Blockmodeling temporal networks described by temporal quantities

using clustering with relational constraint

Vladimir Batagelj

IMFM Ljubljana, IAM UP Koper, and NRU HSE Moscow

EUSN 2021 – 5th European Conference on Social Networks

Naples / on Zoom, September 6-10, 2021

Outline

1. Introduction
2. BM and CRC
3. Example
4. Conclusions
5. References

Vladimir Batagelj: vladimir.batagelj@fmf.uni-lj.si
Current version of slides (September 8, 2021 at 01 : 32): slides PDF
https://github.com/bavla/TQ/blob/master/docs

Temporal networks

Temporal networks described by *temporal quantities* (TQs) were introduced in the paper [1]. We get a *temporal network* $\mathcal{N}_\mathcal{T} = (\mathcal{V}, \mathcal{L}, \mathcal{T}, \mathcal{P}, \mathcal{W})$ by attaching the *time* $\mathcal{T}$ to an ordinary network, where $\mathcal{T} = [T_{min}, T_{max})$ is a linearly ordered set of time points $t \in \mathcal{T}$ which are usually integers or reals.

In a temporal network nodes/links activity/presence, nodes properties, and links weights can change through time. These changes are described with TQs. A TQ is described by a sequence

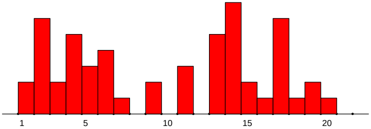
$$\mathbf{a} = [(s_r, f_r, v_r) : r = 1, 2, \ldots, k]$$

where $[s_r, f_r)$ determines a time interval and v_r is the value of the TQ $\mathbf{a}$ on this interval. The set $T_a = \bigcup_r [s_r, f_r)$ is called the *activity set* of $\mathbf{a}$. For $t \notin T_a$ its value is *undefined*, $a(t) = ⌘$. Assuming $a + ⌘ = ⌘ + a = a$ and $a \cdot ⌘ = ⌘ \cdot a = ⌘$ we can extend the addition and multiplication to TQs

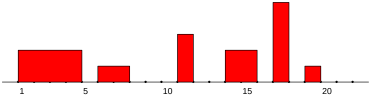
$$(a + b)(t) = a(t) + b(t) \quad \text{and} \quad T_{a+b} = T_a \cup T_b$$
$$(a \cdot b)(t) = a(t) \cdot b(t) \quad \text{and} \quad T_{a \cdot b} = T_a \cap T_b$$

Addition and multiplication of TQs

$a + b$:

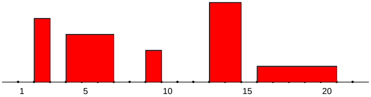


a :



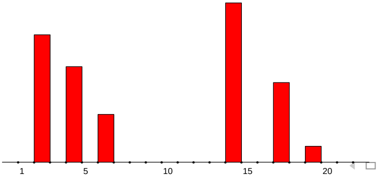
$$a = [(1,5,2),(6,8,1),(11,12,3),(14,16,2),(17,18,5),(19,20,1)]$$

b :



$$b = [(2,3,4),(4,7,3),(9,10,2),(13,15,5),(16,21,1)]$$

$a \cdot b$:



Temporal networks

Let $T_V(v) \subseteq \mathcal{T}$, $T_V \in \mathcal{P}$, be the activity set for a node $v \in \mathcal{V}$ and $T_L(\ell) \subseteq \mathcal{T}$, $T_L \in \mathcal{W}$, the activity set for a link $\ell \in \mathcal{L}$. The following *consistency condition* must be fulfilled for activity sets: If a link $\ell(u, v)$ is active at the time point t then its end-nodes u and v should be active at the time point t :

$$T_L(\ell(u, v)) \subseteq T_V(u) \cap T_V(v).$$

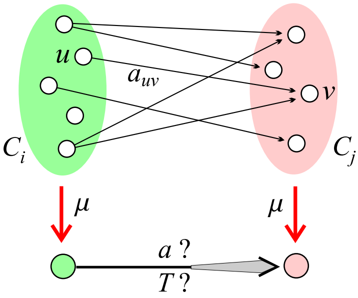
In the following we will need

- *Total* $\quad \text{total}(a) = \sum_i (f_i - s_i) \cdot v_i$
- *Average* $\quad \text{average}(a) = \frac{\text{total}(a)}{|T_a|}$
- *Maximum* $\quad \max(a) = \max_i v_i$

To support the computations with TQs we developed in Python a library TQ, see https://github.com/bavla/TQ.

