



Event  
sequence  
analysis

V. Batagelj &

Introduction

Sequence  
analysis

Oštro

Conclusions

References

References

# Event sequence analysis

## Network analysis of sequences of events

**Vladimir Batagelj, Anuška Ferligoj, Oštro Center**  
IMFM, FDV, UP IAM, Oštro

**1343+1344. sredin seminar**  
Ljubljana, March 6, 2024

# Outline

Event  
sequence  
analysis

V. Batagelj &

Introduction

Sequence  
analysis

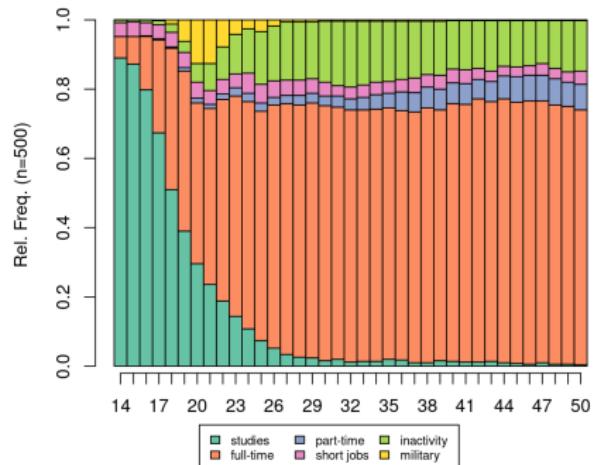
Oštro

Conclusions

References

References

- 1 Introduction
- 2 Sequence analysis
- 3 Oštro
- 4 Conclusions
- 5 References



Vladimir Batagelj: [vladimir.batagelj@fmf.uni-lj.si](mailto:vladimir.batagelj@fmf.uni-lj.si)

Current version of slides (March 13, 2024 at 06:29): [slides PDF](#)

<https://github.com/bavla/TQ/tree/master/trajectories/Feb24>



# Oštro

Event  
sequence  
analysis

V. Batagelj &

Introduction

Sequence  
analysis

Oštro

Conclusions

References

References

Last year in mid-November, Nuša was contacted by the Center for Investigative Journalism **Oštro**, asking if she could advise them on the analysis of data on members of the current Slovenian government and parliamentarians. Nuša agreed to the meeting, which I also attended. It turned out that Oštro collected the CVs of all the mentioned politicians. We found the data interesting. So we agreed to see how we could tackle them.

Person  $X$ 's CV

$$CV_X = (e_1, e_2, \dots, e_{k_x})$$

consists of a sequence of events  $e_i$ . For an event  $e_i$

$$e_i = (s_i, f_i, r_i, q_i, \dots)$$

we at least know its start date  $s_i$ , its end (finish) date  $f_i$ , the type  $r_i$  of the event, the location  $q_i$  of the event, and maybe something else.



# Sequence analysis

## Event sequence analysis

V. Batagelj &

Introduction

Sequence  
analysis

Ostro

Conclusions

References

References

Event sequence data encode sequences of events over time.  
I became aware of this kind of data years ago because for their analysis [1–4] results from my article on the generalized Ward criterion function [5] were used.

This subfield of longitudinal data analysis is closely related to time series analysis but the emphasis is on **categorical** values represented as states.

It is described by several tags: sequence analysis, life course, trajectory, career path, event history, multichannel sequence, discrepancy analysis, etc.

There is **The Sequence Analysis Association (SAA)**.



# Sequence analysis

## Books

Event  
sequence  
analysis

V. Batagelj &

Introduction

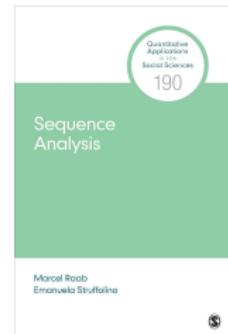
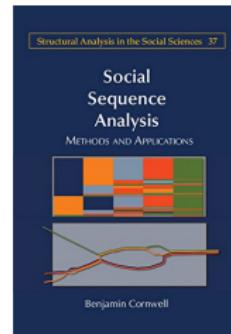
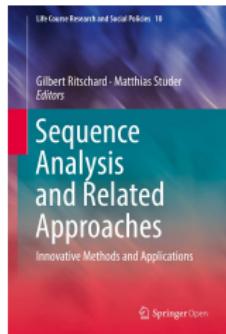
Sequence  
analysis

Ostro

Conclusions

References

References

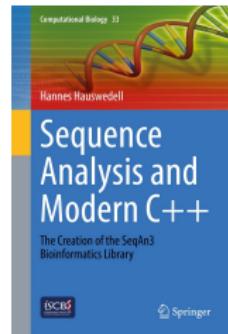
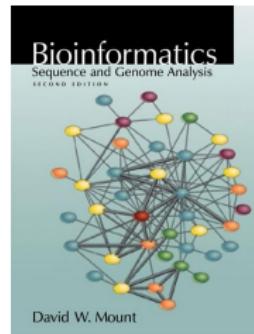
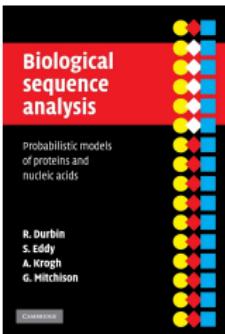
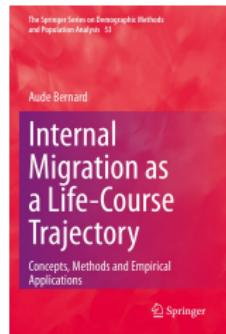


2018

2014

2015

2022



2022

1998

2004

2022



# Sequence analysis

## Fields of application

Event  
sequence  
analysis

V. Batagelj &

Introduction

Sequence  
analysis

Ostro

Conclusions

References

References

Recent overviews [6–9].

Sequence analysis in social sciences.

Fields of application:

- IT processes logs,
- healthcare,
- social science: life course research, career paths, daily patterns, life (family, housing, employment, demographic, etc.) trajectories,
- bank transactions,
- social media,
- sports,
- music, text,
- bibliometrics,
- etc.



# Sequence analysis

## Example gene homologies / Pax6 gene

### Event sequence analysis

V. Batagelj &

Introduction

Sequence  
analysis

Ostro

Conclusions

References

References

Mouse	GTATCCAACGGTTGTGAGTAAAATTCTGGCAGGTATTACGAGACTGGCTCCATCAGA
Fly	GTATCAATGGATGTGTGAGCAAATCCTCGGAGGTATTATGAACAGGAACCATAACGA
Shark	GTGTCCAACGGTTGTGTCACTAAATCTGGCAGATACTATGAACAGGATCCATCAGA
Squid	GTCTCCAACGGCTCGTTAGCAAGATTCTCGGCAGGTACTATGAGACGGGCTCCATAAGA
Worm	GTGTCTAATGGTTGTGTTAGTAAAATACTTGCCTATTATGGAACAGGTTCTATTAAA

video



# Sequence analysis

## Example music [10]

Event  
sequence  
analysis

V. Batagelj &

Introduction

Sequence  
analysis

Ostro

Conclusions

References

References



Silvio Hein, "Maria Cahill's Arab Love Song"



Joseph E. Howard, "I Think I Hear a Woodpecker"



# Sequence analysis

## Example KEDS, Cameo

### Event sequence analysis

V. Batagelj &

Introduction

Sequence analysis

Ostro

Conclusions

References

References

...	910625	YUG	YUGCRO	198	(WITHDRAW FROM) DECLARED INDEPENDENCE
	910625	YUG	SLO	198	(WITHDRAW FROM) DECLARED INDEPENDENCE
	910625	YUGCRO	YUG	081	(MAKE AGREEMENT) IS PLANS
	910625	YUGCRO	YUG	198	(WITHDRAW FROM) DECLARE INDEPENDENCE
	910625	USA	YUG	102	(URGE) REITERATED
	910625	GER	YUG	031	(MEET) TELEPHONED
	910625	YUG	GER	031	(MEET) TELEPHONED
	910625	ITA	YUG	031	(MEET) TELEPHONED
	910625	YUG	ITA	031	(MEET) TELEPHONED
	910625	FRN	SLO	121	(CRITICIZE) DISAPPROVED
	910625	FRN	YUGCRO	121	(CRITICIZE) DISAPPROVED
	910626	SLO	YUG	043	(RALLY) CELEBRATED
	910626	YUGCRO	SERMIL	223	(MIL ENGAGEMENT) KILLED
	910626	USR	YUGCRO	042	(ENDORSE) SOVIET SUPPORT FOR CROATIA
	910626	USR	SLO	042	(ENDORSE) SOVIET SUPPORT FOR CROATIA
	910626	YUG	SLO	182	(MILITARY DEMO) PUT ON ALERT
	910626	YUG	YUGCRO	223	(MIL ENGAGEMENT) KILLED
	910626	SERMIL	SLO	199	(-) COMBAT
	910626	YUGGOV	SLO	197	(CENSOR) OUTLAWED
	910626	YUGGOV	YUGCRO	197	(CENSOR) OUTLAWED
	910626	YUGGOV	SLO	210	(SEIZE) SEIZE
	910626	AUL	YUGCRO	022	(PESSIMIST COMMENT) SAID NOT
	910626	USAGOV	YUG	022	(PESSIMIST COMMENT) SAID NOT
	910626	USA	YUG	112	(REFUSE) OPPOSED
	910626	USA	EUR	160	(WARN) WARNED
	910626	UNK	YUG	094	(CALL FOR) CALLED ON
	910626	NTH	EEC	022	(PESSIMIST COMMENT) SAID NOT
	910626	GERGOV	YUG	102	(URGE) URGED
	910627	SLO	SERMIL	223	(MIL ENGAGEMENT) FIGHTING ERUPTED
	910627	SLO	SERMIL	223	(MIL ENGAGEMENT) SHOT
	910627	SERMIL	SLO	031	(MEET) SENT
	910627	SLO	SERMIL	121	(CRITICIZE) SLOVENIA SAID YUGOSLAV SOLDIER KILLED
	...				



# Sequence analysis

## Example purchase

Event  
sequence  
analysis

V. Batagelj &

Introduction

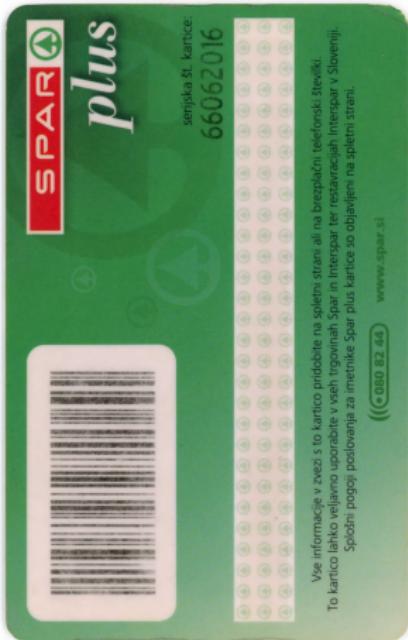
Sequence  
analysis

Ostro

Conclusions

References

References



**SPAR** ♠

SUPERMARKET KAPITELJ, Poljanski nasip 8, 1000 Ljubljana  
Spar Slovenija d.o.o., Letališka c. 26, 1000 Ljubljana  
ID DDV SI32156782

Ljubljana 04.03.2024 15:57 B1 101 Bon 1442 U 226

Izdelek	Količina	Cena	Popust	Znesek
OLJE OLJČNO ILIADA 1L	1	21.98	5.50	16.48 B
STEGNA PIŠČ.PRE.BKK IK	1	5.69	0.00	5.69 B
BLAZINICE NIVEA 80/1	1	2.59	0.00	2.59 C
MASLO PLANIKA 250G	1	3.89	0.00	3.89 B
PAPIR TOAL.ZEWA 10/1	1	5.99	2.00	3.99 C
FIŽOL RJAVA 190G	2	1.15	0.00	2.30 B
SPAR PR. MUSLI MAND.37	1	3.79	0.00	3.79 B
VINO MALVAZ.VK 0,75L	1	4.99	0.00	4.99 C
BANANE	1.020	1.49	0.61	0.91 B
JAJCA M 6/1	1	2.68	0.00	2.68 B
Skupaj		EUR		47.31



