Longitudinal approach to the analysis of temporal networks based on temporal quantities

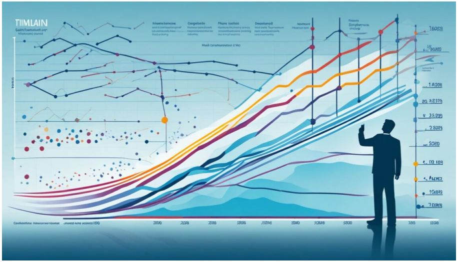
Vladimir Batagelj

IMFM Ljubljana, IAM UP Koper

Workshop on Innovation in Dynamic Network Modelling

Lugano, September 11-13, 2024

Outline



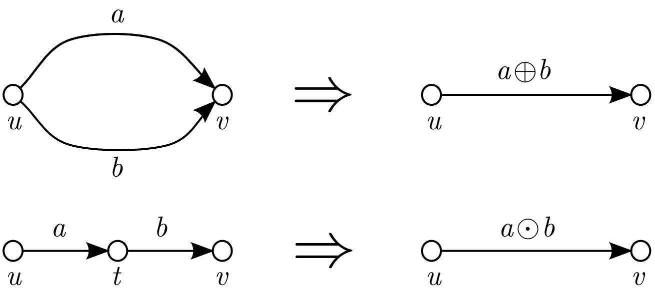
Epidemiology

Vladimir Batagelj: vladimir.batagelj@fmf.uni-lj.si

Current version of slides (September 12, 2024 at 00 : 55): slides PDF

https://github.com/bavla/TQ

Computing with link weights in networks



A semiring is a "natural" algebraic structure to formalize computations with link weights in networks. They allow us to extend the weights from links to nodes, walks (paths) and sets of walks.

Semiring

Let $\mathbb{A}$ be a set and a, b, c elements from $\mathbb{A}$. A *semiring* [11, 4, 15, 3, 1] is an algebraic structure $(\mathbb{A}, \oplus, \odot, 0, 1)$ with two binary operations (addition $\oplus$ and multiplication $\odot$) where:

$(\mathbb{A}, \oplus, 0)$ is an *abelian monoid* with the neutral element 0 (zero):
$a \oplus b = b \oplus a$ – commutativity
$(a \oplus b) \oplus c = a \oplus (b \oplus c)$ – associativity
$a \oplus 0 = a$ – existence of zero

$(\mathbb{A}, \odot, 1)$ is a *monoid* with the neutral element 1 (unit):
$(a \odot b) \odot c = a \odot (b \odot c)$ – associativity
$a \odot 1 = 1 \odot a = a$ – existence of a unit

Multiplication $\odot$ *distributes* over addition $\oplus$:
$a \odot (b \oplus c) = a \odot b \oplus a \odot c \qquad (b \oplus c) \odot a = b \odot a \oplus c \odot a$

In formulas we assume precedence of the multiplication over the addition.

Some examples

Some examples of semirings used in network data analysis:

1. Combinatorial: $(\mathbb{N}, +, \cdot, 0, 1)$ or $(\mathbb{R}_0^+, +, \cdot, 0, 1)$
2. Reachability: $(\{0, 1\}, \vee, \wedge, 0, 1)$
3. Shortest paths [16]: $(\mathbb{R}_0^+, \min, +, \infty, 0)$
4. MaxMin (capacity): $(\mathbb{R}_0^+, \max, \min, 0, \infty)$
5. Pathfinder [21, 20, 22]: $(\overline{\mathbb{R}_0^+}, \min, \boxed{r}, \infty, 0)$
   where $a \boxed{r} b = \sqrt[r]{a^r + b^r}$ (Minkowski)
6. Interval [18, 2]: $[a, A], [b, B] \subset \mathbb{R}_0^+$
   $[a, A] \oplus [b, B] = [a + b, A + B]$ and
   $[a, A] \odot [b, B] = [a \cdot b, A \cdot B]$