Traditional (generalized) blockmodeling scheme

A *blockmodel* (BM) [7] consists of structures obtained by identifying all units from the same cluster of the clustering / *partition* $\mathbf{C} = \{C_i\}$, $\pi(v) = i \Leftrightarrow v \in C_i$. For an exact definition of a blockmodel we have to be precise also about which blocks produce an arc in the *reduced graph* and which do not, what is the *weight* of this arc, and in the case of generalized BM, of what *type*. The reduced graph can be represented by relational matrix, called also *image matrix*.



BM of temporal networks

To the traditional BM scheme we add the time dimension – the BM partition π is described for each node v with a temporal quantity $\pi(v,t)$: $\pi(v,t) = i$ means that in time t node v belongs to cluster i. The structure and activity of clusters $C_i(t) = \{v : \pi(v,t) = i\}$ can change through time, but they preserve their identity.

For the BM μ the clusters are mapped into BM nodes $\mu : C_i \to [i]$. To determine the BM we still have to specify how the links from C_i to C_j are represented in the BM – in general, for the model arc $([i],[j])$, we have to specify two TQs: its weight $a_{ij}(t)$ and, in the case of generalized BM, its type $\tau_{ij}(t)$. The weight can be an object of a different type than the weights of the block links in the original temporal network.

For an early attempt see [2, 3].

. . . BM of temporal networks

We assume that in a temporal network $\mathcal{N} = (\mathcal{V}, \mathcal{L}, \mathcal{T}, \mathcal{P}, \mathcal{W})$ the links weight is described by a TQ $w \in \mathcal{W}$. In the following years we intend to develop methods case by case.

- constant partition – nodes stay in the same cluster all the time
  - Sunbelt 2020: indirect approach based on clustering of TQs: $p(v) = \sum_{u \in N(v)} w(v,u)$, hierarchical clustering and leaders [5];
  - EUSN 2021: indirect approach by conversion to the clustering with relation constraint (CRC);
  - local optimization of the criterion function P over Φ
- dynamic partition – nodes can move between clusters through time. The details are still to be elaborated.

TBM as a CRC problem

To use the clustering in the construction of a nodes partition we have to define a similarity measure $s(u, v)$ (or dissimilarity $d(u, v)$) between nodes. An obvious solution is $s(u, v) = f(w(u, v))$, for example

- Total activity $s_1(u, v) = \text{total}(w(u, v))$
- Average activity $s_2(u, v) = \text{average}(w(u, v))$
- Maximal activity $s_3(u, v) = \max(w(u, v))$

We can transform a similarity $s(u, v)$ into a dissimilarity by

$$d(u, v) = \frac{1}{s(u, v)}$$

We transformed the partitioning problem to the clustering with relational constraint problem [4, 360–369].

Basic steps

1. Using clustering with relational constraint determine the partition π
2. Construct the corresponding BM
3. Interpret the BM

The simplest option is to take for $\mathbf{a}_{ij}$ the sum of all total activities of links from C_i to C_j

$$\mathbf{a}_{ij} = \sum_{\ell \in \mathcal{L}(C_i, C_j)} \mathrm{total}(w(\ell))$$

Another option is to consider the activity of links from C_i to C_j as a TQ

$$\mathbf{a}_{ij} = \sum_{\ell \in \mathcal{L}(C_i, C_j)} w(\ell)$$

September 11th Reuters terror news

The Reuters terror news network was obtained from the CRA (Centering Resonance Analysis) networks produced by Steve Corman and Kevin Dooley at Arizona State University. The network is based on all the stories released during 66 consecutive days by the news agency Reuters concerning the September 11 attack on the U.S., beginning at 9:00 AM EST 9/11/01.