# Sequence analysis

## Example Ris

### Event sequence analysis

V. Batagelj &

Introduction

Sequence  
analysis

Ostro

Conclusions

References

References

TY - JOUR  
AU - Sun, F.  
AU - Chen, W.  
AU - Lin, T.  
TI - Improved Voting Algorithm Based on K-truss for Influence Maximization Problem  
PY - 2022  
T2 - Jisuanji Gongcheng/Computer Engineering  
VL - 48  
IS - 11  
SP - 291  
EP - 298  
DO - 10.19678/j.issn.1000-3428.0063033  
UR - <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85143346303&doi=10.19678%2fj.issn.1000-3428&partnerID=40>  
AD - School of Computer Science, South China Normal University, Guangzhou, 510631, China  
AB - In the Influence Maximization(IM) problem of social networks, approximation algorithms are proposed to solve it. In this paper, we propose a voting algorithm based on K-truss decomposition to solve the problem. The experimental results show that our algorithm can obtain better performance than other algorithms.  
KW - IC model  
KW - Influence Maximization(IM)  
KW - K-truss decomposition  
KW - SIR model  
KW - Social network  
KW - voting algorithm  
PB - Editorial Office of Computer Engineering  
SN - 10003428 (ISSN)  
LA - Chinese  
J2 - Jisuanji Gongcheng  
M3 - Article  
DB - Scopus  
N1 - Export Date: 15 November 2023; Cited By: 0; Correspondence Address: F. Sun; School of Computer Science, South China Normal University, Guangzhou, 510631, China  
ER -

# Event sequences

## Example hospital

### Event sequence analysis

V. Batagelj &

Introduction

Sequence analysis

Ostro

Conclusions

References

References

A

### Event enumeration (OCEL)

22.10.2014 12:13	ER Registration	John Smith
22.10.2014 22:31	ER Registration	William Brown
23.10.2014 08:23	ER Sepsis Triage	John Smith
24.10.2014 08:11	CRP	John Smith
24.10.2014 09:19	ER Sepsis Triage	William Brown
25.10.2014 11:47	Leucocytes	William Brown
26.10.2014 11:48	IV Antibiotic	William Brown
26.10.2014 17:50	Release A	John Smith
28.10.2014 15:01	Release A	William Brown

B

### Case-event enumeration (XES)

John Smith

22.10.2014 12:13	ER Registration
23.10.2014 08:23	ER Sepsis Triage
24.10.2014 08:11	CRP
26.10.2014 17:50	Release A

William Brown

22.10.2014 22:31	ER Registration
24.10.2014 09:19	ER Sepsis Triage
25.10.2014 11:47	Leucocytes
26.10.2014 11:48	IV Antibiotic
28.10.2014 15:01	Release A

Object-Centric Event Log (OCEL)

1849-2023 - IEEE Standard for eXtensible Event Stream (XES)

# Event sequences

Diaries – A weekday of a group of individuals aged 22-30 [11]

## Event sequence analysis

V. Batagelj &

Introduction

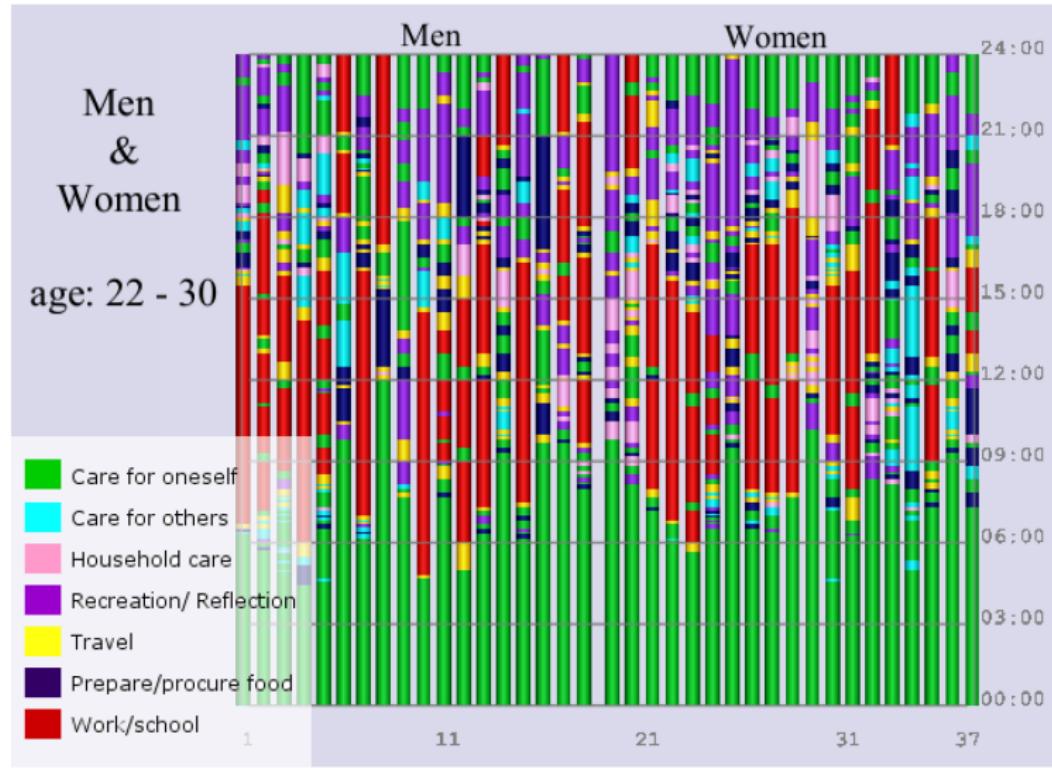
Sequence analysis

Ostro

Conclusions

References

References



# Event sequences

Diarie – A weekday of a group of individuals aged 22-30 [11]

## Event sequence analysis

V. Batagelj &

Introduction

Sequence analysis

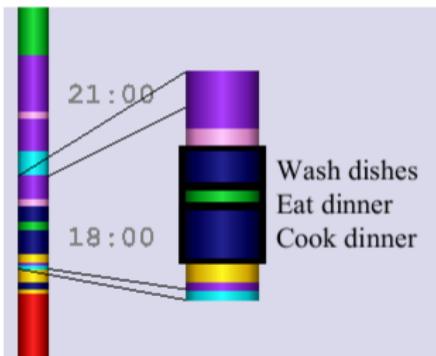
Ostro

Conclusions

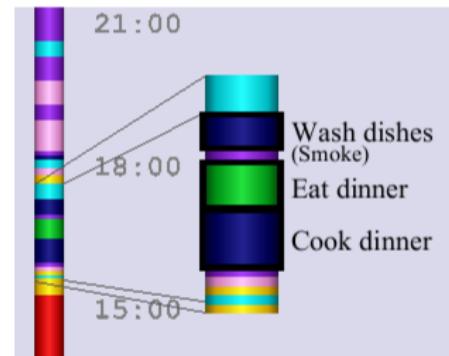
References

References

Examples of the activity sequence (*tuple*) “cook dinner → eat dinner → wash dishes” integrated in different ways in two individuals’ diaries



a) a zero gap match



b) a gap = 1 match



# Sequence analysis

## basic notions

Event  
sequence  
analysis

V. Batagelj &

Introduction

Sequence  
analysis

Ostro

Conclusions

References

References

A *raw event sequence*  $S$  is a list of events

$$S = (e_1, e_2, \dots, e_n)$$

$n$  is the *length* of the sequence  $S$ . Each event is a list of *named attributes*

$$e_i = (ind = i, t = t_i, a_1 = a_{1i}, a_2 = a_{2i}, \dots, a_m = a_{mi})$$

where  $ind$  is the event's *index*,  $t$  is the event's *temporal* information, and  $a_i$ s are *data attributes* describing different event characteristics. Their values can be structured objects.

In an event, some attributes can be missing. Each event has at least one data attribute defined (has value). An attribute  $a_i$  is called *general* if it is defined for all events in the sequence  $S$ . The attribute  $ind$  and usually also the time  $t$  are general attributes.

An alternative is to list for each event all its attributes with “missing” attributes having value *undefined* □.

# Sequence analysis

time

The temporal information can be described in different ways

- implicitly – the time is determined by the event's index. It expresses the ordering of events,
- $t_i \in \mathbb{N}$  – it expresses the ordering of events or starts of events on a discrete timeline (hours, days, years, etc.),
- $t_i = (s_i, f_i)$  – it specifies the time interval of event  $e_i$ ;  $s_i$  is the start and  $f_i$  is the finish time of the event;  $s_i \leq f_i$ ,
- $t_i = (s_i, d_i)$  – it specifies the time interval of event  $e_i$ ;  $s_i$  is the start and  $d_i$  is the duration time of the event;  $d_i \geq 0$ .

Quantities  $s_i$  and  $f_i$  are expressed as numbers or as date-time stamps. In some applications, an event's time can be of different types: *interval* (both values are defined) and *instantaneous* (the second value is undefined).

The time can be measured on a *common* timeline or on a *relative* timeline (for example from a birthday of each person).

# Sequence analysis

Event  
sequence  
analysis

V. Batagelj &

Introduction

Sequence  
analysis

Ostro

Conclusions

References

References

In the dataset that includes events in the careers of 40 professors, we have both: point events (e.g., time they received their bachelor's or master's degrees, published a journal or conference paper) and interval events (e.g., appointments as an assistant or associate professor).

We denote

- $e_i$  – event with  $ind = i$ ,
- $e_i.a$  – value of the attribute  $a$  in the event  $e_i$ .

Often in our analyses, we do not need all the available data attributes and some attributes need to be recoded – we transform the raw event sequence  $S$  into a work event sequence  $S'$ .

Usually, we also have some additional data about the (values of) attributes used in the event description. For example: address, age, gender, etc. of persons.



# Sequence analysis

## event sequences

Event  
sequence  
analysis

V. Batagelj &

Introduction

Sequence  
analysis

Ostro

Conclusions

References

References

Let  $\pi$  be a *partition* of  $1 : n$  into  $k$  classes. A sublist of  $S$

$$S_j^\pi = (e_i : (e_i \in S) \wedge (\pi(i) = j))$$

is the *event sequence* of the class  $j$ . The order of events in a list  $S_j$  has to be compatible with the time  $t$ :  $p < q \Rightarrow s_p \leq s_q$ . We get a *set of event sequences*  $E^\pi = \{S_j^\pi : j \in 1 : k\}$ .

Usually, for a categorical general attribute  $a$  with a range  $C$ , the partition  $\pi_a$  is defined by

$$\pi_a(i) = \text{ind}(a_i)$$

where the function  $\text{ind}$  returns the position (index) of the value  $a_i$  in the list  $C$ .

Note that in a general event sequence, some events can be active in parallel.



# Sequence analysis

## Problems

Event  
sequence  
analysis

V. Batagelj &

Introduction

Sequence  
analysis

Ostro

Conclusions

References

References

Sequence analysis deals with questions about the set of event sequences  $E$

- structure of the set  $E$  (statistics, clustering, visualization)
- models that generate sequences from the set  $E$  (transition matrices, Petri nets, Markov chains)
- frequent patterns – motifs (subsequences) in sequences from  $E$
- networks derived from  $E$
- etc

# Sequence analysis

## History

Event  
sequence  
analysis

V. Batagelj &

Introduction

Sequence  
analysis

Ostro

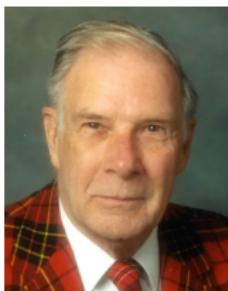
Conclusions

References

References

Event sequences were studied already in the first half of the 20th century in Engineering and Operation research (tasks scheduling, CPM, Pert, transportation) and Computer and Communication sciences (signal processing, coding, language and automata theory [12, 13]). In the second half of the 20th century, they found an application in Biology in studying DNA [14].

Sequence Analysis (SA) was introduced in the field of social sciences in 1983 by Andrew Abbott [15] and by Abbott and Forrest [16] (1986). It has become a very popular tool for studying trajectories.



