Sparse potpensity method

Lavoslav Čaklović
Faculty of Natural Sciences, Dept. of Math.

Biostat 2019 – Zagreb, 5–8 june

Sadržaj

The logic of causation

To formalize the theory we should:

- setup the meaning of notions like:

events, cause, effect, similar world. . .

- setup the meaning of 'conclusion'

$$A \textbf{ causes } B \ (A \rightsquigarrow B)$$

and the **negation of an event**.

- setup the rules of assigning the truth value to hypothetical sentences. For instance:

Had Franz Ferdinand not been shot, WW1 would not have occurred.

Is it truth or not? Is it true in this world or in hypothetical world?

Hume vs. Lewis

Hume:[1]

> . . . *what one does have is the constant conjunction of cause C and effect E and the expectation that E will follow C.*

May be more formally:

> . . . we may define the relation of cause and effect such that *where, if the first object had not been, the second never had existed.*

\- - - - - -

Lewis (1973) — counterfactual approach:

[1] *A Treatise of Human Nature & An Enquiry Concerning Human Understanding.*

Counterfactual (Lewis, 1973)

A is the cause of B ($A \rightsquigarrow B$) if and only if:

(1) $A \to B$ (if A were to occur B would occur) – A implies B.

(2) $A \;\Box\!\!\rightarrow B$ (if B were not to occur A would not occur)

Counterfactual (Lewis, 1973)

A is the cause of B ($A \rightsquigarrow B$) if and only if:
(1) $A \to B$ (if A were to occur B would occur) – A implies B.
(2) $A \;\Box\!\!\to B$ (if B were not to occur A would not occur)

It may be shown that (1) and (2) are necessary conditions for causality, but not sufficient.
(3) $B \;\Box\!\!\to A$ (if A were not to occur B would not occur)

Neyman (1923), Quine (1960), Mill (1843) are also speaking of counterfactuals.

Counterfactual (Lewis, 1973)

A is the cause of B ($A \rightsquigarrow B$) if and only if:
(1) $A \to B$ (if A were to occur B would occur) – A implies B.
(2) $A \;\Box\!\!\to B$ (if B were not to occur A would not occur)

It may be shown that (1) and (2) are necessary conditions for causality, but not sufficient.
(3) $B \;\Box\!\!\to A$ (if A were not to occur B would not occur)

Neyman (1923), Quine (1960), Mill (1843) are also speaking of counterfactuals.
Weakness of Lewis approach:
– his causal relation is symmetric
– early preemption
– late preemption
– trumping
A possible correction: *causal chain*.

Influence (Lewis, 2000)

In his new theory Lewis is talking about *influence* instead of *causality* and introduces the *chain of influence* fro A to B.

Still, there is a problem with backward transitivity.

The principle of individual choice. In the causal history of an event we choose an event as the cause which offers a *reasonable explanation of the causal chain*. 'Reasonable' is context dependent[2]

Explanation is 'epistemic notion', causality is 'metaphysical relation'.

[2] Because of that we are speaking about individual choice.

Influence (Lewis, 2000)

In his new theory Lewis is talking about *influence* instead of *causality* and introduces the *chain of influence* fro A to B.

Still, there is a problem with backward transitivity.

The principle of individual choice. In the causal history of an event we choose an event as the cause which offers a *reasonable explanation of the causal chain*. 'Reasonable' is context dependent[2]

Explanation is 'epistemic notion', causality is 'metaphysical relation'.

Are we are coming back to Hume??!!

[2]Because of that we are speaking about individual choice.

Neyman, 1923. Statistical answer to logical problem.

A – a finite set of entities (population).
T – a treatment with measurable effect Y.
$\bar{Y}$ – a parameter of distribution Y (usualy $E(Y)$)
C – another treatment (control).

$\bar{Y}_{A,T}$ – expectation $E(Y|T)$.

Neyman, 1923. Statistical answer to logical problem.

A – a finite set of entities (population).
T – a treatment with measurable effect Y.
$\bar{Y}$ – a parameter of distribution Y (usualy $E(Y)$)
C – another treatment (control).

$\bar{Y}_{A,T}$ – expectation $E(Y|T)$.
$\bar{Y}_{A,C}$ – expectation $E(Y|C)$ (*counterfactual world*)

Neyman, 1923. Statistical answer to logical problem.

A – a finite set of entities (population).
T – a treatment with measurable effect Y.
$\bar{Y}$ – a parameter of distribution Y (usualy $E(Y)$)
C – another treatment (control).

$\bar{Y}_{A,T}$ – expectation $E(Y|T)$.
$\bar{Y}_{A,C}$ – expectation $E(Y|C)$ (*counterfactual world*)

Definition. *Causal effect of T with respect to C* is the difference

$$\tau = \bar{Y}_{A,T} - \bar{Y}_{A,C}.$$

We are reading $\bar{Y}_{A,T}$ (real world), not $\bar{Y}_{A,C}$ (imaginal world). How to manage such missing data situation?