Histograms

7. Let the set of bins $\mathbf{B} = \{B_1, B_2, \ldots, B_k\}$ be a partition of the set B such that $(\mathbf{B}, \circ)$ is a semigroup. A *histogram* $h : \mathbf{B} \to \mathbb{N}$
$h_i = h(B_i) = |\{X : v(X) \in B_i\}|$

$$h \oplus g = h + g \quad (h \oplus g)(i) = h(i) + g(i)$$

$h \odot g = h * g$ convolution [13, 10]

$$(h * g)(i) = \sum_{p \circ q = i} h(p) \cdot g(q)$$

Temporal quantities

8 A temporal quantity (TQ) a is a function $a : \mathcal{T} \to A \cup \{⌘\}$ where ⌘ denotes the value *undefined*. $(A, +, \cdot, 0, 1)$ is a semiring. The *activity time set* T_a of a consists of instants $t \in T_a$ in which a is defined $T_a = \{t \in \mathcal{T} : a(t) \in A\}$.

We can extend both operations to the set $A_⌘ = A \cup \{⌘\}$ by requiring that for all $a \in A_⌘$ it holds $a + ⌘ = ⌘ + a = a$ and $a \cdot ⌘ = ⌘ \cdot a = ⌘$.

The structure $(A_⌘, +, \cdot, ⌘, 1)$ is also a semiring.

Let $A_⌘(\mathcal{T})$ denote the set of all TQs over $A_⌘$ in time $\mathcal{T}$. To extend the operations to networks and their matrices we first define the *sum* (parallel links) $a + b$ as

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

The *product* (sequential links) $a \cdot b$ is defined as

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

Temporal quantities

Let us define TQs $\mathbf{0}$ and $\mathbf{1}$ with requirements $\mathbf{0}(t) = ⌘$ and $\mathbf{1}(t) = 1$ for all $t \in \mathcal{T}$. Again, the structure $(A_{⌘}(\mathcal{T}), +, \cdot, \mathbf{0}, \mathbf{1})$ is a semiring.

To produce a software support for computation with TQs we limit it to TQs that can be described as a sequence of disjoint time intervals with a constant value

$$a = [(s_i, f_i, v_i)]_{i \in 1..k}$$

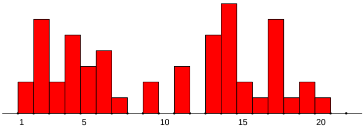
where s_i is the starting time and f_i the finishing time of the i-th time interval $[s_i, f_i)$, $s_i < f_i$ and $f_i \leq s_{i+1}$, and v_i is the value of a on this interval (over combinatorial semiring). Outside the intervals the value of TQ a is undedined, ⌘.

See also [19].

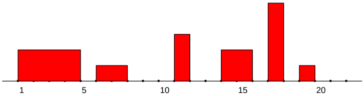
Another approach based on semirings [17, 14]

Sum and product of temporal quantities

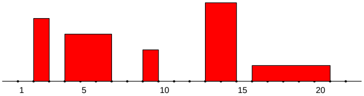
$a + b$:



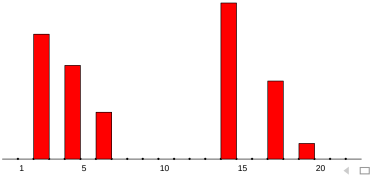
a :



b :



$a \cdot b$:



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

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

Temporal quantities

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

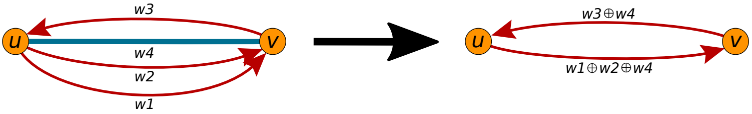
In the following, we will need

1. *Total*: $\text{total}(a) = \sum_i (f_i - s_i) \cdot v_i$
2. *Average*: $\text{average}(a) = \dfrac{\text{total}(a)}{|T_a|}$ where $|T_a| = \sum_i (f_i - s_i)$
3. *Maximum*: $\max(a) = \max_i v_i$

To support the computations with TQs we developed in Python the libraries TQ and Nets, see `https://github.com/bavla/TQ` .

