Analysis of Peer Review data from WoS

part 4: an approach to temporal analysis

Vladimir Batagelj and Anuška Ferligoj

University of Ljubljana, IMFM Ljubljana and IAM UP Koper

PEERE

Vilnius – March 7-9, 2017

Outline

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

Work in progress

Temporal networks / paper

Soc. Netw. Anal. Min. (2016)6:28
DOI 10.1007/s13278-016-0330-4

CrossMark

ORIGINAL ARTICLE

An algebraic approach to temporal network analysis based on temporal quantities

Vladimir Batagelj[1] · Selena Praprotnik[2]

Received: 10 October 2015 / Revised: 6 April 2016 / Accepted: 8 April 2016

Abstract In a temporal network, the presence and activity of nodes and links can change through time. To describe temporal networks we introduce the notion of temporal

Mathematics Subject Classification 91D30 · 16Y60 · 90B10 · 68R10 · 93C55

Temporal networks

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.

Temporal quantities

We introduce a notion of a *temporal quantity*

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

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 ⌘ denotes the value *undefined*.

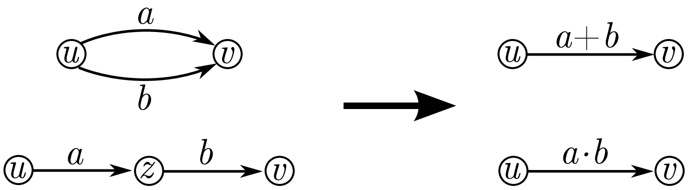
We assume that the values of temporal quantities belong to a set A which is a *semiring* $(A, +, \cdot, 0, 1)$ for binary operations $+ : A \times A \to A$ and $\cdot : A \times A \to A$.
We can extend both operations to the set $A_{⌘} = A \cup \{⌘\}$ by requiring that for all $a \in A_{⌘}$ it holds

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

The structure $(A_{⌘}, +, \cdot, ⌘, 1)$ 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*.

Operations with temporal quantities

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

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

Let us define the temporal quantities $\mathbf{0}$ 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.

Semiring for temporal quantities

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 ⌘) 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 $a_{ik} \cdot b_{kj}$.

Temporal quantities on intervals

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

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

where I_i is a time-interval, $T_a = \bigcup_{i=1}^{k} I_i$, 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

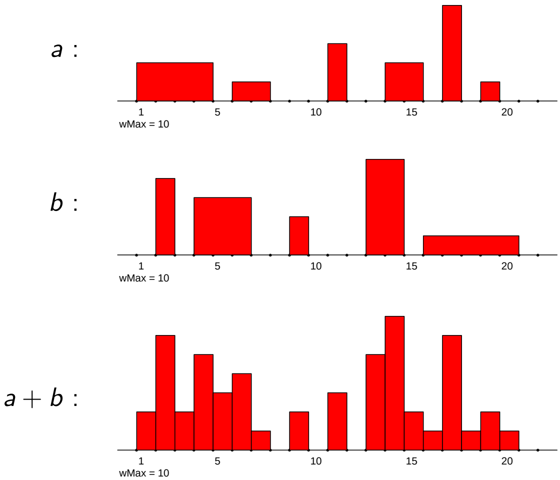

```
a = [(1, 5, 2), (6, 8, 1), (11, 12, 3), (14, 16, 2),
     (17, 18, 5), (19, 20, 1)]
b = [(2, 3, 4), (4, 7, 3), (9, 10, 2), (13, 15, 5), (16, 21, 1)]
```