Neyman – A replacement for counterfactual world?

A – the population exposed to the treatment T and $\bar{Y}_{A,C}$ is not measurable.

B[3] – another population exposed to the (control) C. Let us consider $\bar{Y}_{A,T} - \bar{Y}_{B,C}$.
Ideal situation: $\forall i \in A$ there is a twin $i' \in A$. The first is treated, the second is controlled, and we have a difference

$$Y(i|T) - Y(i'|C),\ i, i' \text{ tweens.}$$

If we 'forget' the twin i' we think of two *potential values* $\{Y_{iT}, Y_{iC}\}$ of which only one is 'readable' depending if the entity is treated or controlled. T is then the indicator for treated group. Let us define $Y_i := (1 - T_i)Y_{iC} + T_i Y_{iT}$, and

$$\tau := E(Y_i|T_i = 1) - E(Y_i|T_i = 0).$$

[3] Should be 'close' to the real world.

Randomization (Rubin)

$$\tau := E(Y_i | T_i = 1) - E(Y_i | T_i = 0). \quad (1)$$

Formula (1) is meaningful if the *group affiliation* is independent of Y, i.e. $E(Y_i | T) = E(Y_i)$. Then

$$\tau := E(Y_{iT}) - E(Y_{iC}).$$

\- - - - - -

Independence may be guaranteed ba random[4] choice of treated units.

See also: v. Russo, Wunsch, Mouchart, *Inferring Causality through Counterfactuals in Observational Studies, Some epistemological issues (2010)*

[4]Which is not allways possible

Additional constraints

- *Stable Unit Treatment Value Assumption (SUTVA).*
- (A) $\{Y_{iT}, Y_{iC}\} \perp\!\!\!\perp T | X$ i (B) $0 < Pr(T = 1 | X) < 1$, for each covariate X (A–independence, B–overlapping)
- balance

  $$\tau | (T = 1) = E[(E(Y_{iT} | X, T = 1) - E(Y_{iC} | X, T = 0)) | T = 1],$$

  where the outer expectation is taken over the restriction $X | T == 1$.

Matching, i.e. looking for twins may be done by NN, Mahalanobis distance or by using the distance on some scale like *propensity scale*.

A recent paper: *Why Propensity Scores Should Not Be Used for Matching* (Gary King-Nielsen-November, 2018)

Potential method: Working plan

Observational data:

item id	T or C	covariates X	Y
i	T	...	Y_i
j	C	...	Y_j
...	...	...	...
n	C	...	Y_n

much more C's than T's

A rough procedure...

1. $\forall i$ (treated) find a *twin* i' (controlled) and observe the difference $\tau_i = Y_i - Y_{i'}$. What is the definition of *twin*?
   - If X's have the same values for i and i' — we have a *twin*.
   - If not, find a set of *proxy twins*.
2. Construct some scale on the set of *treated + proxy twins*.
3. From the distance matrix create optimal matching.
4. Calculate mean effect.

Working plan – details

- Some factors (covariates) generate a stratified population.
- The *proxy twins* should be in the same strata.
- We will use *potential* as the scale. Usually it is generated by generalized (logistic) regression.
  The good side of potential is that it allows missing data – sparse covariates values.
  The difficulty is that PM forces the user to specify the trade off between the covariate units. This may be avoided by standardization of data, normalizing each column to the same *flownorm*[5].
- Now we have potential scale on each strata.
- The final step is to use the Hungarian method for matching.

[5] Analogy with dividing by SD.

An example: lalonde data

```
# A tibble: 16,289 x 8
item  treated   age education married nodegree   ejump etnic
<chr> <fct>   <dbl>     <dbl> <fct>   <fct>      <dbl> <fct>
 E1       1      37        11     1       1       9930.   B
 E2       1      22         9     0       1       3596.   H
 E3       1      30        12     0       0      24909.   B
 E4       1      27        11     0       1       7506.   B
 E7       1      23        12     0       0          0    B
 E8       1      32        11     0       1       8472.   B
 E9       1      22        16     0       0       2164.   B
E10       1      33        12     1       0      12418.   O
# ... with 16,279 more rows

treated     n
<fct>     <int>
    0     15992
    1       297
```

