



Temporal analysis of bibliographic networks

Vladimir Batagelj

IMFM Ljubljana and IAM UP Koper

Sunbelt XXXV

Hilton Metropole, Brighton, UK, June 23–28, 2015

- 1 Temporal networks
- 2 Temporal quantities
- 3 Operations
- 4 Bibliographic temporal networks
- 5 Conclusions



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

A *temporal network* $\mathcal{N}_T = (\mathcal{V}, \mathcal{L}, \mathcal{P}, \mathcal{W}, \mathcal{T})$ is obtained by attaching the *time*, \mathcal{T} , to an ordinary network where \mathcal{T} is a set of *time points*, $t \in \mathcal{T}$.

In a temporal network, nodes $v \in \mathcal{V}$ and links $l \in \mathcal{L}$ are not necessarily present or active in all time points. Let $T(v)$, $T \in \mathcal{P}$, be the *activity set* of time points for node v and $T(l)$, $T \in \mathcal{W}$, the activity set of time points for link l .

Besides the presence/absence of nodes and links also their properties can change through time.

We introduce a notion of a *temporal quantity*

$$a(t) = \begin{cases} a'(t) & t \in T_a \\ \mathbb{K} & t \in \mathcal{T} \setminus T_a \end{cases}$$

where T_a is the *activity time set* of a and $a'(t)$ is the value of a in an instant $t \in T_a$, and \mathbb{K} denotes the value *undefined*.

We assume that the values of temporal quantities belong to a set A which is a *semiring* $(A, +, \cdot, 0, 1)$ for binary operations $+$: $A \times A \rightarrow A$ and \cdot : $A \times A \rightarrow A$.

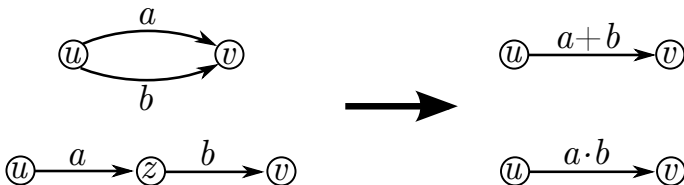
We can extend both operations to the set $A_{\mathbb{K}} = A \cup \{\mathbb{K}\}$ by requiring that for all $a \in A_{\mathbb{K}}$ it holds

$$a + \mathbb{K} = \mathbb{K} + a = a \quad \text{and} \quad a \cdot \mathbb{K} = \mathbb{K} \cdot a = \mathbb{K}.$$

The structure $(A_{\mathbb{K}}, +, \cdot, \mathbb{K}, 1)$ is also a semiring.

Semiring addition and multiplication in networks.

The “default” semiring is the *combinatorial* semiring $(\mathbb{R}_0^+, +, \cdot, 0, 1)$ where $+$ and \cdot are the usual addition and multiplication of real numbers.



In applications of semirings in analysis of graphs and networks the addition $+$ describes the composition of values on parallel paths and the multiplication \cdot describes the composition of values on sequential paths. For a combinatorial semiring these two schemes correspond to basic principles of combinatorics *Rule of Sum* and *Rule of Product*.

Let $A_{\mathfrak{K}}(\mathcal{T})$ denote the set of all temporal quantities over $A_{\mathfrak{K}}$ 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

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

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

A semiring is also the set of square matrices of order n over it for addition $\mathbf{A} \oplus \mathbf{B} = \mathbf{S}$

$$s_{ij} = a_{ij} + b_{ij}$$

and multiplication $\mathbf{A} \odot \mathbf{B} = \mathbf{P}$

$$p_{ij} = \sum_{k=1}^n a_{ik} \cdot b_{kj}.$$

The matrix multiplication is closely related to traveling on networks. For a value p_{ij} to be defined (different from \mathbb{K}) there should exist at least one node k such that both link (i, k) and link (k, j) exist (at the same time) – the transition from the node i to the node j through a node k is possible. Its contribution is $a_{ik} \cdot b_{kj}$.