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)]
```

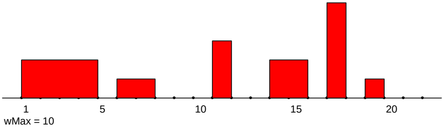
They are visually displayed at the bottom half of figures on the following slides.

Addition of temporal quantities.

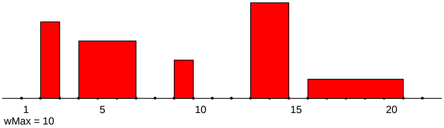


Multiplication of temporal quantities.

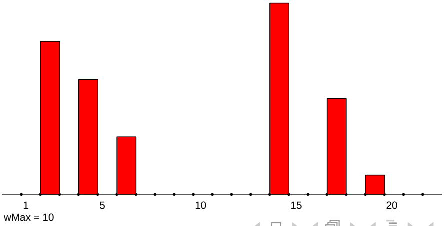
a :



b :



$a \cdot b$:



The aggregated value

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 $\Sigma a + \Sigma b = \Sigma(a + b)$.

Networks from bibliographic data

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 × authors (**WA**),
– works × keywords (**WK**);
and from some data also the network
– works × classification (**WC**), and the
– one-mode citation network works × 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).

Temporal co-occurrence networks

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

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

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

- **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

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

Example

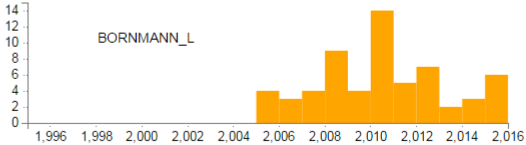
BORNMANN_L(2007)1:83 Bornmann, L., Daniel, H.-D. (2007). Gatekeepers of science – Effects of external reviewers' attributes on the assessments of fellowship applications. Journal of Informetrics, 1(1), 83-91.
DANIEL_H(2007):71 Daniel, H.-D., Mittag, S., Bornmann, L. (2007). The potential and problems of peer evaluation in higher education and research. In: A. Cavalli (Ed.), Quality Assessment for Higher Education in Europe (p. 71-82). London, UK: Portland Press.
BORNMANN_L(2007):149 Bornmann, L., Daniel, H.-D. (2007). Functional use of frequently and infrequently cited articles in citing publications. A content analysis of citations to articles with low and high citation counts. In Daniel Torres-Salinas, Henk F. Moed (Eds.), Proceedings of the 11th International Conference of the International Society for Scientometrics and Informetrics, 2007 (pp. 149-153). Madrid, Spain: Spanish Research Council (CSIC).
BORNMANN_L(2006)15:209 Bornmann, L., Daniel, H.-D. (2006). Potential sources of bias in research fellowship assessments. Effects of university prestige and field of study on approval and rejection of fellowship applications. Research Evaluation, 15(3), 209-219.
BORNMANN_L(2006)20:347 Bornmann, L. (2006). Peer-Review zur Auswahl von Forschungsstipendiaten. Eine Analyse der Fairness und prognostischen Validität des Auswahlprozesses mittels CHAID und GLM. Empirische Pädagogik, 20(4), 347-368.
MITTAG_S(2006)28:6 Mittag, S., Bornmann, L., Daniel, H.-D. (2006). Qualitätssicherung und -verbesserung von Studium und Lehre durch Evaluation. Akzeptanz und Folgen mehrstufiger Evaluationsverfahren. Beiträge zur Hochschulforschung, 28(2), 6-27.
BORNMANN_L(2006)68:427 Bornmann, L., Daniel, H.-D. (2006). Selecting scientific excellence through committee peer review – A citation analysis of publications previously published to approval or rejection of post-doctoral research fellowship applicants. Scientometrics, 68(3), 427-440.

Example WAi network

#(BORNMANN_L)=61

	BORNMANN_L	DANIEL_H	MITTAG_S
BORNMANN_L(2007)1:83}	(2007,2008,1)	(2007,2008,1)	
DANIEL_H(2007):71}	(2007,2008,1)	(2007,2008,1)	(2007,2008,1)
BORNMANN_L(2007):149}	(2007,2008,1)	(2007,2008,1)	
BORNMANN_L(2006)15:209}	(2006,2007,1)	(2006,2007,1)	
BORNMANN_L(2006)20:347}	(2006,2007,1)		
MITTAG_S(2006)28:6}	(2006,2007,1)	(2006,2007,1)	(2006,2007,1)
BORNMANN_L(2006)68:427}	(2006,2007,1)	(2006,2007,1)	