Hamming



Levenshtein



Abbott

# Sequence analysis

## Journal publication and citation trends [6]

Event  
sequence  
analysis

V. Batagelj &

Introduction

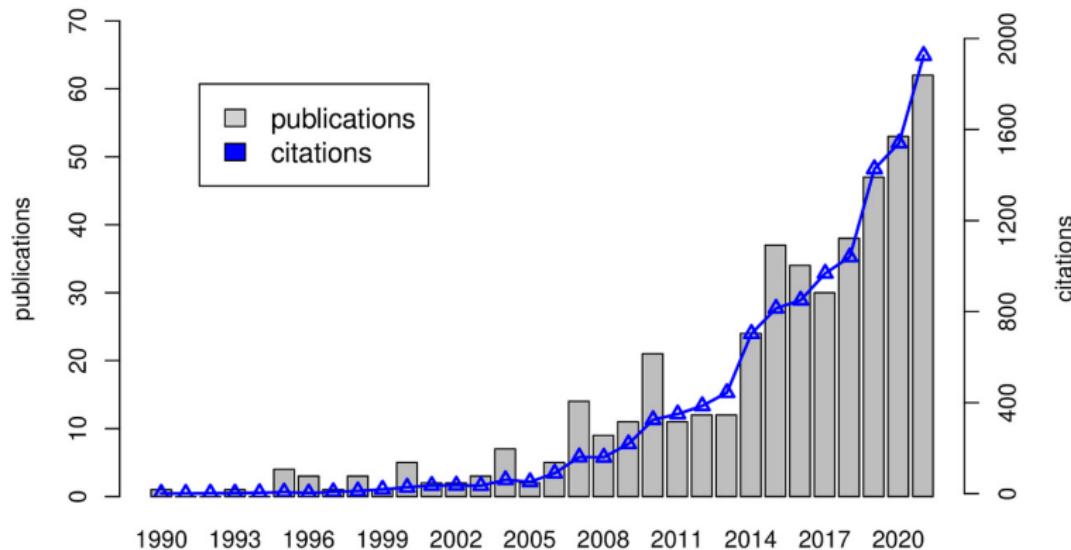
Sequence  
analysis

Ostro

Conclusions

References

References





# Sequence analysis

Event  
sequence  
analysis

V. Batagelj &

Introduction

Sequence  
analysis

Ostro

Conclusions

References

References

A typical application of SA, according to Abbott and Tsay [17], involves three steps: coding narratives or processes as sequences, measuring pairwise dissimilarities between sequences, and some form of data reduction such as cluster analysis, and post-analysis of the obtained results.

Often in the data-cleaning phase, we need to reduce the number of states – alphabet reduction: state  $< 2\% \rightarrow$  “other”. Also in the case of multiple states at the same time point are replaced by the dominant state.

For details see:

Fasang, A, Struffolino, E (2022) *Sequence Analysis for Social Science* [18]

Robette, N (2023) *Tutorial on sequence analysis* [19]

# Sequence analysis

## Basic steps [video](#)

Event  
sequence  
analysis

V. Batagelj &

Introduction

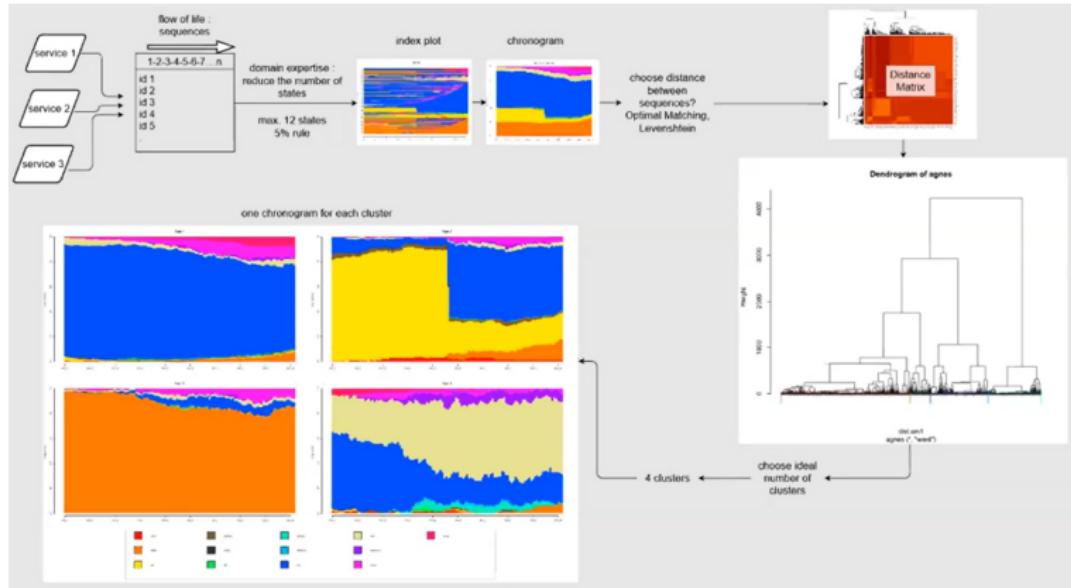
Sequence  
analysis

Ostro

Conclusions

References

References





# Sequence analysis

## Sequence (dis)similarity measures [20]

Event  
sequence  
analysis

V. Batagelj &

Introduction

Sequence  
analysis

Ostro

Conclusions

References

References

- Hamming distance (Hamming, 1950) and dynamic HD (Lesnard, 2010)
- Optimal matching (OM) distance based on state editing (substitution, insertion, deletion; Abbott & Forrest, 1986 [16]) – essentially the Levenshtein distance (Levenshtein, 1965 [13])
- Elzinga's combinatorial measures (Elzinga, 2003, 2005, 2007)
- TWED – Time warp edit distance (Marteau, 2009)
- DTW – Dynamic time warping distance (Rabiner & Juang, 1993)
- LCS – Longest common subsequence (Wagner & Fischer, 1974)



# Sequence analysis

## Datasets

Event  
sequence  
analysis

V. Batagelj &

Introduction

Sequence  
analysis

Ostro

Conclusions

References

References

- Sample Datasets
- UC Irvine Machine Learning Repository
- IEEE Task Force Event Log repository
- International Event Data Sets; books
- Kaggle / event sequence



# Sequence analysis

## Software [21]

Event  
sequence  
analysis

V. Batagelj &

Introduction

Sequence  
analysis

Ostro

Conclusions

References

References

### The Sequence Analysis Association / software

- TraMineR/CRAN,
- EventFlow, EventAction,
- Stata: SADI, SQ,
- Gantt Chart, LiveGantt,
- Flow, LifeFlow, OutFlow, DecisionFlow,
- StoryLines, Narrative, StoryFlow,
- MatrixFlow, MatrixWave,
- Episogram, CarePathway, Frequency, Peekquence, Logan, TreatmentExplorer,
- EventThread, PatternFinder, Similan2,
- VISUAL-TimePACTS,

# Event sequences

Clusters, state proportion plots, modal plots [6]

## Event sequence analysis

V. Batagelj &

Introduction

Sequence analysis

Ostro

Conclusions

References

References

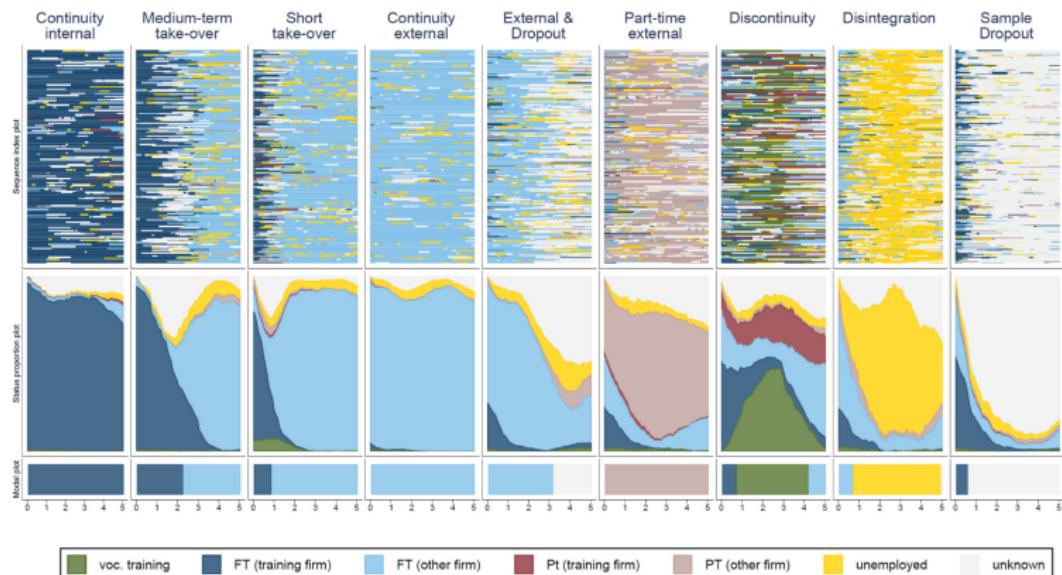


Fig. 3. Example of combination of sequence index plots, state proportion plots and modal plots; taken from Brzinsky-Fay et al. (2016).

# Event sequences

Job mobility sequences ordered by start level [22]

## Event sequence analysis

V. Batagelj &

Introduction

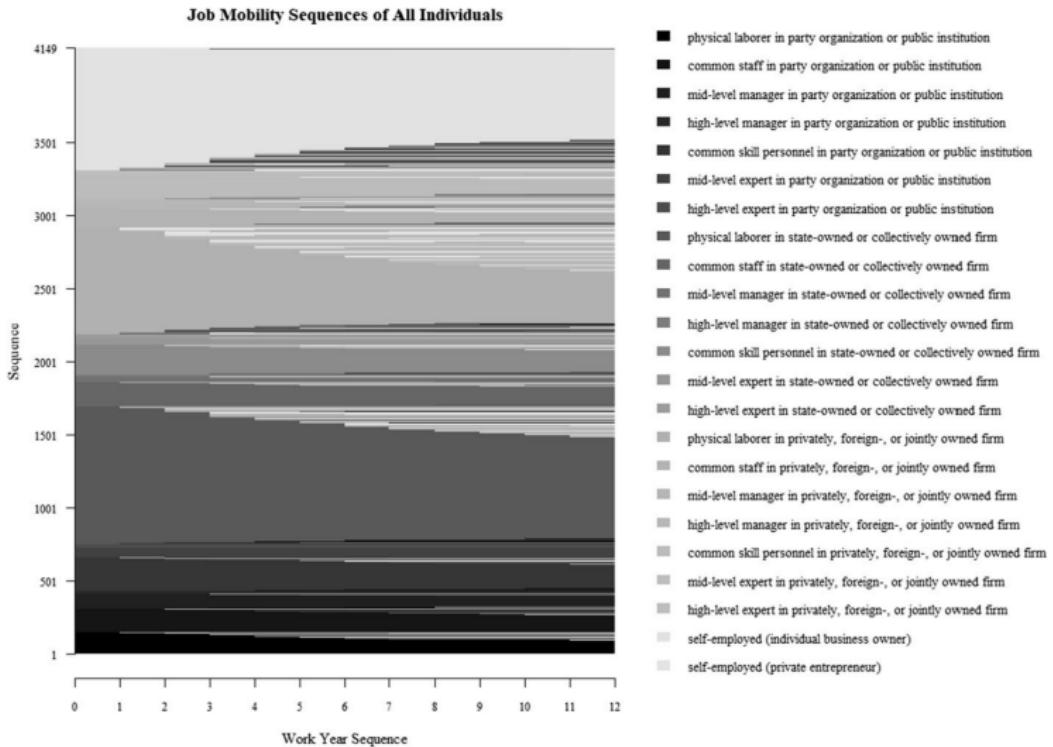
Sequence analysis

Ostro

Conclusions

References

References



# Event sequences

Interpreted dendrogram, Ellersgaard (2019)