In the following we shall limit our discussion to temporal quantities that can be described in the form of time-interval/value sequences

$$a = ((l_i, v_i))_{i=1}^k$$

where l_i is a time-interval, $T_a = \bigcup_{i=1}^k l_i$, and v_i is a constant value of a on this interval. To simplify the exposition we will assume in the following that all the intervals are of the form $[s_i, f_i)$, s_i is the starting time and f_i is the finishing time. Therefore we can describe the temporal quantities with sequences

$$a = ((s_i, f_i, v_i))_{i=1}^k$$

To provide a computational support for the proposed approach we are developing in Python a library TQ (Temporal Quantities). The following are two temporal quantities a and b represented in Python

```
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)]
```

The temporal quantity a has on interval $[1, 5)$ value 2, on interval $[6, 8)$ value 1, on interval $[11, 12)$ value 3, etc. Outside the specified intervals its value is undefined, ⌘.

Sum and product of temporal quantities

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

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

The following are the sum $s = a + b$ and the product $p = a \cdot b$ of temporal quantities a and b over combinatorial semiring.

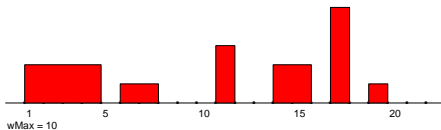
$$s = [(1, 2, 2), (2, 3, 6), (3, 4, 2), (4, 5, 5), (5, 6, 3), \\ (6, 7, 4), (7, 8, 1), (9, 10, 2), (11, 12, 3), \\ (13, 14, 5), (14, 15, 7), (15, 16, 2), (16, 17, 1), \\ (17, 18, 6), (18, 19, 1), (19, 20, 2), (20, 21, 1)]$$

$$p = [(2, 3, 8), (4, 5, 6), (6, 7, 3), (14, 15, 10), \\ (17, 18, 5), (19, 20, 1)]$$

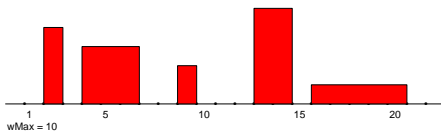
They are visually displayed at the bottom half of figures on the following slides.

Addition of temporal quantities.

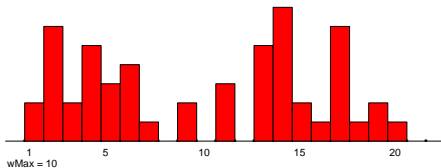
$a :$



$b :$

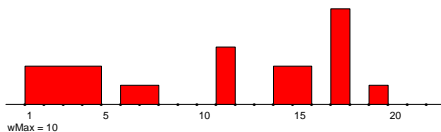


$a + b :$

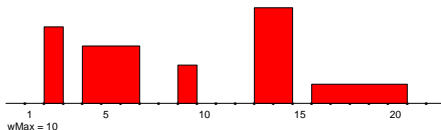


Multiplication of temporal quantities.

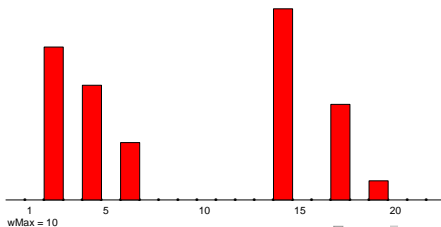
a :



b :



$a \cdot b$:



In some applications over the combinatorial semiring we shall use the *aggregated value* of a temporal quantity $a = ((s_i, f_i, v_i))_{i=1}^k$. It is defined as

$$\Sigma a = \sum_{i=1}^k (f_i - s_i) \cdot v_i$$

and is computed using the procedure *total*. For example $\Sigma a = 23$ and $\Sigma b = 30$. Note that $\Sigma a + \Sigma b = \Sigma(a + b)$.