```
BORNMANN_L          (2006,2007,4), (2007,2008,3)
DANIEL_H            (2006,2008,3)
MITTAG_S            (2006,2007,1), (2007,2008,1)
```



Multiplication of co-occurence networks

Instantaneous

Instantaneous **A** on $P \times A$ and **B** on $P \times B$. $\mathbf{C} = \mathbf{A}^T.\mathbf{B}$ 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

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

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 $\mathbf{A}^T.\mathbf{B}$ 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-occurence networks

Cumulative

Cumulative **A** on $P \times A$ and **B** on $P \times B$. $\mathbf{C} = \mathbf{A}^T.\mathbf{B}$ 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

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

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

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

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

Temporal co-authorship networks

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

$$\mathbf{Ci} = \mathbf{Ai}^T \cdot \mathbf{Ai} \qquad \text{and} \qquad \mathbf{Cc} = \mathbf{Ac}^T \cdot \mathbf{Ac}$$

A typical example of such a matrix is the papers authorship matrix **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.

Peer review temporal co-authorship networks

From a collection WoS peer review network we extracted the data about works with complete description: pCiteD, pWAd, pWKd, pWJd, pYearD, pNPd and transformed the network pWAd into corresponding temporal networks pWAdInst pWAdCum in netJSON format.

In the network pWAd we have $|W| = 22104$, $|A| = 62106]$, and $nArcs = 80021$.
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-autorship temporal network and the cumulative co-autorship temporal network.

Fractional versions of temporal co-authorship networks can be also computed.

Temporal co-authorship networks

multiply.py

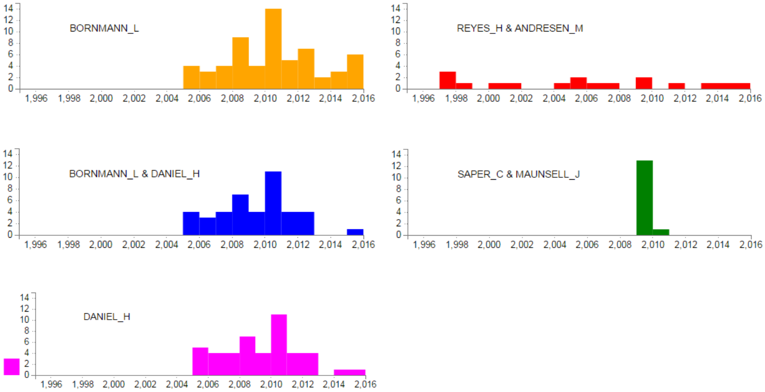
```
gdir = 'c:/users/batagelj/work/python/graph/graph'
wdir = 'c:/users/batagelj/work/python/graph/JSON/peere'
cdir = 'c:/users/batagelj/work/python/graph/chart'
import sys, os, datetime, json
sys.path = [gdir]+sys.path; os.chdir(wdir)
import TQ
from GraphNew import Graph
# file = 'C:/Users/batagelj/work/Python/graph/JSON/peere/pWAdCum.json'
file = 'C:/Users/batagelj/work/Python/graph/JSON/peere/pWAdInst.json'
t1 = datetime.datetime.now()
print("started: ",t1.ctime(),"\n")
G = Graph.loadNetJSON(file)
t2 = datetime.datetime.now()
print("\nloaded: ",t2.ctime(),"\ntime used: ", t2-t1)
# T = G.transpose()
# Co = Graph.TQmultiply(T,G,True)
# CR = G.TQtwo2oneRows()
CC = G.TQtwo2oneCols()
t3 = datetime.datetime.now()
print("\ncomputed: ",t3.ctime(),"\ntime used: ", t3-t2)
ia = { v[3]['lab']: k for k,v in CC._nodes.items() }
```

The most collaborating co-authors