Hungarian matching (strata="10B")

```
get_strata_data(data, strata=c("1", "0", "B"))

# A tibble: 16 x 9
item  treated   age  edu   marr nodeg  ejump  etnic      X
<chr> <fct>   <dbl> <dbl> <fct> <fct>  <dbl> <fct>  <dbl>
E1450 0          23   12  1         0  3210.      B -0.886
E8055 0          42   14  1         0  1261.      B  4.11
E1893 0          23   12  1         0   331.      B -0.886
E862  0          27   12  1         0 -2274.      B -0.360
E1965 0          26   12  1         0   -83.6     B -0.492
E1701 0          27   12  1         0 -2702.      B -0.360
E40   1          23   12  1         0  5912.      B -0.886
E61   1          42   14  1         0 13168.      B  4.11
E183  1          23   12  1         0 -4796.      B -0.886
E239  1          27   12  1         0 -5029.      B -0.360
E271  1          26   12  1         0 -4370.      B -0.492
E283  1          27   12  1         0  -334.      B -0.360
```

Effect by strata

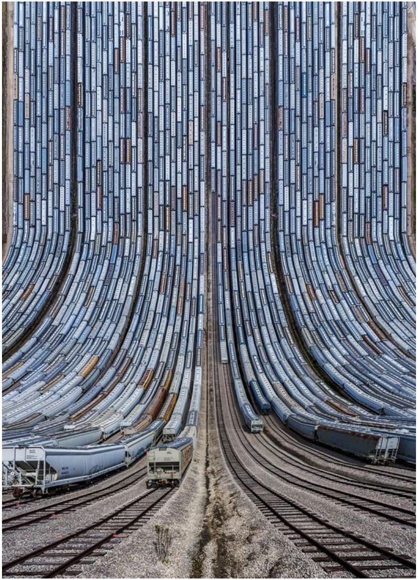
```
Size by strata:          Effect by strata:                Overall effect:
---------------          --------------------             ----------------
1  00B       203         eff(00B) = 2382.36               513.2903.
2  00H        22         eff(00H) = 5168.931
3  000       376         eff(000) = -56.78306
4  01B       276         eff(01B) = -163.8093
5  01H        70         eff(01H) = -2023.542
6  010       250         eff(010) = -570.09
7  10B        57         eff(10B) = 2644.577
8  10H        14         eff(10H) = 3927.748
9  100       236         eff(100) = -4383.368
10  11B       121        eff(11B) = 2794.743
11  11H         8        eff(11H) = -8109.565
12  110        39        eff(110) = 1507.999
```

Olmos & Govindasamy, Propensity Scores: A Practical Introduction Using R (2015), propensity score (effect): 326.3214

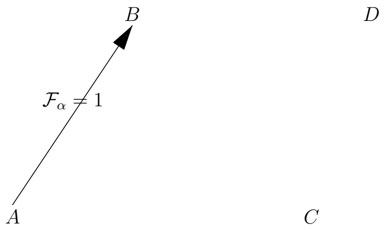
The final effect

We are interpreting the effect of the cause,
not he cause of the effect.

Reality or perception?



Potential Method[6]



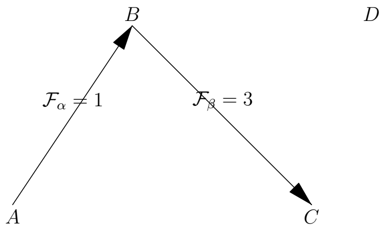
Preference flow $\mathcal{F}$

Incidence matrix $A \in \mathbb{R}^{m \times n}$

	nodes$_n$				flow
arcs$_m$	A	B	C	D	$\mathcal{F}$
α	-1	1	0	0	1

[6]Čaklović (2012); Čaklović and Kurdija (2017)

Potential Method[6]



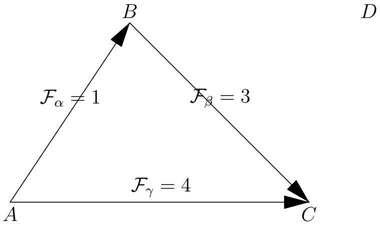
Preference flow $\mathcal{F}$

Incidence matrix $A \in \mathbb{R}^{m \times n}$

	nodes$_n$				flow
arcs$_m$	A	B	C	D	$\mathcal{F}$
α	-1	1	0	0	1
β	0	-1	1	0	3

[6] Čaklović (2012); Čaklović and Kurdija (2017)

Potential Method[6]



Preference flow $\mathcal{F}$

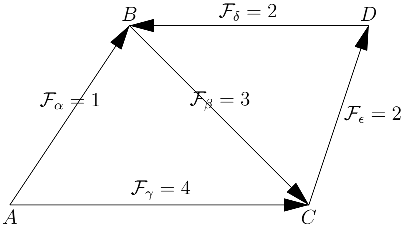
$$\mathcal{F}_\alpha + \mathcal{F}_\beta - \mathcal{F}_\gamma = 0$$

Incidence matrix $A \in \mathbb{R}^{m \times n}$

	nodes$_n$				flow
arcs$_m$	A	B	C	D	$\mathcal{F}$
α	-1	1	0	0	1
β	0	-1	1	0	3
γ	-1	0	1	0	4

