMSCD follows the steps:
1. Obtain inter-layer weights $\omega_i^{t,t+1}$.
2. Construct multilayer network $\mathcal{A}$ and supra-Laplacian matrix $\mathcal{L}$.
3. Obtain spectral components of $\mathcal{L}$, and discover position of λ^*.
4. Construct spectral graph wavelets $\psi_{s,i}^t$ using wavelet filter g.
5. At each scale s, in the range $s \in \{s_{min}, \ldots, s_{max}\}$, obtain distances between wavelets $\psi_{s,i}^t$ and group nodes in communities using hierarchical clustering with an average maximal gap height cut.
6. At each scale s, obtain instability $\gamma_a(s)$ of the detected communities.

3.Comparative Performance

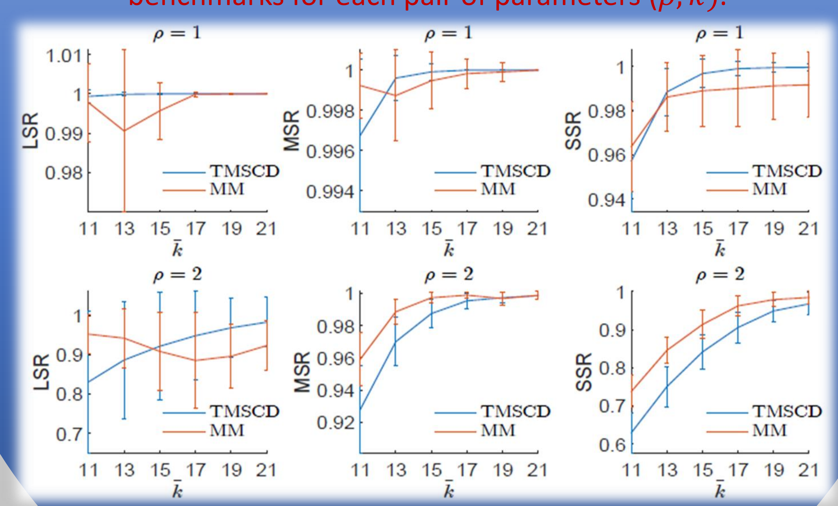
Benchmarks with Three Community Scales
Based on the Sales-Pardo benchmark graph, we construct three types of temporal benchmarks:
1. With Small Scale Community (SSC) changes
2. With Medium Scale Community (MSC) changes
3. With Large Scale Community (LSC) changes

TMSCD on One SSC Benchmark


MM on One SSC Benchmark


TMSCD vs MM SSC Results
The Sales-Pardo graph is based on two parameters: ρ quantifies the scale separation and $\bar{k}$ is the average node degree.

Results below are obtained on 100 simulations of SSC benchmarks for each pair of parameters $(\rho, \bar{k})$.



Benchmarks with One Community Scale
Stochastic Block Models temporal benchmarks with
1. Growing and shrinking communities over time
2. Merging and Splitting communities over time
3. Mixed growing and merging communities

Results
Results below are obtained on 100 simulations of each benchmark type.

	Grow	Merge	Mixed
TMSCD	1.00±0.00	0.87±0.19	0.95±0.10
MM	0.99±0.01	0.68±0.13	0.84±0.14

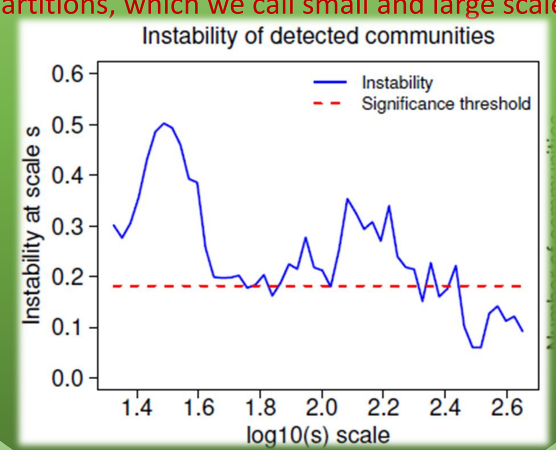
Discussion
The TMSCD method performs competitively well compared to the MM method for both types of benchmarks.