## Event sequence analysis

V. Batagelj &

Introduction

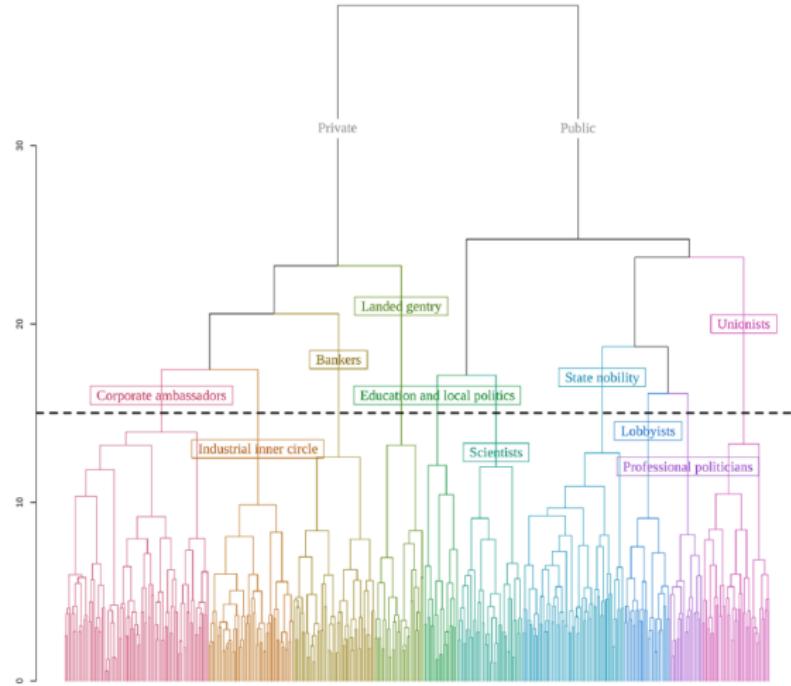
Sequence analysis

Ostro

Conclusions

References

References



# Event sequences

## Partition based on selected attribute [2]

### Event sequence analysis

V. Batagelj &

Introduction

Sequence analysis

Ostro

Conclusions

References

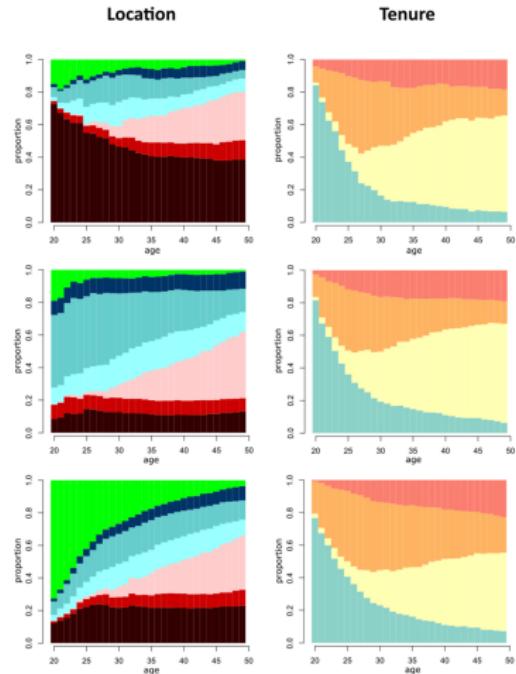
References

People socialized  
during childhood ...

... in Paris

... in the suburbs

... out of the Paris region



- Dense subway
- Scarce subway
- Express train
- Railway and motorway
- Railway
- No railway
- Out of Paris region
- At parents' or other status
- Homeownership
- Rental
- Social housing

# Event sequences

Basketball Play-by-play Analysis, Monroe (2013, *slides*)

## Event sequence analysis

V. Batagelj &

Introduction

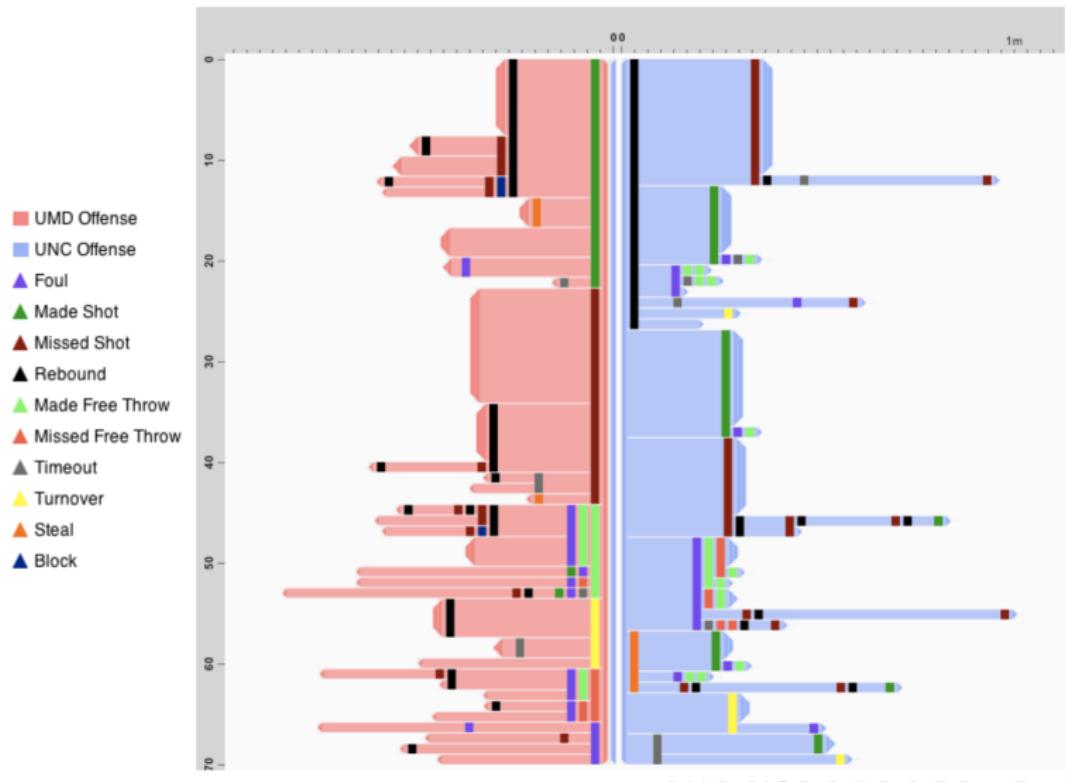
Sequence analysis

Ostro

Conclusions

References

References



# Event sequences

Hierarchical plot, Stanojevic (2021)

## Event sequence analysis

V. Batagelj &

Introduction

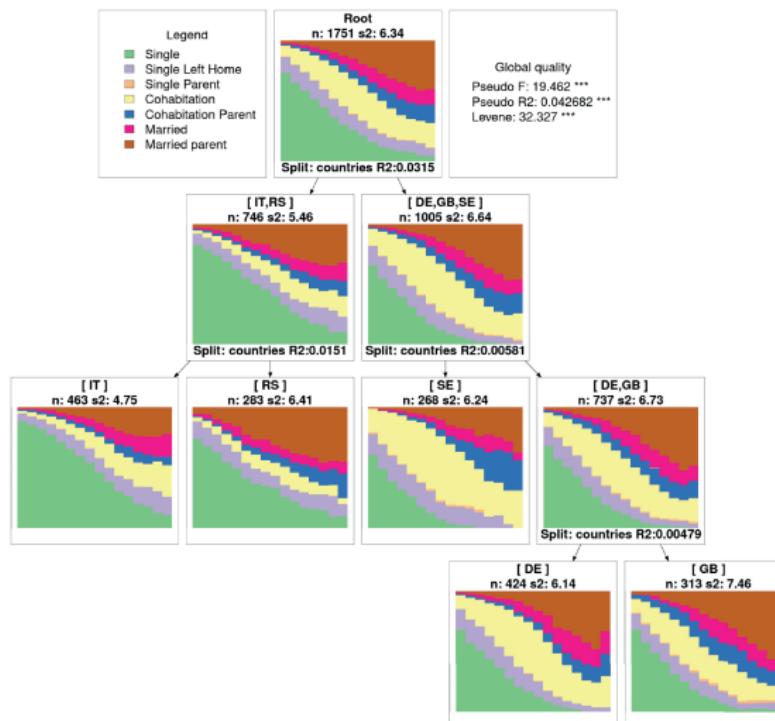
Sequence analysis

Ostro

Conclusions

References

References



# Event sequences

Sankey Diagram for Customer Journey, Warudkar (2020, page)

## Event sequence analysis

V. Batagelj &

Introduction

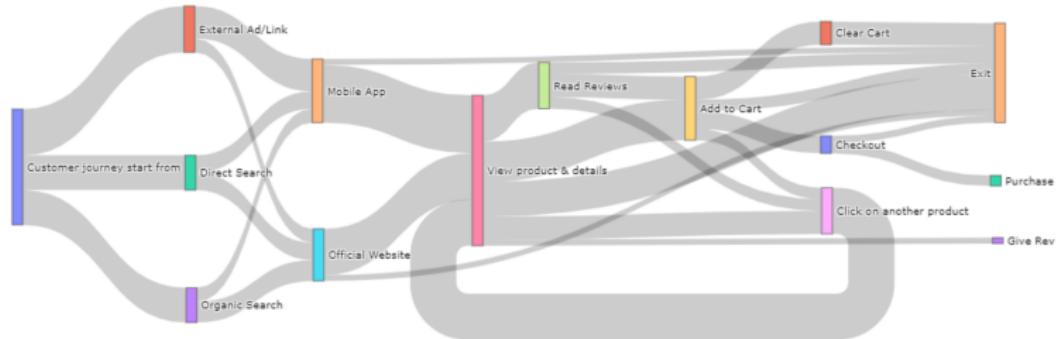
Sequence analysis

Ostro

Conclusions

References

References





# Oštro

## Problem

Event  
sequence  
analysis

V. Batagelj &

Introduction

Sequence  
analysis

Oštro

Conclusions

References

References

Oštro, Center for Investigative Journalism in the Adriatic Region compiled data on the career paths of all current public officials (prime minister, ministers, secretaries, and deputies in the National Assembly) from their CVs and then verified and supplemented it with the help of registers, specialized databases, media archives, independent investigative work, etc.

- ① who [person name] is related to whom [person name]? The longer the common time period [start\_day/start\_month\_start\_year - end\_day/end\_month\_end\_year] in the common institution [institution\_si], the stronger the connection
- ② how many different institutions [institution\_si] connect two individuals [name]
- ③ how much time [start\_day/start\_month\_start\_year - end\_day/end\_month\_end\_year] individuals [name] spent in political positions [affiliation type = politician]
- ④ number of management positions before and after political office [affiliation type = manager in a public company], [affiliation type = manager in a company], [position = state secretary], [position = minister], [position = Prime Minister]



# Oštro

## Problem

### Event sequence analysis

V. Batagelj &

Introduction

Sequence  
analysis

Oštro

Conclusions

References

References

- ⑤ the number of advisory positions [part of CV = advisory and supervisory functions etc.] before and after the political function [affiliation type = politician]
- ⑥ what is the last position [position\_si] before starting the political office [affiliation type = politician] and what is the first position [position\_si] after the end of the political office
- ⑦ party power - how "profitable" is an individual party [PERSONS\_LIVE tab: party\_si], in terms of what positions [position\_si] the functionaries move to after the end of the political function [affiliation type = politician]
- ⑧ the highest level of education achieved [part of CV = education] - is there a difference between MPs, secretaries and ministers [PERSONS LIVE: position tab]
- ⑨ gender [PERSONS LIVE tab: gender]
- ⑩ newbies [PERSONS LIVE tab: IS\_FIRST\_TIME\_IN\_OFFICE]

We will try to analyze the data following the research question: Do the former connections in different areas affect political connections among the current political office holders? What is the structure of the current set of Slovenian public officials?



# Oštro

## Raw data

Event  
sequence  
analysis

V. Batagelj &

Introduction

Sequence  
analysis

Oštro

Conclusions

References

References

The raw **Zvezoskop** data table was prepared in Excel for “human inspection”. It contains 3157 events. It had to be normalized for “computer processing”

- uppercase and lowercase initials
- extra spaces at the end of strings
- masculine/feminine form of words (poslanec / poslanka, etc.)
- mixed date format (dd/mm/yyyy / mm/dd/yyyy)
- incomplete dates (month and year only or year only)

Oštro event sequences dataset has some specifics: real-time events, parallel events in the same sequence, and different types of events.