The nodes, $n = 13332$, of this network are important words (terms). For a given day, there is an edge between two words iff they appear in the same utterance (for details see the paper [6]). The weight of an edge is its frequency. There are no loops in the network.

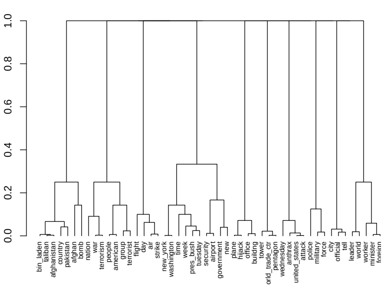
As an example we will analyze the subnetwork of 50 the most active terms available as Terror50.json.

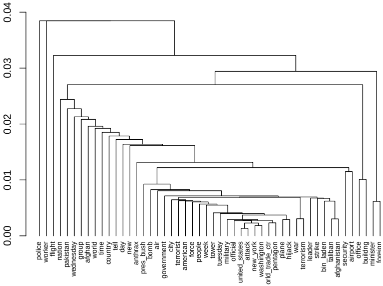
The network Terror50 is undirected – so will be also the BM.

Dendrograms

max/tolerant and min/tolerant

The network file `Terror50tot.net` was created in Python. Hierarchical CRC was done in Pajek. The dendrograms were produced in R.

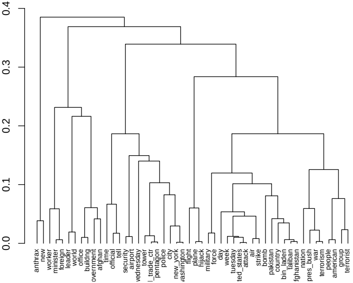


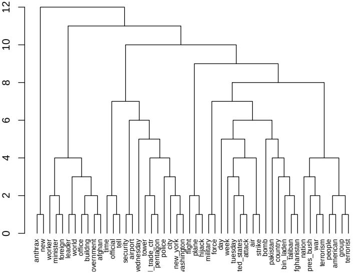


- max/tolerant: many clusters
- min/tolerant: chaining of small (singleton) clusters

Dendrograms

average/tolerant and its rank levels





- average/tolerant: homogeneity of the space of units?
- rank levels: same tree, complexity/nestedness of clusters

Partition

$C_1 = \{$ united_states, attack, taliban, afghanistan, bin_laden, military, country, tuesday, force, day, week, air, strike, pakistan, bomb $\}$

$C_2 = \{$ people, pres_bush, american, war, terrorism, group, terrorist, nation $\}$

$C_3 = \{$ plane, hijack, flight $\}$

$C_4 = \{$ new_york, washington, official, world_trade_ctr, security, city, pentagon, time, tell, airport, tower, wednesday, police $\}$

$C_5 = \{$ government, leader, world, worker, office, minister, afghan, buildng, foreign $\}$

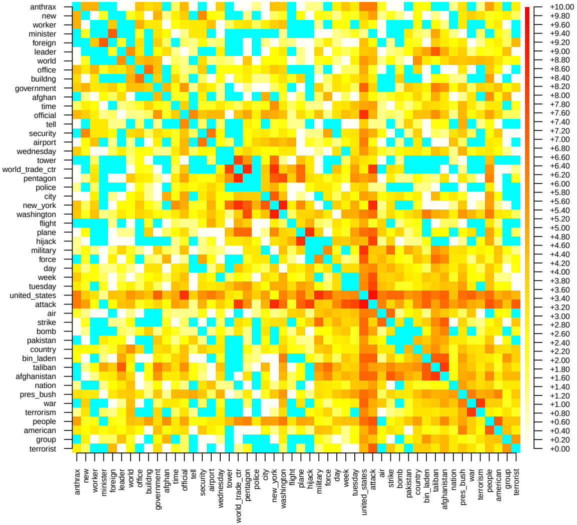
$C_6 = \{$ anthrax, new $\}$

(1) Actors, (2) Reaction, (3) Act, (4) Place, (5) Official, (6) Other

The obtained partition was in Python transformed in a temporal BM network that was further analyzed in R.