```
>>> CC._links[(ia['BORNMANN_L'],ia['DANIEL_H'])][4]['tq']
[(2005, 2006, 4), (2006, 2007, 3), (2007, 2008, 4), (2008, 2009, 7), (2009, 2010, 4),
 (2010, 2011, 11), (2011, 2013, 4), (2015, 2016, 1)]
>>> CC._links[(ia['BROWN_D'],ia['RAFF_H'])][4]['tq']
[(2013, 2014, 11)]
>>> CC._links[(ia['SAPER_C'],ia['MAUNSELL_J'])][4]['tq']
[(2009, 2010, 13), (2010, 2011, 1)]
>>> CC._links[(ia['REYES_H'],ia['ANDRESEN_M'])][4]['tq']
[(1997, 1998, 3), (1998, 1999, 1), (2000, 2002, 1), (2004, 2005, 1), (2005, 2006, 2),
 (2006, 2008, 1), (2009, 2010, 2), (2011, 2012, 1), (2013, 2016, 1)]
>>> CC._links[(ia['KRAVITZ_R'],ia['FELDMAN_M'])][4]['tq']
[(2010, 2016, 1)]
>>> CC._links[(ia['DEANGELI_C'],ia['FONTANAR_P'])][4]['tq']
[(2000, 2002, 1), (2003, 2004, 1), (2005, 2012, 1)]
>>> CC._links[(ia['DANIEL_H'],ia['DANIEL_H'])][4]['tq']
[(1993, 1994, 3), (2005, 2006, 5), (2006, 2008, 4), (2008, 2009, 7), (2009, 2010, 4),
 (2010, 2011, 11), (2011, 2013, 4), (2014, 2016, 1)]
>>> CC._links[(ia['BORNMANN_L'],ia['BORNMANN_L'])][4]['tq']
[(2005, 2006, 4), (2006, 2007, 3), (2007, 2008, 4), (2008, 2009, 9), (2009, 2010, 4),
 (2010, 2011, 14), (2011, 2012, 5), (2012, 2013, 7), (2013, 2014, 2), (2014, 2015, 3),
 (2015, 2016, 6)]
>>> bb = CC._links[(ia['BORNMANN_L'],ia['BORNMANN_L'])][4]['tq']

>>> tit = 'BORNMANN_L'
>>> TQmax = 15; Tmin = 1995; Tmax = 2016; w = 600; h = 150
>>> Graph.TQshow(bb,cdir,TQmax,Tmin,Tmax,w,h,tit,fill='orange')
```

Co-authors / pictures



Co-authors / cumulative

Peer Review from WoS

V. Batagelj, A. Ferligoj

Temporal networks

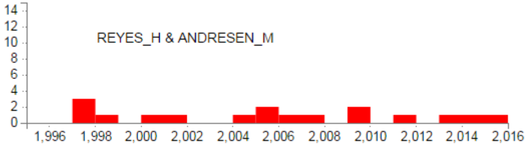
Temporal quantities

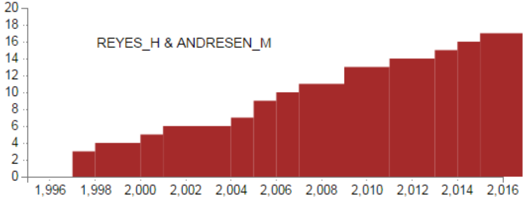
Operations

Bibliographic temporal networks

Conclusions

```
file = 'C:/Users/batagelj/work/Python/graph/JSON/peere/pWAdCum.json'
---
>>> rac = CC._links[(ia['REYES_H'],ia['ANDRESEN_M'])][4]['tq']
>>> tit = 'REYES_H & ANDRESEN_M'
>>> TQmax = 20; Tmin = 1995; Tmax = 2016; w = 600; h = 200
>>> Graph.TQshow(rac,cdir,TQmax,Tmin,Tmax,w,h,tit,fill='brown')
```





Authors and keywords

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

$$\mathbf{AKi} = \mathbf{WAi}^T \cdot \mathbf{WKi} \qquad \text{and} \qquad \mathbf{AKc} = \mathbf{WAc}^T \cdot \mathbf{WKc}$$

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).

In September 2016 we developed algorithms for determining temporal ordinary and P_S-cores that allow us to identify evolving groups in temporal networks.

Temporal citation networks

A citation matrix **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 **Cii**:

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

and its cumulative version **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 $\mathbf{Ca} = \mathbf{WA}^T \cdot \mathbf{Ci} \cdot \mathbf{WA}$.

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

Conclusions

- 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). A much faster version of TQ library based on a graph representation is under development.
  $|A| = 62106$, $nEdges(CC) = 633977$, $633977/1928608671 = 0.0003287$