# Oštro

## Raw data

### Event sequence analysis

V. Batagelj &

Introduction

Sequence  
analysis

Oštro

Conclusions

References

References

,person\_id, person\_name, part\_of\_cv, show\_in\_network, start\_day, start\_month,  
start\_year, institution\_si, institution\_standardized\_si, institution\_department\_si,  
affiliation\_type\_si, position\_si, end\_day, end\_month, end\_year, notes\_institution\_si, notes\_position\_si  
...  
2241,112,Matej Tonin,izobraževanje,True,,9,1998,gimnazija in srednja šola rudolfa maistra kamnik  
2242,112,Matej Tonin,izobraževanje,True,,10,2002,ul fakulteta za družbene vede,,,dodiplomski št  
2243,112,Matej Tonin,izobraževanje,True,,10,2007,ul fakulteta za družbene vede,,,magistrski štu  
2244,112,Matej Tonin,delovne izkušnje,True,16/4/2007,,2007,državni zbor.,,poslanska skupina NSi,j  
2245,112,Matej Tonin,strankarska pozicija,True,,,2006,nsi,,občinski odbor Kamnik,,predsednik,,,2  
2246,112,Matej Tonin,strankarska pozicija,True,10/9/2010,,2010,nsi,,,kanididat za župana v obč  
2247,112,Matej Tonin,delovne izkušnje,True,29/11/2006,,2006,občina kamnik,,politik,občinski sve  
2248,112,Matej Tonin,strankarska pozicija,True,,,2007,nsi,,Mlada Slovenija,,podpredsednik,,,2009  
2249,112,Matej Tonin,delovne izkušnje,True,3/11/2008,,2008,vizija agencija za komuniciranje mate  
2250,112,Matej Tonin,strankarska pozicija,True,,,2010,nsi,,,podpredsednik,,,2012,,  
2251,112,Matej Tonin,delovne izkušnje,False,21/12/2011,,2011,državni zbor,,politik,poslanec,13/  
2253,112,Matej Tonin,delovne izkušnje,True,22/6/2018,,2018,državni zbor,,politik,predsednik,23/  
2254,112,Matej Tonin,delovne izkušnje,False,13/5/2022,,2022,državni zbor,,politik,poslanec,,21  
2255,112,Matej Tonin,strankarska pozicija,True,22/8/2008,,2008,nsi,,,kandidat za poslanca,21/9/  
2256,112,Matej Tonin,delovne izkušnje,True,13/3/2020,,2020,ministrstvo za obrambo,,politik,minis  
2257,112,Matej Tonin,delovne izkušnje,True,13/3/2020,,2020,vlada,,politik,podpredsednik,1/6/202  
2258,112,Matej Tonin,strankarska pozicija,True,1/2/2018,,2018,nsi,,,v d predsednika ,20/4/2018,  
2259,112,Matej Tonin,strankarska pozicija,False,21/4/2018,,2018,nsi,,,predsednik,,2100,,  
2260,112,Matej Tonin,svetovalne in nadzorne funkcije etc.,True,,2015,institut dr janeza evangel  
2261,112,Matej Tonin,prostočasne aktivnosti,True,27/11/2006,,2006,športno društvo tuhinj,,  
2262,112,Matej Tonin,strankarska pozicija,False,,12,2001,nsi,,,član,,2100,,  
2263,112,Matej Tonin,strankarska pozicija,False,1/2/2018,,2018,nsi,,izvršilni odbor,,predsednik,  
...



The cleaned data table was first reduced to variables considered in the analysis.

- R – row (event) number (index)
- ID – person\_name: person's name
- s – start\_day|month|year: start date
- f – end\_day|month|year: finish date
- S – part\_of\_cv: type of activity
- T – institution\_si: location (institution, party, organization, etc.)



# Ostro

## data

### Event sequence analysis

V. Batagelj &

Introduction

Sequence  
analysis

Ostro

Conclusions

References

References

```
raw2events <- function(D,Tday="",CSV="events.csv"){
  if(Tday=="") Tday <- format(Sys.time(),"%d/%m/%Y")
  csv <- file(CSV,"w",encoding="UTF-8")
  cat("R ID s f S T\n",file=csv)
  n <- nrow(D); OK <- TRUE; k <- 0; td <- datum(Tday,0,0)
  for(i in 1:n){add <- TRUE
    ID <- trimws(D$person_name[i]); rel <- trimws(tolower(D$part_of_cv[i]))
    test <- trimws(tolower(D$institution_si[i]))
    ds <- D$start_day[i]; ms <- D$start_month[i]; ys <- D$start_year[i]
    sd <- datum(ds,ms,ys)
    if(sd>=td) add <- FALSE
    if(!OK) {cat(i,":",ID,ds,ms,ys,rel,'*** wrong date\n')
      flush.console(); OK <- TRUE; add <- FALSE}
    de <- D$end_day[i]; me <- D$end_month[i]; ye <- D$end_year[i]
    ed <- if(ye==2100) datum(Tday,NA,0) else datum(de,me,ye)
    ed <- min(ed,td)
    if(!OK) {cat(i,":",ID,de,me,ye,rel,'*** wrong date\n')
      flush.console(); OK <- TRUE; add <- FALSE}
    if(add) cat(i,'"',ID,'"',sd,'"',ed,'"',rel,'"',"",test,'"'\n',sep=' ',file=csv) else
      k <- k+1
  }
  close(csv)
  if(k>0) {cat(k," events skipped\n"); flush.console()}
}
```



# Oštro

## Data – co-presence

Event  
sequence  
analysis

V. Batagelj &

Introduction

Sequence  
analysis

Oštro

Conclusions

References

References

R	ID	s	f	S	T
64	Igor Papič	1973-09-01	1981-08-01	education	oš v kočevju
1116	Robert Golob	1979-09-01	1983-08-01	education	gim nova gorica
65	Igor Papič	1981-09-01	1985-06-01	education	vegova ljubljana
1068	Robert Golob	1983-10-01	1989-09-29	education	ul fe
66	Igor Papič	1986-10-01	1992-01-01	education	ul fe
1070	Robert Golob	1989-11-15	1994-11-14	work	ul fe
1069	Robert Golob	1992-10-01	1994-10-17	education	ul fe
67	Igor Papič	1992-10-01	1995-01-01	education	ul fe
1115	Robert Golob	1994-01-01	1994-01-01	education	tu georgia atlanta us
70	Igor Papič	1994-01-01	1996-01-01	work	siemens de
1071	Robert Golob	1994-11-15	1995-10-31	work	ul fe
69	Igor Papič	1993-03-01	1996-01-01	work	ul fe
68	Igor Papič	1995-10-01	1998-01-01	education	ul fe
1072	Robert Golob	1995-11-01	1997-09-30	work	ul fe
1073	Robert Golob	1997-10-01	2001-06-30	work	ul fe
1078	Robert Golob	1998-11-01	1999-08-31	work	min za gospodarstvo
71	Igor Papič	1999-01-01	2007-01-01	work	ul fe
96	Igor Papič	2000-01-01	2002-06-30	work	ul fe
94	Igor Papič	2001-01-01	2002-12-31	work	ul fe
...					

u	v	s	f	rel	d	Ru	Rv	
1	Robert Golob	Igor Papič	1986-10-01	1989-09-29	education	1094	1068	66
2	Robert Golob	Igor Papič	1993-03-01	1994-11-14	work	623	1070	69
3	Robert Golob	Igor Papič	1992-10-01	1994-10-17	education	746	1069	67
4	Igor Papič	Robert Golob	1994-11-15	1995-10-31	work	350	69	1071
5	Igor Papič	Robert Golob	1995-11-01	1996-01-01	work	61	69	1072
7	Robert Golob	Igor Papič	1999-01-01	2001-06-30	work	911	1073	71
10	Robert Golob	Igor Papič	2000-01-01	2001-06-30	work	546	1073	96
12	Robert Golob	Igor Papič	2001-01-01	2001-06-30	work	180	1073	94
...								

We decided to base our analysis on the corresponding *co-presence network* – a weighted multi-relational temporal network  $N = (V, L, w, t)$ .

The set of nodes  $V$  consists of persons from the data table.

There is a link (edge)  $\ell = (u : v; S)$  of type (relation)  $S$  between persons  $u, v \in V$  iff there exist events

$e_u = (ID = u, s = s_u, f = f_u, S = S_u, T = T_u)$  and

$e_v = (ID = v, s = s_v, f = f_v, S = S_v, T = T_v)$  such that

$S_u = S_v = S$  and  $T_u = T_v$  and  $[s, f] = [s_u, f_u] \cap [s_v, f_v] \neq \emptyset$ .

The weight of the link  $\ell \in L$  is  $w(\ell) = f - s$ , that is the length (in days) of the corresponding time interval  $[s, f]$ .

# Oštro

## Co-presence network: Mestna občina Nova Gorica

Event  
sequence  
analysis

V. Batagelj &

Introduction

Sequence  
analysis

Oštro

Conclusions

References

References

MO Nova Gorica

LEGENDA

točna datumna

točni in ocenjeni datum

ocenjena datumna

zadnja pravdopodobnost podatkov 1. 2. 2024



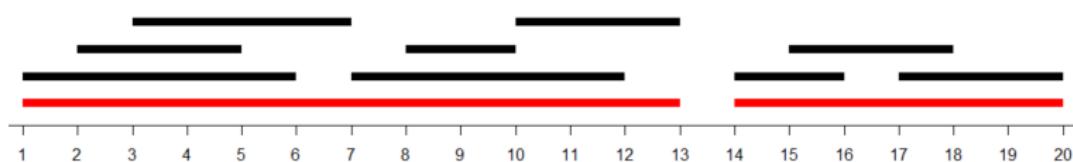


In preliminary experiments, we found that parallel events of the same kind can greatly increase the similarity between persons. One way to reduce this influence is to count parallel events of the same type between two persons only once.

We can do this by creating an associated set of time intervals for each quadruple  $(u, v, S, T)$  and computing its union, which we use to determine links. An efficient procedure is obtained if the data is ordered according to  $(u, v, S, T, s, f)$ .

The weight of a link (similarity of a pair of persons) is based on

- only pairs with different names are considered - the network has no loops
- in the results, the pairs are ordered ( $u < v$ )
- parallel connections of the same type are counted only once



All time intervals  $\{[s_i, f_i] : i \in 1 : k\}$  of edges between nodes  $u$  and  $v$  of kind  $S$  we replace with a single edge with weight

$$w(u, v; S) = |\bigcup_{i=1}^k [s_i, f_i]|$$



```
intUnion <- function(N){  
  k <- nrow(N); z <- rep(0,k); L <- data.frame(s=z,f=z)  
  i <- 1; r <- 0; s <- N$s[i]; f <- N$f[i]  
  while(i<k){i <- i+1  
    if(N$s[i]<=(t+1)){ if(N$f[i]>f) f <- N$f[i]  
      } else { r <- r+1; L[r,] <- list(s=s,f=f); s <- N$s[i]; f <- N$f[i] }  
    }  
  r <- r+1; L[r,] <- list(s=s,f=f)  
  return(L[1:r,])  
}  
  
traj2Pajek <- function(E,kMax,Net){  
  I <- order(E$s,E$f)  
  n <- length(I); k <- 0; r <- 0; sent <- "?????"  
  cn <- rep(0,kMax); cs <- rep("",kMax)  
  N <- data.frame(u=cs,v=cs,s=cn,f=cn,rel=cs,test=cs)  
  cat("% traj2Pajek",date(),"\nnevents",n,"\\n")  
  for(p in 1:(n-1)){i <- I[p]; tm <- E$f[i]  
    cat("."); if(p%%50==0) {cat(p,k,date(),"\\n"); flush.console()  
      for(q in (p+1):n){j <- I[q]; r <- r+1  
        if(E$s[j]>tm) break  
        uID <- E$ID[i]; vID <- E$ID[j]  
        if(uID!=vID) {  
          if(uID > vID) {vID <- E$ID[i]; uID <- E$ID[j]}  
          if(E$S[i]==E$S[j]) if(E$T[i]==E$T[j]) {  
            fm <- min(E$f[i],E$f[j]); sm <- max(E$s[i],E$s[j]); T <- fm-sm+1  
            if(T>0){k <- k+1; if(k>kMax) stop("kMax too small")  
              N[k,] <- list(u=uID,v=vID,s=sm,f=fm,rel=E$S[i],test=E$T[i])}  
          }  
        }  
      }  
    }  
  }  
  cat("\n",date(),"\\ndensity R =",2*r/n/(n-1)," tests =",r,  
  "\\ndensity E =",2*k/n/(n-1)," edges =",k,"\\n"); flush.console()
```