Heatmap of activity matrix

average/tolerant order

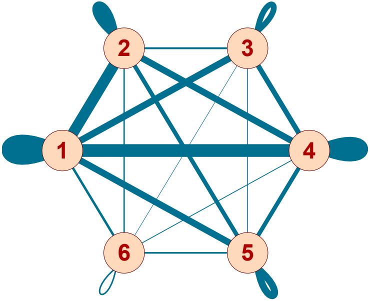


cyan – no link

Blockmodel – activities

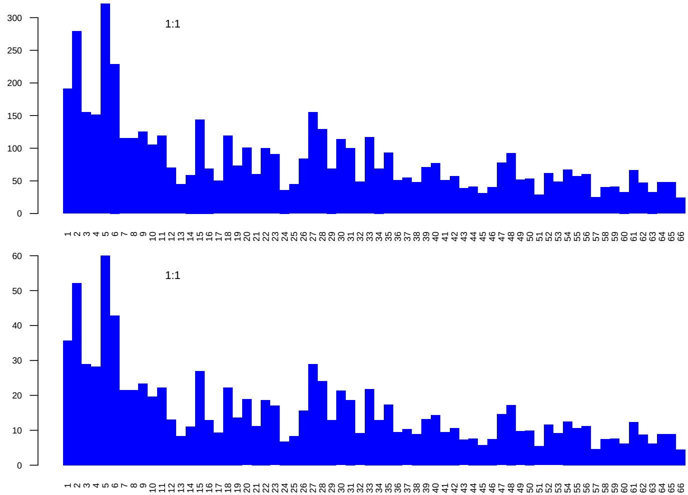
sqrt weights on graph

	1	2	3	4	5	6
1	5511	1398	539	1931	514	69
2	1398	1001	44	503	211	22
3	539	44	382	301	7	1
4	1931	503	301	3043	206	8
5	514	211	7	206	576	32
6	69	22	1	8	32	26