From special bibliographies (**BibTeX**) and bibliographic services (**Web of Science**, **Scopus**, **SICRIS**, **CiteSeer**, **Zentralblatt MATH**, **Google Scholar**, **DBLP Bibliography**, **US patent office**, and others) we can

derive some two-mode networks on selected topics:

- works \times authors (**WA**),
- works \times keywords (**WK**);

and from some data also the network

- works \times classification (**WC**), and the
- one-mode citation network works \times works (**Ci**);

where works include papers, reports, books, patents etc.

Besides this we get also at least the partition of works by the journal or publisher and the partition of works by the publication year.

For converting WoS file into networks in Pajek's format a program **WoS2Pajek** was developed (in Python).

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

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

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

- **instantaneous** $\mathbf{A}_i = [a_{i_{ep}}]$, where

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

- **cumulative** $\mathbf{A}_c = [a_{c_{ep}}]$, where

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

Instantaneous **A** on $P \times A$ and **B** on $P \times B$. **C** = **A**^T·**B** on $A \times B$.

$$c_{ij}(t) = \sum_{p \in P} a_{pi}(t)^T \cdot b_{pj}(t)$$

$a_{pi} = [(d_{pi}, d_{pi} + 1, v_{pi})]$ and $b_{pj} = [(d_{pj}, d_{pj} + 1, v_{pj})]$
for $t = d$ we get

$$c_{ij} = [(d, d + 1, \sum_{p \in P: d_{pi}=d_{pj}=d} v_{pi} \cdot v_{pj})]_{d \in \mathcal{T}}$$

for $v_{pi} = v_{pj} = 1$ we finally get

$$v_{ij}(d) = |\{p \in P : d_{pi} = d_{pj} = d\}|$$

For binary temporal two-mode networks **A** and **B** the value $v_{ij}(d)$ of the product **A**^T·**B** is equal to the number of different members of P with which both i and j have contact in the instant d .

Multiplication of co-occurrence networks

Cumulative

Cumulative **A** on $P \times A$ and **B** on $P \times B$. **C** = **A**^T·**B** on $A \times B$.

$$c_{ij}(t) = \sum_{p \in P} a_{pi}(t)^T \cdot b_{pj}(t)$$

$a_{pi} = [(d_{pi}, last + 1, v_{pi})]$ and $b_{pj} = [(d_{pj}, last + 1, v_{pj})]$
for $t = d$ we get

$$c_{ij} = [(d, d + 1, \sum_{p \in P: (d_{pi} \leq d) \wedge (d_{pj} \leq d)} v_{pi} \cdot v_{pj})]_{d \in \mathcal{T}}$$

for $v_{pi} = v_{pj} = 1$ we finally get

$$v_{ij}(d) = |\{p \in P : (d_{pi} \leq d) \wedge (d_{pj} \leq d)\}|$$



Temporal co-authorship networks

Temporal
Bibliographic
Networks

V. Batagelj

Temporal
networks

Temporal
quantities

Operations

Bibliographic
temporal
networks

Conclusions

Using the multiplication of temporal matrices over the combinatorial semiring we get the corresponding instantaneous and cumulative co-occurrence matrices

$$\mathbf{C}_i = \mathbf{A}_i^T \cdot \mathbf{A}_i \quad \text{and} \quad \mathbf{C}_c = \mathbf{A}_c^T \cdot \mathbf{A}_c$$

A typical example of such a matrix is the papers authorship matrix \mathbf{WA} where E is the set of papers W , P is the set of authors A and d is the publication year.

The triple (s, f, v) in a temporal quantity ci_{pq} tells that in the time interval $[s, f)$ there were v events in which both p and q took part.

The triple (s, f, v) in a temporal quantity cc_{pq} tells that in the time interval $[s, f)$ there were in total v accumulated events in which both p and q took part.

The diagonal matrix entries ci_{pp} and cc_{pp} contain the temporal quantities counting the number of events in the time intervals in which the participant p took part.



Temporal co-authorship networks

Temporal Bibliographic Networks

V. Batagelj

Temporal
networks

Temporal
quantities

Operations

Bibliographic
temporal
networks

Conclusions

For an example, from a collection SN5 of network data about publications on social networks till 2008 we extracted data about 92 the most active researchers and transformed them to corresponding temporal networks: CiteInst, CiteCum, WAinst, WAcum, WKinst, WKcum and a partition W92 with outdegrees of works in the original WA network.