Simplification of weighted networks

For a subset $S \subseteq L$ of links, we define its *total* weight as $\Sigma(S) = \bigoplus_{\ell \in S} w(\ell)$.



A *simplification* of a weighted network $\mathcal{N} = (V, L, w)$ over an abelian monoid $(\mathbb{A}, \oplus, 0)$ is its transformation into a *simple* directed weighted network $\widehat{\mathcal{N}} = (V, A, \widehat{w})$ such that for
$L(u,v) = \{\ell \in L : \text{init}(\ell) = u \land \text{term}(\ell) = v\}$
it holds $L(u,v) \neq \emptyset \Rightarrow (u,v) \in A \land \widehat{w}(u,v) = \Sigma(L(u,v))$.

For undirected weighted networks, a special undirected simplification can be defined similarly.

Activity of nodes

$$\text{inStar}(v) = \{\ell \in L : \text{term}(\ell) = v\}, \quad \text{outStar}(v) = \{\ell \in L : \text{init}(\ell) = v\}$$

Using the link weight w of the network $\mathcal{N}$ we can compute some interesting node properties such as *in-sum* (weighted indegree):

$$inS(\mathcal{N}, v) = \Sigma(\text{inStar}(v))$$

and *out-sum* (weighted outdegree):

$$outS(\mathcal{N}, v) = \Sigma(\text{outStar}(v))$$

of a node v. They are describing the *activity* of the node v.

Activity of sets

For a simple weighted network $\mathcal{N} = (V, A, w)$ and a subset of nodes $C \subseteq V$ we can define the *boundaries* of C:

$$\mathrm{inBd}(C) = \{\ell \in A : \mathrm{init}(\ell) \notin C \wedge \mathrm{term}(\ell) \in C\}$$
$$\mathrm{outBd}(C) = \{\ell \in A : \mathrm{init}(\ell) \in C \wedge \mathrm{term}(\ell) \notin C\}$$
$$\mathrm{Bd}(C) = \mathrm{inBd}(C) \cup \mathrm{outBd}(C)$$

the *interior* of C:

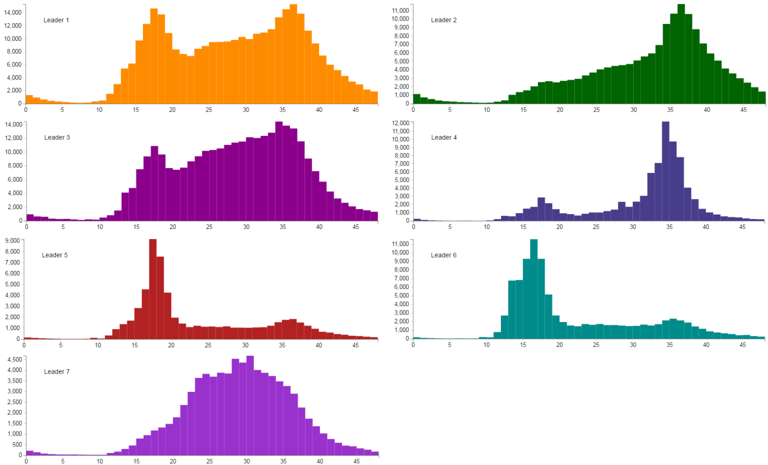
$$\mathrm{Int}(C) = \{\ell \in A : \mathrm{init}(\ell) \in C \wedge \mathrm{term}(\ell) \in C\}$$

and the corresponding *activities*:

$$\mathrm{inPut}(C) = \Sigma(\mathrm{inBd}(C)) \quad \text{and} \quad \mathrm{outPut}(C) = \Sigma(\mathrm{outBd}(C))$$
$$\mathrm{Act}(C) = \Sigma(\mathrm{Int}(C))$$

Bike-sharing data analysis – Citi bike

Clustering of flows / 7 clusters [7]