### Event sequence analysis

V. Batagelj &

Introduction

Sequence analysis

Ostro

Conclusions

References

References

```
N <- N[1:k,]; I <- order(N$u,N$v,N$rel,N$test,N$s,N$f); N <- N[I,]
N[k+1,] <- list(u=sent,v=sent,s=0,f=0,rel="",test="")
cn <- rep(0,k); cs <- rep("",k); r <- 0; q <- 0; k <- k+1
M <- data.frame(u=cs,v=cs,s=cn,f=cn,rel=cs,test=cs)
Q <- data.frame(s=cn,f=cn); i <- 1; H <- N[i,c("u","v","rel","test")]
for(p in 1:k){
  G <- N[p,c("u","v","rel","test")]
  if(all(H==G)) {q <- q+1; Q[q,] <- list(s=N$s[p],f=N$f[p])} else {
    L <- intUnion(Q[1:q,])
    for(h in 1:nrow(L)){ r <- r+1
      M[r,] <- list(u=H$u,v=H$v,s=L$s[h],f=L$f[h],rel=H$rel,test=H$test)
    }
    if(p==k) break
    H <- G; i <- p; q <- 0
  }
}
M <- M[1:r,]; sf <- as.matrix(M[,c("s","f")])
cat("density E' =",2*r/n/(n-1)," edges' =",r,"\\n"); flush.console()
uvrwt2net(M$u,M$v,w=M$f-M$s+1,r=M$rel,t=sf,directed=FALSE,Net=Net)
cat("% finished",date(),"\\n")
}

> D <- read.csv("./ostro_podatki.csv",sep=",",head=TRUE)
> raw2events(D,Tday="",CSV="Slo4all.csv")
> raw2events(D,Tday="24/4/2022",CSV="Slo4apr22.csv")
> Ea <- read.csv("Slo4all.csv",sep="")
> traj2Pajek(Ea,500000,"./Slo4all.net")
> Ee <- read.csv("Slo4apr22.csv",sep="")
> traj2Pajek(Ee,500000,"./Slo4apr22.net")
```



# Oštro

## Explanations

Event  
sequence  
analysis

V. Batagelj &

Introduction

Sequence  
analysis

Oštro

Conclusions

References

References

### R library `trajector.R`

```
> source("https://raw.githubusercontent.com/bavla/TQ/master/trajectories/traector4.R")
> E <- read.csv("Slo4.csv",sep="")
> p1 <- "Mojca Šetinc Pašek"; p2 <- "Tanja Fajon"
> (X <- explain(E,p1,p2))
      u           v           s           f           d       rel       test
1 Mojca Šetinc Pašek Tanja Fajon 1996-11-18 2001-09-03 1751     work    rtv slo
2 Mojca Šetinc Pašek Tanja Fajon 2001-10-01 2009-07-13 2843     work    rtv slo
3 Mojca Šetinc Pašek Tanja Fajon 1996-11-18 1998-05-03 532 education ul fdv
> p1 <- "Igor Papič"; p2 <- "Robert Golob"
> (X <- explain(E,p1,p2))
      u           v           s           f           d       rel       test
1 Igor Papič Robert Golob 1993-03-01 1996-01-01 1037     work ul fe
2 Igor Papič Robert Golob 1999-01-01 2021-01-01 8037     work ul fe
3 Igor Papič Robert Golob 1992-10-01 1994-10-17 747 education ul fe
```

# Oštro data

Event  
sequence  
analysis

V. Batagelj &

Introduction

Sequence  
analysis

Oštro

Conclusions

References

References

Two co-presence networks were constructed: before the elections on April 24, 2022, and till February 1, 2024.

relation	apr22	feb24
1 work	1767	5925
2 education	470	470
3 leisure activities	6	8
4 party	1086	2714
5 counseling & supervision	10	10
total	3339	9127
nodes	154	160

The link values in the network February 2024 are in the interval [1, 14344]. Applying the square root we transform them into the interval [1, 120]. Another application of the square root gives the interval [1, 11].

Monotonicity?!



We first combine all Pajek files related to a co-presence network into a Pajek project file. We also replace full names of persons with corresponding short names.

```
read network Slo4all.net
network Info button
Network/Multiple relations network/Info
Network/Create new network/Transform/Add/vertex labels/default [Yes]
Network/Create new network/Transform/Add/vertex labels/from file(s) [ostro_short.nam]
select network 1 Slo4all.net
File/Network/Dispose
select network 2 default
File/Network/Dispose
read clustering gender.clu
read clustering party.clu
read clustering position.clu
File/Network/Save [Slo4allShort.paj]
```

Procedures for automatically determining the network layout produce node placements that make sense, but are rather unreadable. The layout of the total co-presence network till April 2022 displayed on the next slide is created using the VOS Mapping procedure.

VOS april 2022

In the data on persons, I chose three partitions that describe the characteristics of an individual person: party (party membership), position (role in the country), and gender. We determine their use in Pajek by selecting them as current partitions:

1. Partition: party
2. Partition: position
3. Partition: gender

We request the display of the network with Draw/Network+First Partition. Initially, only party affiliation will be displayed as the node color. I chose the following bright colors



1: Undefined ; 2: Freedom (Svoboda) Movement; 3: Left; 4: NSi;  
5: IT+HU; 6: SD; 7: SDS.

Event  
sequence  
analysis

V. Batagelj &

Introduction

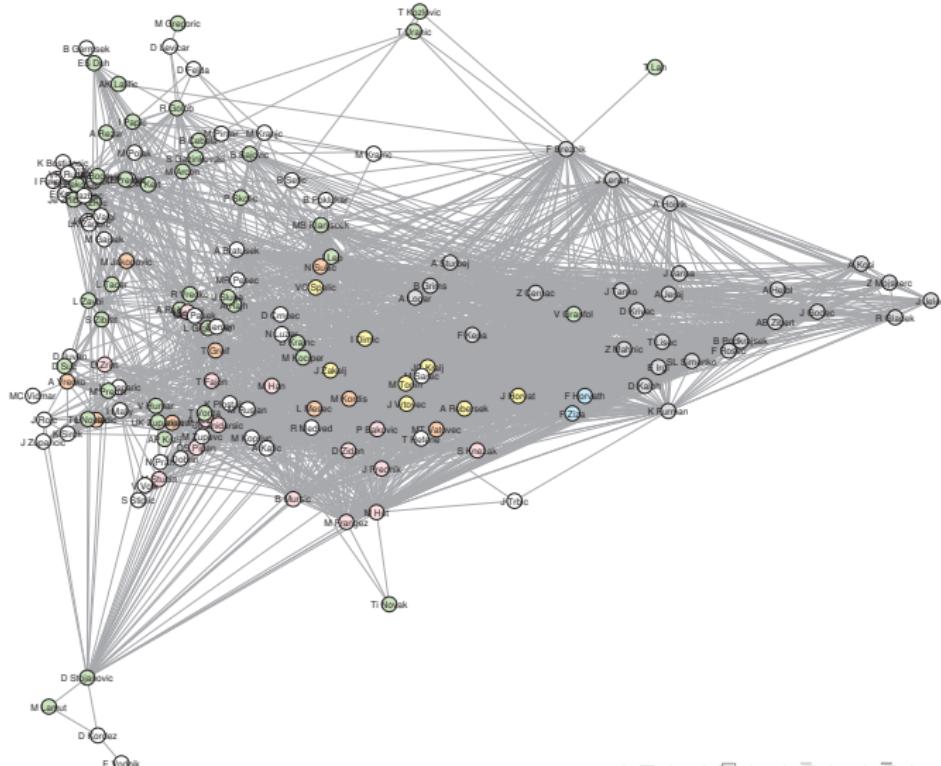
Sequence  
analysis

Oštro

Conclusions

References

References





*Link cutout* (pruning) at a threshold level  $t$  in a given network preserves only the strongest links whose weight is at least  $t$ . It also removes any isolated nodes. In Pajek, we get it like this:

```
Network/Create new network/Transform/Remove/Multiple lines/Sum values [No]
Network/Create new network/Transform/Remove/Lines with Value/lower than [t, OK]
Network/Create Partition/Degree/All
Operations/Network+Partition/Extract/Subnetwork ... [1-*, OK]
```

If we want to show some of the partitions in the picture, we have to adjust it to the obtained cutout

```
select the selected partition as the First partition
select Degree All partition as the Second partition
Partitions/Extract Subpartition
```

Cutout at level 5000 on whole April 2022 network

Cutout at level 3000 on work April 2022 network

# Oštro

Cutout at level 5000 on whole April 2022 network

Event  
sequence  
analysis

V. Batagelj &

Introduction

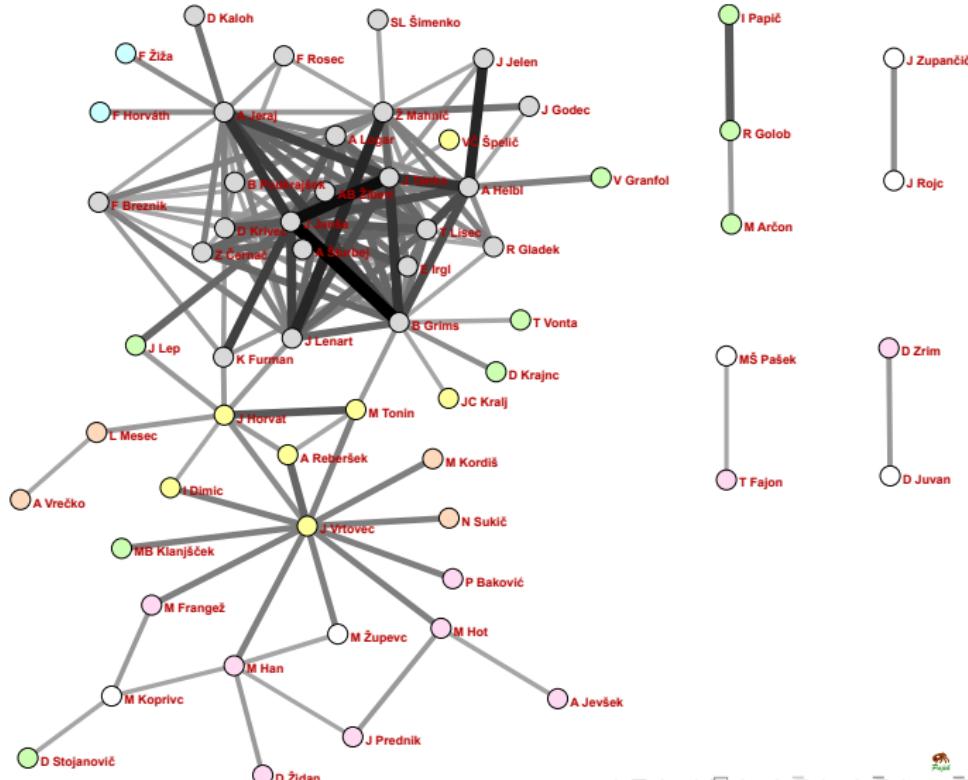
Sequence  
analysis

Oštro

Conclusions

References

References





For a quick inspection of the structure of weighted networks, we use various skeletons, in which less important elements (nodes, links) are discarded. An example of a skeleton is a *k-neighbors* network, in which we keep only the  $k$  most important neighbors (highest weights) in each node.

First, we create a simple temporal network from the basic multi-relational temporal network: if we keep all/selected relations, we use

```
Network/Multiple relations network/Change Relation Number-Label [1-*,1,All together,OK] [Yes]
```

but if we only want the  $r$ -th relation, we can use

```
Network/Multiple relations network/Extract Relation(s) into Separate Network(s) [r,OK] [Yes]
```

The *k*-neighbors procedure goes like this in Pajek:

```
Network/Create new network/Transform/Remove/Multiple lines/Sum values [No]  
Network/Create new network/Transform/Line values/Abs+sqrt
```

```
File/Network/Change label [Sqrt]
```

```
Network/Create new network/Transform/Edges -> arcs [yes]
```

```
File/Network/Change label [Sqrt directed]
```

```
Network/Create new network/Transform/Remove/all arcs ... except/k with highest ...[k]
```

```
Network/Create partition/Components/Weak [2]
```



Finally, we draw the network and export the layout to the selected format

```
Draw/Network+First partition
Layout/Energy/Kamada-Kawai/Separate components
manually improve the picture
set Options and Export/Options
Export/2D/SVG/General [sosedsi.svg]
```

1-neighbors for party links February 2024

Whole network for party links February 2024 on the 1-neighbors nodes layout

We display the second partition with the selected symbols with Options/Mark vertices using/Cluster symbols of second partition. The selected symbols are defined with Options/Symbols for partition clusters/Change. I chose the following symbols

 : Member of Parliament

 : Secretary

 : Minister

 : Prime Minister

The third partition will be represented with the color of the symbol. We enable its display with Options/Colors/Use third partition for symbol color. I chose (Options/Colors/Partition colors/for Symbols) the usual colors - red: female, blue: male.