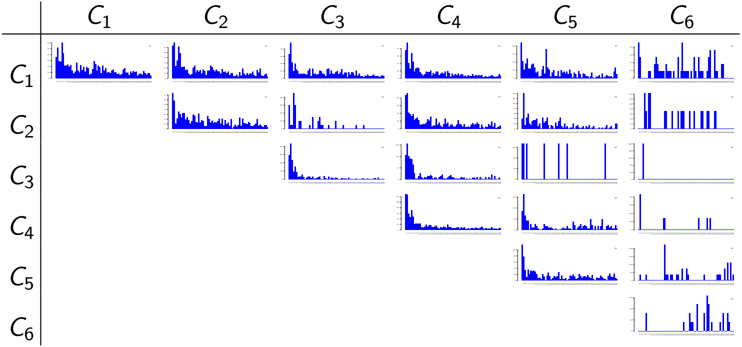
TQs on model links

totals and per-milles



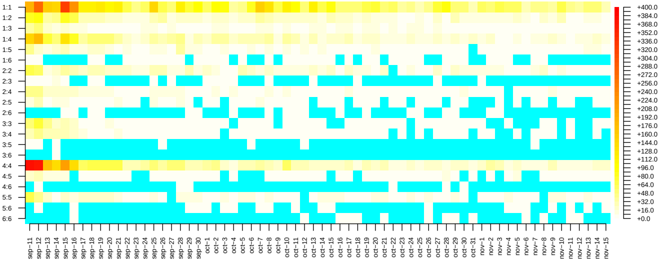
Block model with TQs

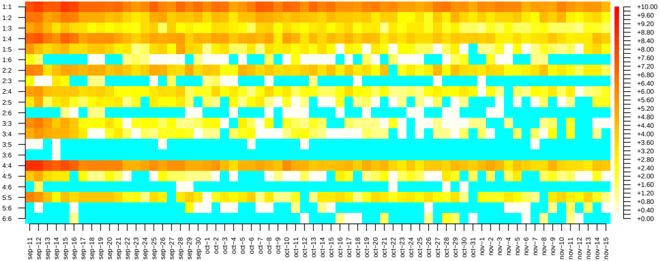
per-milles



Heatmap of BM TQs

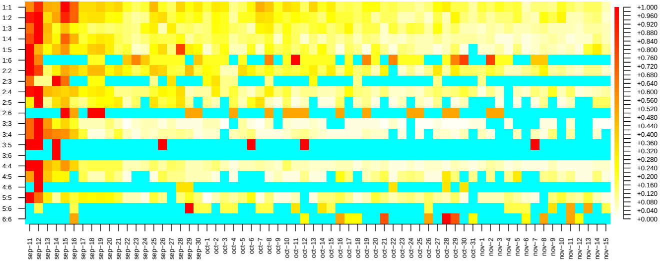
totals and log2 totals



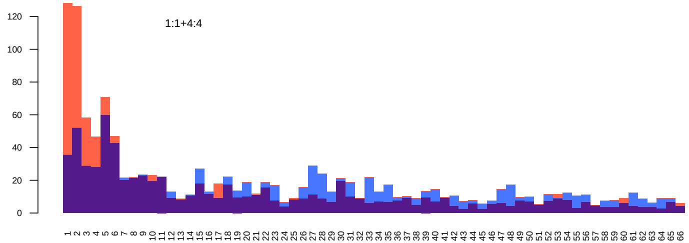


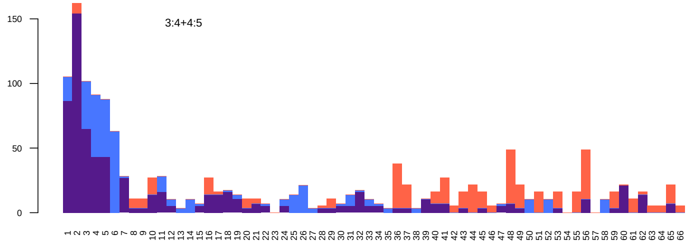
Heatmap of BM TQs

max normalization



Comparing pairs of TQs





Conclusions

1. normalization of model weights TQs when comparing
2. detection of interesting patterns in TQs
3. dynamic temporal networks viewer
4. implementation of the approach in a single programming language (R or Python or Julia)
5. a collection of interesting, well documented temporal networks is needed

Acknowledgments

The computational work reported on in this presentation was performed using Python, R and Pajek. The code and data are available at https://github.com/bavla/TQ/wiki/BMRC .

This work is supported in part by the Slovenian Research Agency (research program P1-0294 and research projects J1-9187 and J5-2557), and prepared within the framework of the HSE University Basic Research Program.

References I

Batagelj, V., Praprotnik, S. (2016) An algebraic approach to temporal network analysis based on temporal quantities. Social Network Analysis and Mining, 6(1), pages 1-22.

Batagelj, V., Ferligoj, A. (2000) Clustering Relational Data. In Gaul, W., Opitz, O., Schader, M. (Eds.) Data Analysis / Scientific Modeling and Practical Application. Springer, pages 3-15

Batagelj, V., Ferligoj, A., Doreian, P. (2007) Indirect Blockmodeling of 3-Way Networks. In Brito, P., Bertrand, P., Cucumel, G., de Carvalho, F. (eds.) Selected Contributions in Data Analysis and Classification. Springer, pages 151–159.

Batagelj, V., Doreian, P., Ferligoj, A., Kejžar, N. (2014) Understanding Large Temporal Networks and Spatial Networks: Exploration, Pattern Searching, Visualization and Network Evolution. Wiley.

References II

- Batagelj, V. (2020) TQ / clusTQ – clustering temporal quantities. https://github.com/bavla/TQ

- Corman, S.R., Kuhn, T., McPhee, R.D., Dooley, K.J. (2002) Studying complex discursive systems: Centering resonance analysis of communication. Human Communication Research, 28(2), 157-206.

- Doreian, P., Batagelj, V., Ferligoj, A. (2005) Generalized Blockmodeling. Cambridge University Press.