The matrices

$$\mathbf{Coi} = \mathbf{WAI}^T \cdot \mathbf{WAI} \quad \text{and} \quad \mathbf{Coc} = \mathbf{WAc}^T \cdot \mathbf{WAc}$$

describe the instantaneous co-authorship temporal network and the cumulative co-authorship temporal network.



92 authors

Temporal Bibliographic Networks

V. Batagelj

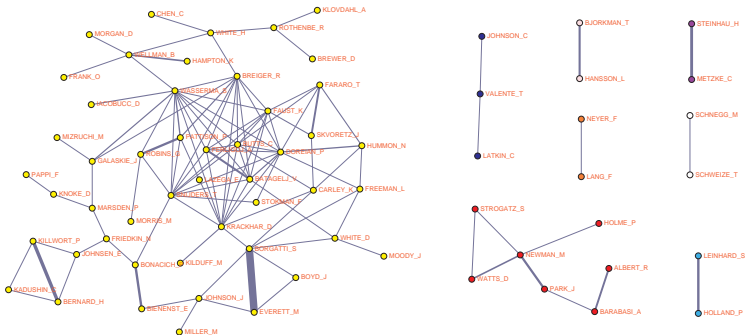
Temporal networks

Temporal quantities

Operations

Bibliographic temporal networks

Conclusions



- COHEN_C
- HAMMER_M
- MILARDO_R
- LITWIN_H
- YAMAGUCHI_K
- BERKMAN_L
- KIM_Y
- MARTIN_J
- SEKKULA_J
- SCOTT_J
- LEYDESDO_L
- MIKA_P
- ESTRADA_E
- PESCOSOL_B
- HIRSCH_B
- BURT_R
- SEIDMAN_S
- ANON_L
- COHEN_S
- LIN_N
- ROBERTS_J
- SMITH_R
- SMALL_H



Temporal co-authorship networks

Temporal
Bibliographic
Networks

V. Batagelj

Temporal
networks

Temporal
quantities

Operations

Bibliographic
temporal
networks

Conclusions

```
>>> import os, sys, datetime
>>> os.chdir("C:/Users/batagelj/work/Python/WoS/SN5/ten")
>>> from TQ import *
>>> wai = TQ.Ianus2Mat("WAinst.ten")
>>> wac = TQ.Ianus2Mat("WAcum.ten")
>>> list(wai.keys())
['dim', 'met', 'typ', 'nam', 'mat', 'til', 'tin', 'tit']
>>> wai['dim']
(1346, 92, 1970, 2008)
>>> WAi = wai['mat']; WAc = wac['mat']
>>> AWi = TQ.MatTrans(WAi); AWc = TQ.MatTrans(WAc)
>>> Coi = TQ.MatProd(AWi, WAi); Coc = TQ.MatProd(AWc, WAc)
>>> auNames = wai['nam'][nr:]
>>> ia=dict(zip(auNames, range(92)))
>>> Coi[ia['BORGATTI_S']][ia['EVERETT_M']]
[(1988, 1989, 1), (1989, 1990, 2), (1990, 1991, 4),
 (1991, 1992, 1), (1992, 1995, 2), (1996, 1998, 1),
 (1999, 2000, 3), (2003, 2004, 1), (2005, 2007, 1)]
>>> Coc[ia['BORGATTI_S']][ia['EVERETT_M']]
[(1988, 1989, 1), (1989, 1990, 3), (1990, 1991, 7),
 (1991, 1992, 8), (1992, 1993, 10), (1993, 1994, 12),
 (1994, 1996, 14), (1996, 1997, 15), (1997, 1999, 16),
 (1999, 2003, 19), (2003, 2005, 20), (2005, 2006, 21),
 (2006, 2008, 22)]
```

Using the multiplication of temporal matrices over the combinatorial semiring on bibliographic matrices **WA** and **WK** we get the corresponding instantaneous and cumulative matrices

$$\mathbf{AKi} = \mathbf{Wai}^T \cdot \mathbf{WKi} \quad \text{and} \quad \mathbf{AKc} = \mathbf{Wac}^T \cdot \mathbf{WKc}$$

The triple (s, f, v) in a temporal quantity aki_{ak} tells that in the time interval $[s, f)$ the author a used the keyword k v times (in v works).