Event  
sequence  
analysis

V. Batagelj &

Introduction

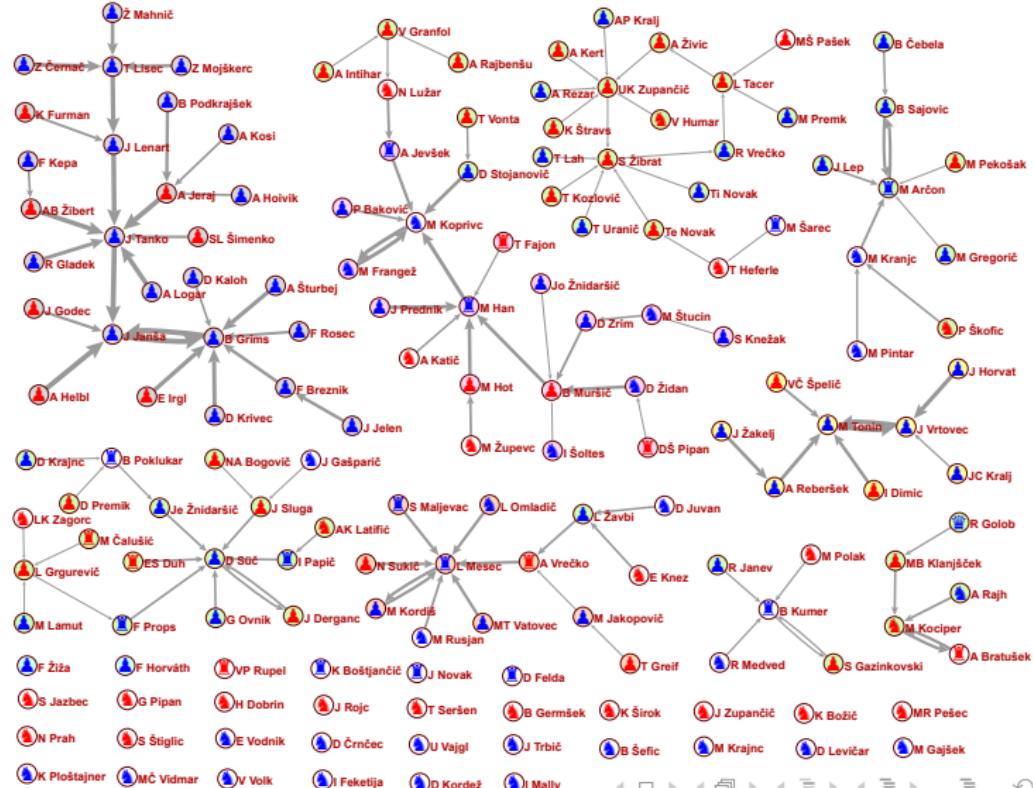
Sequence  
analysis

Oštro

Conclusions

References

References



Event  
sequence  
analysis

V. Batagelj &

Introduction

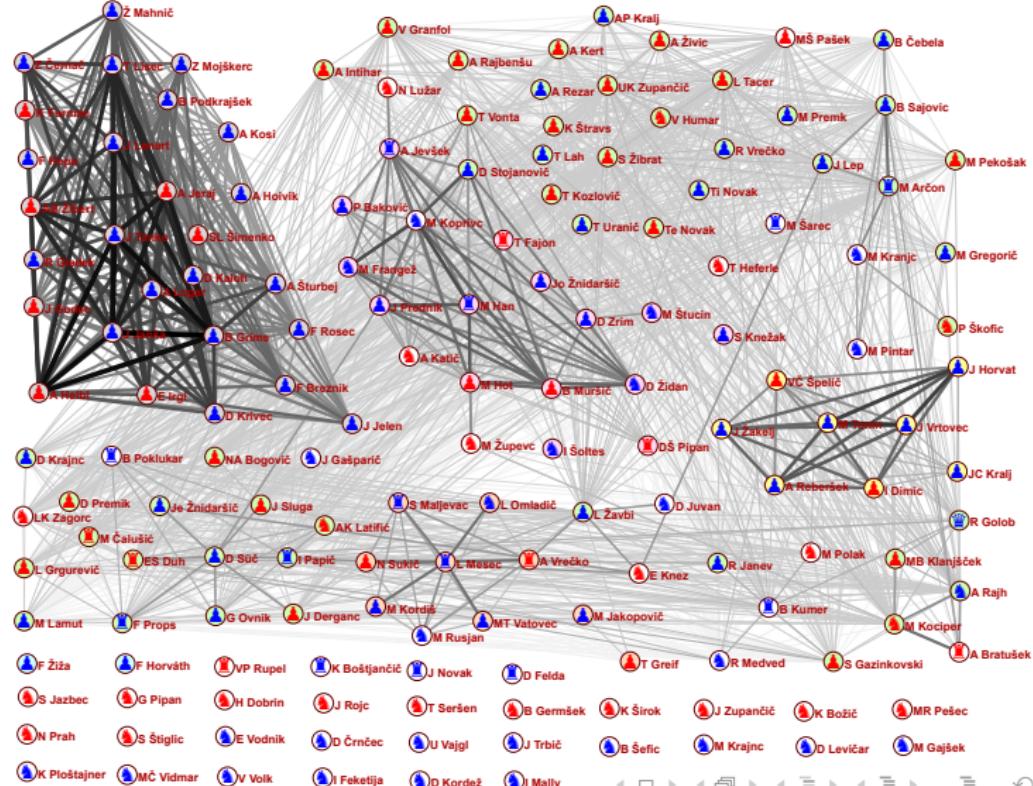
Sequence  
analysis

Oštro

Conclusions

References

References





In larger (more than 10 nodes) dense (higher average degree) networks, its display with a graph is very unreadable. For networks of size up to some thousands, a matrix display is a more suitable solution. In a matrix display, the order of the nodes is very important – a good order will create patterns in the display that reveal the structure of the network.

Let the current network be a simplified network (a simple network without parallel links). Then we can determine an ordering of the nodes by clustering them. Since the range of weight values is very large, we reduce it using a monotonic mapping (eg  $\ln(w)$  or  $\sqrt{\sqrt{w}}$  ). In our case, we will use double square rooting

```
Network/Create new network/Transform/Line values/Abs + Sqrt
Network/Create new network/Transform/Line values/Abs + Sqrt
Cluster/Create complete cluster [160] [OK]
Operations/Network+Cluster/Dissimilarity*/Network based/d5-Corrected Euclidean [1] [OK]
select network Sqrt(Sqrt)
File/Network/Export as Matrix to/EPS/using permutation
```



Event  
sequence  
analysis

V. Batagelj &

Introduction

Sequence  
analysis

Oštro

Conclusions

References

References

We open the resulting files `dendro.eps` and `matrix.eps` in Acrobat reader and save them in PDF format. We then open the matrix display in Inkscape and save it in SVG format.

In Pajek, individual clusters can be rearranged (change the order of individual branches in the tree) and get an improved display. We can also specify clusters and get the corresponding partition.