NY Citi Bike one-year data from October 2015 to September 2016: 13266296 trips, 678 stations.
flow$(u, v; k) = \#$ of trips starting in a node u in the k-th half hour and finishing in a node v

Application – temporal bibliographic networks

Let the binary *affiliation* matrix $\mathbf{A} = [a_{ep}]$ describe a two-mode network on the set of events E and the set of of participants P:

$$a_{ep} = \begin{cases} 1 & p \text{ participated at the event } e \\ 0 & \text{otherwise} \end{cases}$$

The function $d : E \to \mathcal{T}$ assigns to each event e the date $d(e)$ when it happened. Assume $\mathcal{T} = [\mathit{first}, \mathit{last}] \subset \mathbb{N}$. Using these data we can construct two temporal affiliation matrices [6]:

- **instantaneous Ai** $= [ai_{ep}]$, where

$$ai_{ep} = \begin{cases} [(d(e), d(e)+1, 1)] & a_{ep} = 1 \\ [\,] & \text{otherwise} \end{cases}$$

- **cumulative Ac** $= [ac_{ep}]$, where

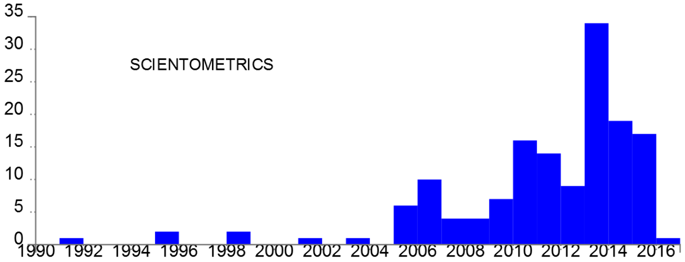
$$ac_{ep} = \begin{cases} [(d(e), \mathit{last}+1, 1)] & a_{ep} = 1 \\ [\,] & \text{otherwise} \end{cases}$$

Networks are from the collection of bibliographic networks on peer review from WoS till 2017.

The derived network describing citations between journals is obtained as

$$\mathbf{JCJ} = \mathbf{WJi}^T \cdot \mathbf{CiI} \cdot \mathbf{WJc}$$

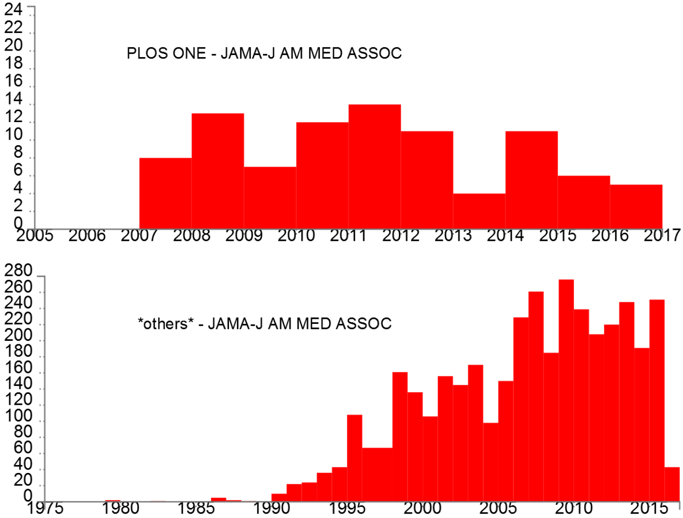
Note that the third network in the product is cumulative.



Selfcitations

Application – temporal bibliographic networks

Temporal citations between journals



Citations between journals

Conclusions

Work in progress.

Most of the approaches to temporal network analysis are cross-sectional – based on time slices.
Another approach (used in most talks here) is based on event sequences. In applications, an event has additional properties besides the pair of nodes and starting time such as duration, type, location, etc. [9].
The longitudinal approach is an alternative to be considered.

Collections of interesting, understandable, and well-documented datasets (bikes, world trade, bibliographic data, political events (KEDS, WEIS), etc.). Sources of problems and ideas.

New analytical procedures: temporal clustering of nodes, temporal 1-neighbors, temporal cores, etc [8].