The triple (s, f, v) in a temporal quantity akc_{ak} tells that in an instant t in the time interval $[s, f)$ the author a used cumulatively (till time t) the keyword k v times (in v works).

```
>>> wki = TQ.Ianus2Mat("WKinst.ten")
>>> AKi = TQ.MatProd(AWi,wki['mat'])
>>> kwNames = wki['nam'][nr:]
>>> len(kwNames)
8571
>>> ik=dict(zip(kwNames,range(8571)))
>>> Bc = [ AKi[i][ik['centrality']] for i in range(92)]
>>> [auNames[i] for i in range(92) if Bc[i]!=[]]
['BORGATTI_S', 'CARLEY_K', 'GALASKIE_J', 'BURT_R', 'FREEMAN_L',
'NEWMAN_M', 'BARABASI_A', 'WELLMAN_B', 'KNOKE_D', 'PAPPI_F',
'HOLME_P', 'WATTS_D', 'JOHNSON_C', 'WHITE_D', 'BREWER_D',
'MARSDEN_P', 'ROTHENBE_R', 'VALENTE_T', 'SNIJDERS_T',
'KRACKHAR_D', 'WHITE_H', 'KILDUFF_M', 'LEYDESDO_L',
'KLOVDAHL_A', 'MOODY_J', 'FRANK_O', 'BONACICH_P', 'BATAGELJ_V',
'JOHNSON_J', 'FAUST_K', 'MIZRUCHI_M', 'YAMAGUCH_K',
'FRIEDKIN_N', 'LAZEGA_E', 'CHEN_C', 'KILLWORT_P', 'ESTRADA_E',
'BUTTS_C', 'EVERETT_M', 'FERLIGOJ_A', 'IACOBUCD_D']
```

```
>>> T = [ (i,TQ.total(Bc[i])) for i in range(92) ]
>>> I = sorted(T,key=lambda e:e[1],reverse=True)
>>> [[auNames[i],v,Bc[i]] for (i,v) in I[:5]]
[['BORGATTI_S', 11, [(1991, 1992, 1), (1994, 1995, 1),
(1997, 1998, 1), (1999, 2000, 2), (2003, 2004, 1),
(2005, 2007, 2), (2007, 2008, 1)]],
['NEWMAN_M', 9, [(2001, 2002, 2), (2002, 2003, 1),
(2004, 2005, 2), (2005, 2006, 1), (2006, 2007, 2),
(2007, 2008, 1)]],
['BONACICH_P', 7, [(1986, 1988, 1), (1991, 1992, 1),
(1998, 1999, 1), (2001, 2002, 1), (2004, 2005, 2)]],
['EVERETT_M', 6, [(1997, 1998, 1), (1999, 2000, 2),
(2004, 2007, 1)]],
['CARLEY_K', 5, [(1999, 2000, 1), (2003, 2004, 1),
(2006, 2007, 3)]]]
```

A citation matrix \mathbf{Ci} describes the citation relation p cites q . Note that p cites $q \Rightarrow d(p) \geq d(q)$.

Then we can construct its instantaneous version \mathbf{Cii} :

$$cii_{pq} = [(d(p), d(p) + 1, 1)] \quad \text{iff} \quad ci_{pq} = 1$$

and its cumulative version \mathbf{Cic} :

$$cic_{pq} = [(d(p), last + 1, 1)] \quad \text{iff} \quad ci_{pq} = 1$$

Temporal versions of:

Bibliographic coupling $\mathbf{biCo} = \mathbf{Ci} \cdot \mathbf{Ci}^T$.