Matrix display whole April 2022 network/ $\sqrt{w}$ ; Dendro PDF; Matrix PDF

Matrix display whole February 2024 network/ $\sqrt{w}$ ; Dendro PDF; Matrix PDF

Zvezoskop analysis

## Oštro

Dendrogram April 2022

## Event sequence analysis

V. Batagelj &

Introduction

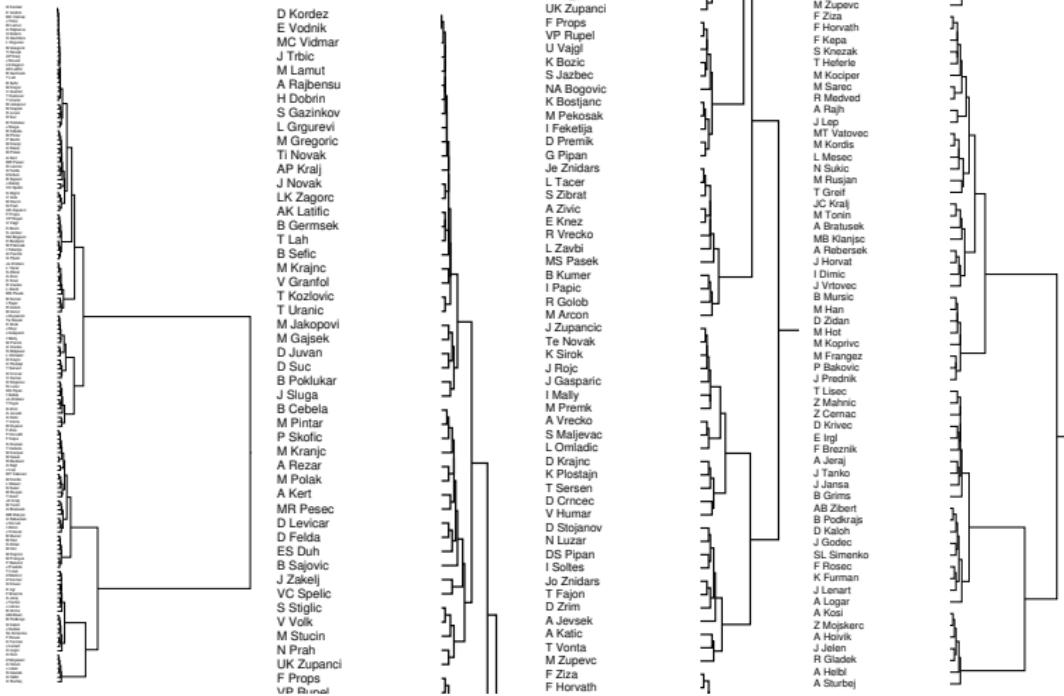
## Sequence analysis

Ostro

### Conclusions

## References

### References



Event  
sequence  
analysis

V. Batagelj &

Introduction

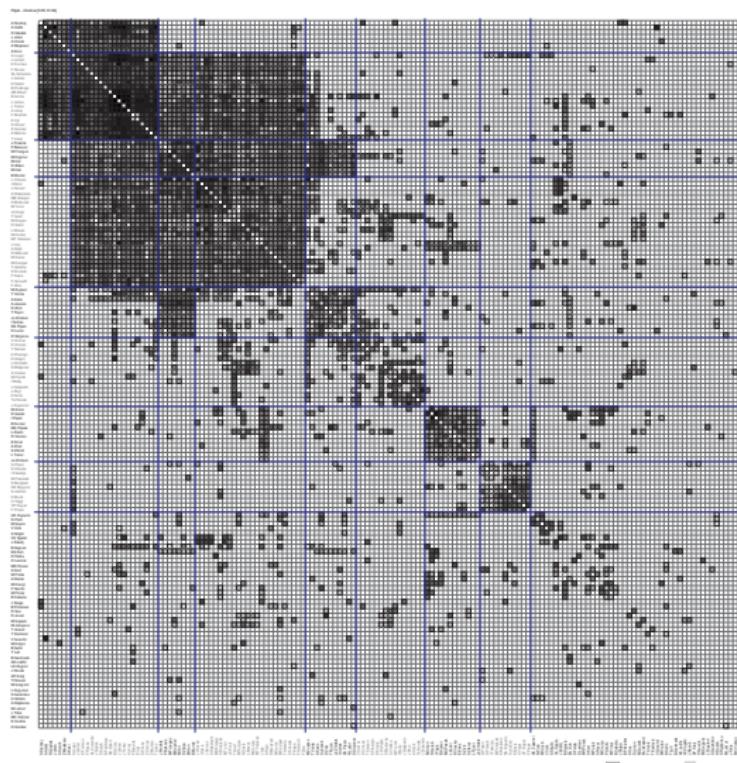
Sequence  
analysis

Oštro

Conclusions

References

References



Event  
sequence  
analysis

V. Batagelj &

Introduction

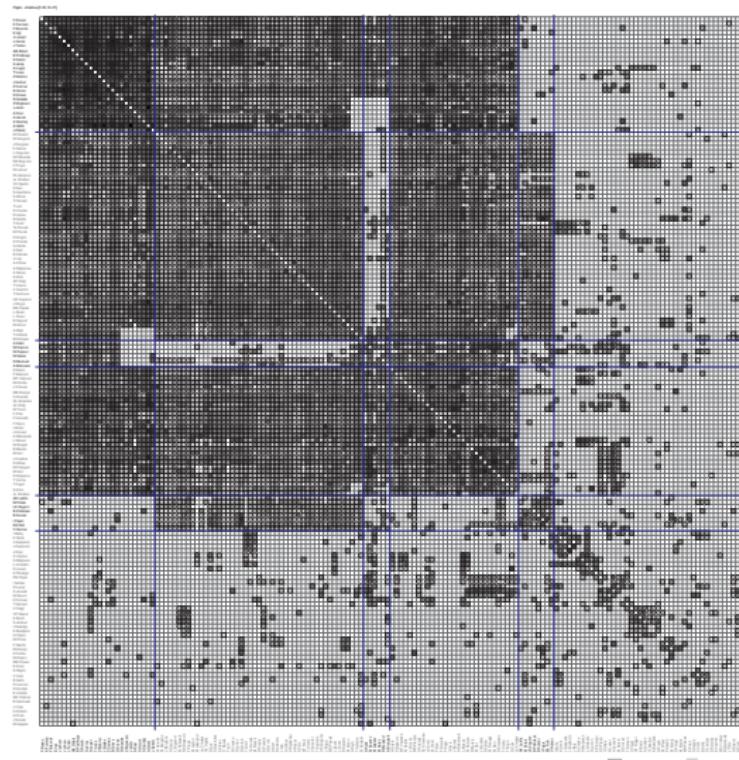
Sequence  
analysis

Oštro

Conclusions

References

References





# Oštro

## Blockmodel

Event  
sequence  
analysis

V. Batagelj &

Introduction

Sequence  
analysis

Oštro

Conclusions

References

References

For a given clustering (partition), we can obtain the corresponding partition contraction matrix

```
select SUM network as the First network
select the clustering partition as the First partition
Operations/Network+Partition/Shrink network
select the clustering partition as the First partition
Partition/Count
```

For the April 2022 network we get (click on the relevant register) the matrix

	size	1	2	3	4	5	6	7	8	9	Label
1.	47	51934	23424	72685	14499	36016	45941	13235	9347	9787	#AK Latifić
2.	11	23424	34047	36714	0	42487	10418	26731	106358	1442	#A Jevšek
3.	24	72685	36714	334267	5562	522702	34194	58636	192011	5373	#A Reberšek
4.	7	14499	0	5562	43790	348917	0	5593	458	0	#A Helbl
5.	19	36016	42487	522702	348917	927203	6737	22678	121872	7797	#A Jeraj
6.	12	45941	10418	34194	0	6737	55999	4821	31	8993	#A Živic
7.	15	13235	26731	58636	5593	22678	4821	59184	4508	828	#A Vrečko
8.	8	9347	106358	192011	458	121872	31	4508	94293	1494	#B Muršič
9.	11	9787	1442	5373	0	7797	8993	828	1494	36673	#D Premik

In a text editor we put the obtained data into a table and read it into R.

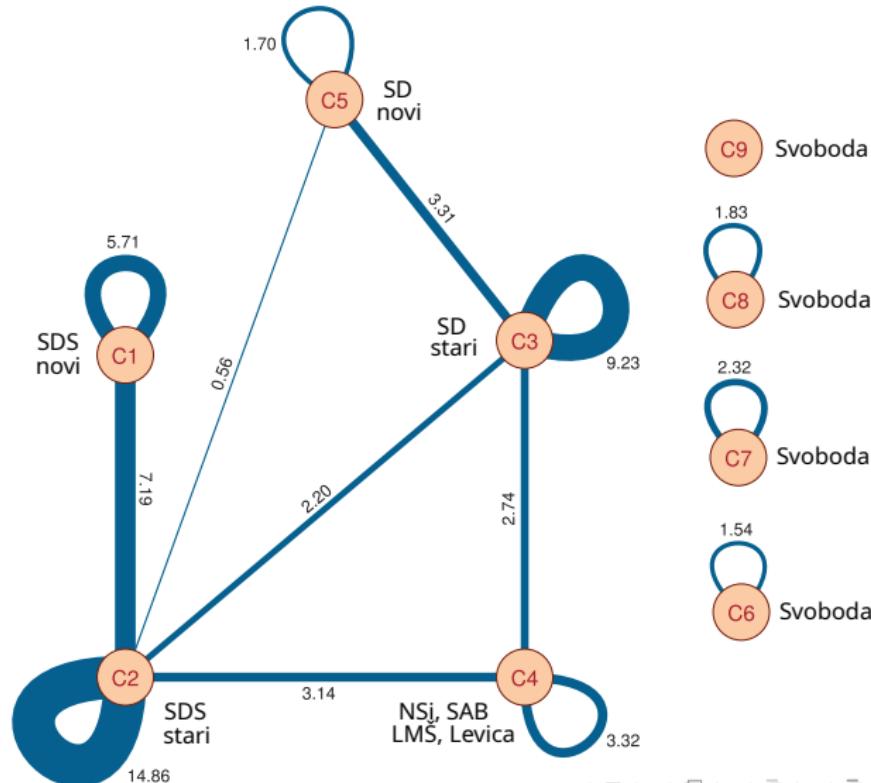
We still need to normalize the values in the matrix  $M$ . For  $i \neq j$  the normalized weight is  $Q[i, j] = \frac{M[i, j]}{n[i] \cdot n[j]}$  and for  $i = j$  it is  $Q[i, i] = \frac{M[i, i]}{n[i] \cdot (n[j] - 1)}$  (for a directed network; or its double for an undirected network).

```
S <- read.table("./pics/shrink22.csv", head=TRUE)
M <- as.matrix(S[, 1:9])
B <- matrix(0, nrow=9, ncol=9)
for(i in 1:9) for(j in 1:9) B[i, j] <- M[i, j] / S$n[i] / (S$n[j] - (i == j))
diag(B) <- 2 * diag(B)
matrix2net(B, Net="BM22.net")
```

The resulting normalized matrix is exported as a Pajek network and displayed in Pajek.

PS1. During normalization, we could also divide by the number of links in the block – we get them by first setting all weights to 1 in the network and then shrinking them.

PS2. Full normalization can be done in Pajek (my *Introduction to Network Analysis* slides).





# Conclusions

Event  
sequence  
analysis

V. Batagelj &

Introduction

Sequence  
analysis

Ostro

Conclusions

References

References

- ① Apply traditional sequence analysis to Zvezoskop data.  
Extend the dataset at least to public officials from the previous government.
- ② Networks and sequence analysis [23], [24], [25], [26], [27].
- ③ Find interesting and well documented datasets.
- ④ The Sequence Analysis Association (SAA).



# Acknowledgments

Event  
sequence  
analysis

V. Batagelj &

Introduction

Sequence  
analysis

Oštros

Conclusions

References

References

The computational work reported in this paper was performed using a collection of R functions `trajectoR` and the program **Pajek** for analysis of large networks. The code and data are available at [Github/Bavla/trajectories](#).

This work is supported in part by the Slovenian Research Agency (research program P1-0294, research program CogniCom (0013103) at the University of Primorska, and research projects J5-2557, J1-2481, and J5-4596), and prepared within the framework of the COST action CA21163 (HiTEc).

YouTube/Oštros: [video 1](#), [video 2](#).



# References |

Event  
sequence  
analysis

V. Batagelj &

Introduction

Sequence  
analysis

Ostro

Conclusions

References

References

1. Gabadinho, A. et al. Package '*TraMineR*'. 2024.
2. Le Roux, G., Studer, M., Bringé, A. & Bonvalet, C. Selecting qualitative cases using sequence analysis: A mixed-method for in-depth understanding of life course trajectories. *Advances in Life Course Research* **56**, 100530 (2023).
3. Studer, M. Validating sequence analysis typologies using parametric bootstrap. *Sociological Methodology* **51**, 290–318 (2021).
4. Studer, M., Ritschard, G., Gabadinho, A. & Müller, N. S. Discrepancy analysis of state sequences. *Sociological methods & research* **40**, 471–510 (2011).
5. Batagelj, V. Generalized Ward and related clustering problems. *Classification and related methods of data analysis* **1** (1988).
6. Liao, T. F. et al. Sequence analysis: Its past, present, and future. *Social science research* **107**, 102772 (2022).
7. Yeshchenko, A. & Mendling, J. A survey of approaches for event sequence analysis and visualization. *Information Systems*, 102283 (2023).

# References II

8. Van der Linden, S., de Fouw, E., van den Elzen, S. & Vilanova, A. A survey of visualization techniques for comparing event sequences. *Computers & Graphics* (2023).
9. Peiris, Y., Barth, C.-M., Huang, E. M. & Bernard, J. *A Data-Centric Methodology and Task Typology for Time-Stamped Event Sequences*. in *2022 IEEE Evaluation and Beyond-Methodological Approaches for Visualization (BELIV)* (2022), 66–76.
10. Han, S.-K. in *Advances in sequence analysis: Theory, method, applications* 21–38 (Springer, 2014).
11. Vrotsou, K., Ellegård, K. & Cooper, M. Exploring time diaries using semi-automated activity pattern extraction. *electronic International Journal of Time Use Research (eIJTUR)* 6, 1–25 (2009).
12. Hamming, R. W. Error detecting and error correcting codes. *The Bell system technical journal* 29, 147–160 (1950).
13. Levenshtein, V. I. *Binary codes capable of correcting deletions, insertions, and reversals*. in *Soviet physics doklady* 10. (translated from Doklady Akademii Nauk SSSR, 163(1965)4, 845–848) (1966), 707–710.



# References III

Event  
sequence  
analysis

V. Batagelj &

Introduction

Sequence  
analysis

Ostro

Conclusions

References

References

14. Sankoff, D. & Kruskal, J. *Time Warps, String Edits, and Macromolecules: The Theory and Practice of Sequence Comparison*. (Chicago: CSLI publications, 1983).
15. Abbott, A. Sequences of social events: Concepts and methods for the analysis of order in social processes. *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 16, 129–147 (1983).
16. Abbott, A. & Forrest, J. Optimal matching methods for historical sequences. *The Journal of Interdisciplinary History* 16, 471–494 (1986).
17. Abbott, A. & Tsay, A. Sequence analysis and optimal matching methods in sociology: Review and prospect. *Sociological methods & research* 29, 3–33 (2000).
18. Fasang, A. & Struffolino, E. *Sequence Analysis for Social Science*. <https://pdhp.isr.umich.edu/wp-content/uploads/2022/02/PDHP-Sequence-Analysis-Slides.pdf>. slides, Accessed: (04-Mar-24). 2022.
19. Robette, N. *Tutorial on sequence analysis*. <https://cran.r-project.org/web/packages/seqhandbook/vignettes/Tutorial.html>. Accessed: (04-Mar-24). 2023.



# References IV

Event  
sequence  
analysis

V. Batagelj &

Introduction

Sequence  
analysis

Ostro

Conclusions

References

References

20. Studer, M. & Ritschard, G. What matters in differences between life trajectories: A comparative review of sequence dissimilarity measures. *Journal of the Royal Statistical Society Series A: Statistics in Society* **179**, 481–511 (2016).
21. Yanni, P., Xiaoping, F., Ying, Z. & Fangfang, Z. A survey of visualization of temporal event sequence data. *Journal of Computer-Aided Design & Computer Graphics* **31**, 1698–1710 (2019).
22. Zhou, Y. Work trajectories and status attainment process: a study using sequence analysis. *The Journal of Chinese Sociology* **10**, 1 (2023).
23. Mattsson, C. E. & Takes, F. W. Trajectories through temporal networks. *Applied Network Science* **6**, 35 (2021).
24. Heffernan, T. Academic networks and career trajectory: There's no career in academia without networks. *Higher Education Research & Development* **40**, 981–994 (2021).
25. Wu, K., Tang, J. & Zhang, C. *Where Have You Been? Inferring Career Trajectory from Academic Social Network.* in *IJCAI* (2018), 3592–3598.



# References V

Event  
sequence  
analysis

V. Batagelj &

Introduction

Sequence  
analysis

Ostro

Conclusions

References

References

26. Hamberger, K. Relational sequence networks as a tool for studying gendered mobility patterns. *Sequence Analysis and Related Approaches: Innovative Methods and Applications*, 121–146 (2018).
27. Bison, I. in *Advances in sequence analysis: Theory, method, applications* 231–248 (Springer, 2014).