Acknowledgments

The computational work reported in this presentation was performed in Python using libraries TQ and Nets. The code and data are available at https://github.com/bavla/TQ.

This work is supported in part by the Slovenian Research Agency (research program P1-0294 and research projects J5-2557, J1-2481, and J5-4596), and prepared within the framework of the COST action CA21163 (HiTEc).

References I

- Abdali SK, Saunders BD (1985) Transitive closure and related semiring properties via eliminants. Theor Com-put Sci 40:257—274
- Alves HFC (2020) Interval-Weighted Networks: Community Detection and Centrality Measures. PhD thesis Faculdade de Ciências da Universidade do Porto, Universidade de Aveiro e Universidade do Minho Matemática Aplicada.
- Baras JS, Theodorakopoulos G (2010) Path problems in networks. Morgan & Claypool, Berkeley
- Batagelj V (1994) Semirings for social networks analysis. J Math Soc 19(1):53-–68
- Batagelj V, Praprotnik S (2016) An algebraic approach to temporal network analysis based on temporal quantities. Social Network Analysis and Mining 6(1): 1–22
- Batagelj V, Maltseva D(2020) Temporal bibliographic networks. Journal of Informetrics, 14 (1), 101006.

References II

- Batagelj, V (2017) Symbolic network analysis of bike-sharing data – Citi Bike. 6th Symbolic Data Analysis Workshop, Ljubljana, 12–14. June 2017. slides
- Batagelj, V.: Clustering and Blockmodeling Temporal Networks - Two Indirect Approaches. In: Paula Brito, José G. Dias, Berthold Lausen, Angela Montanari, Rebecca Nugent (eds.) Classification and Data Science in the Digital Age. Proceedings of IFCS 2022, Porto. Springer 2023, paper, book.
- Batagelj, V, Ferligoj, A, Oštro Center (2024) Event sequence analysis - Network analysis of sequences of events. 1343+1344. sredin seminar, Ljubljana, March 6, 2024. slides
- Brunotte H (2013) Discrete Convolution Rings. Romanian Journal of Mathematics and Computer Science, Volume 3, Issue 2, p.155–159
- Carré B (1979) Graphs and networks. Clarendon, Oxford
- Cerinšek M, Batagelj V (2017) Semirings and Matrix Analysis of Networks. In: Alhajj R., Rokne J. (eds) Encyclopedia of Social Network Analysis and Mining. Springer, New York, NY.

References III

- Dongol B, Hayes IJ, Struth G (2016) Convolution as a Unifying Concept: Applications in Separation Logic, Interval Calculi and Concurrency ACM Transactions on Computational Logic, Article No.: 15
- Falzon L, Quintane E, Dunn J, Robins G (2018) Embedding time in positions: Temporal measures of centrality for social network analysis. Social Networks 54, 168–178
- Gondran M, Minoux M (2008) Graphs, dioids and semirings: new models and algorithms. Springer, New York
- Harary F, Norman RZ, Cartwright D (1967) Structural Models: An Introduction to the Theory of Directed Graphs. New York: John Wiley and Sons
- Kontoleon N, Falzon L, Pattison P (2013) Algebraic structures for dynamic networks. Journal of Mathematical Psychology 57, 310–319
- Moore RE, Kearfott RB, Cloud MJ (2009). Introduction to Interval Analysis. Philadelphia: SIAM.
- Praprotnik S, Batagelj V (2016) Semirings for temporal network analysis. arXiv:1603.08261 [cs.SI]

References IV

- Quirin A, Cordón O, Santamaria J, Vargas-Quesada B, Moya-Anegón F (2008) A new variant of the Pathfinder algorithm to generate large visual science maps in cubic time. Inf Process Manag 44(4): 1611–1623
- Schvaneveldt RW, Dearholt DW, Durso FT (1988) Graph theoretic foundations of Pathfinder networks. Comput Math Appl 15(4):337–345
- Vavpetič A, Batagelj V, Podpečan V (2009) An implementation of the Pathfinder algorithm for sparse networks and its application on text networks. SiKDD.