Co-citation $\mathbf{coCi} = \mathbf{Ci}^T \cdot \mathbf{Ci}$.

Citations between authors $\mathbf{Ca} = \mathbf{WA}^T \cdot \mathbf{Ci} \cdot \mathbf{WA}$.

$$\mathbf{ACA} = \mathbf{WAI}^T \cdot \mathbf{Cii} \cdot \mathbf{WAc}$$



Citations between authors

Temporal
Bibliographic
Networks

V. Batagelj

Temporal
networks

Temporal
quantities

Operations

Bibliographic
temporal
networks

Conclusions

```
>>> cite = TQ.Ianus2Mat("CiteInst.ten")
>>> Cite = cite['mat']; WAc = wac['mat']
>>> ACA = TQ.MatProd(TQ.MatProd(AWi,Cite),WAc)
>>> ACA[ia['WASSERMA_S']] [ia['HOLLAND_P']]
[(1977, 1978, 1), (1980, 1981, 5), (1981, 1982, 2),
 (1984, 1985, 2), (1985, 1986, 1), (1987, 1989, 2),
 (1990, 1991, 1), (1991, 1992, 2), (1992, 1994, 3),
 (1995, 1996, 2), (1996, 1997, 3), (1999, 2000, 5),
 (2000, 2001, 1), (2006, 2008, 1)]
>>> D = [(i,TQ.total(ACA[ia['DOREIAN_P']][i])) for i in range(92)]
>>> J = sorted(D,key=lambda e:e[1],reverse=True)
>>> [[auNames[i],v,ACA[ia['DOREIAN_P']][i]] for (i,v) in J[:5]]
[['DOREIAN_P', 69, [(1980, 1983, 1), (1984, 1985, 2),
 (1985, 1986, 1), (1986, 1987, 3), (1987, 1988, 2),
 (1988, 1989, 7), (1989, 1990, 5), (1990, 1991, 2),
 (1992, 1993, 6), (1994, 1995, 8), (1995, 1996, 2),
 (1996, 1997, 4), (2000, 2001, 3), (2001, 2004, 4),
 (2004, 2005, 6), (2006, 2007, 3)]],
 ['BREIGER_R', 26, [(1980, 1981, 3), (1984, 1986, 1),
 (1986, 1987, 2), (1987, 1988, 1), (1988, 1989, 4),
 (1989, 1990, 1), (1992, 1993, 3), (1994, 1995, 2),
 (1995, 1996, 1), (1996, 1997, 2), (2000, 2001, 1),
 (2004, 2005, 2), (2007, 2008, 2)]]]
```



Citations between authors

Temporal Bibliographic Networks

V. Batagelj

Temporal
networks

Temporal
quantities

Operations

Bibliographic
temporal
networks

Conclusions

```
['BURT_R', 20, [(1985, 1986, 3), (1986, 1987, 1),  
                (1987, 1988, 2), (1988, 1989, 5), (1989, 1990, 2),  
                (1992, 1993, 4), (1994, 1995, 1), (2000, 2001, 1),  
                (2004, 2005, 1)]]],  
['BATAGELJ_V', 17, [(1992, 1993, 2), (1994, 1995, 2),  
                    (1996, 1997, 4), (2000, 2001, 4), (2004, 2005, 5)]]],  
['FARARO_T', 15, [(1984, 1985, 1), (1985, 1986, 2),  
                  (1988, 1989, 2), (1989, 1990, 1), (1992, 1993, 1),  
                  (1995, 1996, 1), (2001, 2002, 3), (2002, 2003, 2),  
                  (2003, 2004, 1), (2006, 2007, 1)]]]]  
>>>
```

- temporal networks approach can give additional insight into bibliographic networks;
- we presented only some examples to show that it works. Many options have still to be elaborated;
- temporal networks methods produce large results. Special methods for identifying and presenting (visualizing) interesting parts need to be developed;
- current version of TQ library is based on matrices. This limits the application of the proposed methods to some thousands of nodes (space and time complexity). The limits can be partially extended by switching to graph representation.