Main advantages to using TMSCD include:
1. Automatic selection of scales at which communities should be sought. This means that no scales with irrelevant community structure will be investigated.
2. The proposed inter-layer weights lead to the best results in terms of balance between uncovering multi-scale communities and the stability of those communities at the relevant scales.
3. The stability procedure used by TMSCD is more sensitive than the modularity maximization one.

4. Real-Life Application

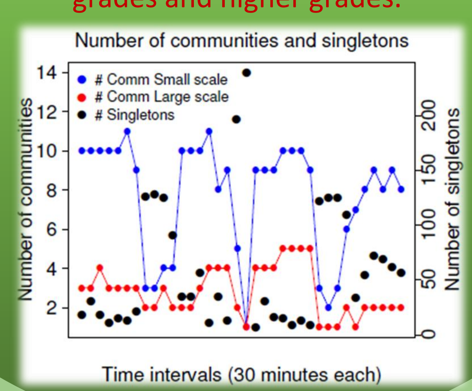
Primary School Data
Data consists of temporal social patterns appearing in a primary school [9]. Data on face-to-face interactions between 242 students and teachers from 10 classes were collected for two days, which we subdivide into 36 intervals of 30 minutes. For each interval, the network of interactions has a link between two individuals if those individuals had at least one contact during the corresponding interval.

Significant Communities
We detect statistically significant community scales by comparing the instability scale of the temporal network to that of a random graph [10]. There appear to be two scales of partitions, which we call small and large scale.



Number of Communities
At the small scale: for pre- and after-lunch class periods students are grouped in around 9-10 communities corresponding to their respective classes. Fewer defined communities appear at lunch due to short face-to-face contacts.

At the large scale: fewer communities are detected as these reflect the split between lower grades and higher grades.



Conclusion
These results validate previous findings on the data set [9], which observe that the majority of communication between students appears in class.

This application on real social patterns demonstrates the strength of TMSCD to detect multi-scale communities in temporal networks: we discovered class-specific communities at the small scale, but we also identified segregation between different grades at a larger scale.

The TMSCD shows great potential for an improved automated community detection procedure, which requires minimum parameter inputs and detects communities at different scales with an increased accuracy.

REFERENCES

1. Bhatia, R.: Matrix Analysis, Graduate Texts in Mathematics, vol. 169. Springer New York, New York, NY (1997)
2. Chung, F.: Spectral Graph Theory. CBMS (1996)
3. Hammond, D.K., Vandergheynst, P., Gribonval, R.: Wavelets on Graphs via Spectral Graph Theory. Appl. Comput. Harmon. Anal. 30(2), 129–150 (mar 2011)
4. Holme, P., Saram¨aki, J.: Temporal Networks. Phys. Rep. 519(3), 97–125 (oct 2012)
5. Kivel¨a, M., Arenas, A., Barthelemy, M., Gleeson, J.P., Moreno, Y., Porter, M.A.: Multilayer Networks. Multilayer Networks 2(3), 203–271 (2014)
6. Kuncheva, Z., Montana, G.: Multi-scale Community Detection in Temporal Networks Using Spectral Graph Wavelets (aug 2017)
7. Mucha, P.J., Richardson, T., Macon, K., Porter, M.A., Onnela, J.P.: Community Structure in Time-Dependent, Multiscale, and Multiplex Networks. Science (80-.). 328 (2010)
8. Newman, M.E.J.: Modularity and Community Structure in Networks. Proc. Natl. Acad. Sci. U. S. A. 103(23), 8577–82 (jun 2006)
9. Stehl´e, J., Voirin, N., Barrat, A., Cattuto, C., Isella, L., Pinton, J.F., Quaggiotto, M., Van den Broeck, W., R´egis, C., Lina, B., Vanhems, P.: High-Resolution Measurements of Face-to-Face Contact Patterns in a Primary School. PLoS One 6(8), e23176 (aug 2011)
10. Tremblay, N., Borgnat, P.: Graph Wavelets for Multiscale Community Mining. IEEE Trans. Signal Process. 62(20), 5227–5239 (oct 2014)

ADDITIONAL INFORMATION

Corresponding author: zhana.kuncheva@c4xdiscovery.com

The author ZK acknowledges partial support by the Engineering and Physical Sciences Research Council (EPSRC), by C4X Discvoery Ltd and by grant no. I 02=19 of Bulgarian NSF.