[6] Čaklović (2012); Čaklović and Kurdija (2017)

Potential Method[6]



Incidence matrix $A \in \mathbb{R}^{m \times n}$

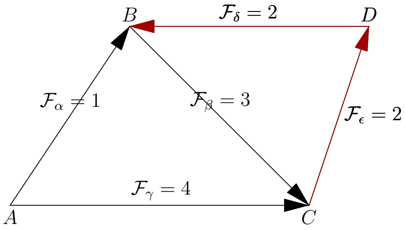
	nodes$_n$				flow
arcs$_m$	A	B	C	D	$\mathcal{F}$
α	-1	1	0	0	1
β	0	-1	1	0	3
γ	-1	0	1	0	4
δ	0	1	0	-1	2
ϵ	0	0	-1	1	2

Preference flow $\mathcal{F}$

$$\mathcal{F}_\alpha + \mathcal{F}_\beta - \mathcal{F}_\gamma = 0$$

[6] Čaklović (2012); Čaklović and Kurdija (2017)

Potential Method[6]



Preference flow $\mathcal{F}$

$$\mathcal{F}_\alpha + \mathcal{F}_\beta - \mathcal{F}_\gamma = 0$$
$$\mathcal{F}_\epsilon + \mathcal{F}_\delta + \mathcal{F}_\beta = 7$$

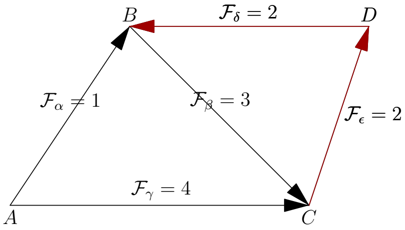
$\mathcal{F}$ cycle DBCD is not consistent!

Incidence matrix $A \in \mathbb{R}^{m \times n}$

	nodes$_n$				flow
arcs$_m$	A	B	C	D	$\mathcal{F}$
α	-1	1	0	0	1
β	0	-1	1	0	3
γ	-1	0	1	0	4
δ	0	1	0	-1	2
ϵ	0	0	-1	1	2

[6] Čaklović (2012); Čaklović and Kurdija (2017)

Potential Method[6]



Preference flow $\mathcal{F}$

$$\mathcal{F}_\alpha + \mathcal{F}_\beta - \mathcal{F}_\gamma = 0$$

$$\mathcal{F}_\epsilon + \mathcal{F}_\delta + \mathcal{F}_\beta = 7$$

$\mathcal{F}$ cycle DBCD is not consistent!

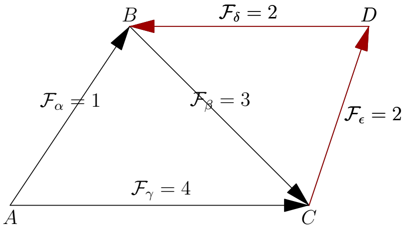
Incidence matrix $A \in \mathbb{R}^{m \times n}$

	nodes$_n$				flow
arcs$_m$	A	B	C	D	$\mathcal{F}$
α	-1	1	0	0	1
β	0	-1	1	0	3
γ	-1	0	1	0	4
δ	0	1	0	-1	2
ϵ	0	0	-1	1	2

$$N(A^\tau) \oplus R(A) = \mathbb{R}^m$$

[6] Čaklović (2012); Čaklović and Kurdija (2017)

Potential Method[6]



Preference flow $\mathcal{F}$

$$\mathcal{F}_\alpha + \mathcal{F}_\beta - \mathcal{F}_\gamma = 0$$
$$\mathcal{F}_\epsilon + \mathcal{F}_\delta + \mathcal{F}_\beta = 7$$

$\mathcal{F}$ cycle DBCD is not consistent!

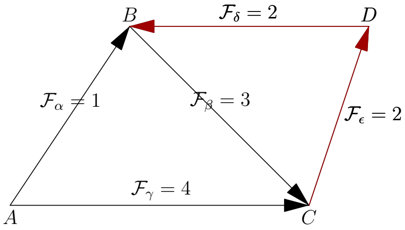
Incidence matrix $A \in \mathbb{R}^{m \times n}$

	nodes$_n$				flow
arcs$_m$	A	B	C	D	$\mathcal{F}$
α	-1	1	0	0	1
β	0	-1	1	0	3
γ	-1	0	1	0	4
δ	0	1	0	-1	2
ϵ	0	0	-1	1	2

$$N(A^\tau) \oplus R(A) = \mathbb{R}^m$$
$$c \oplus \mathcal{F}_o = \mathcal{F}$$

[6] Čaklović (2012); Čaklović and Kurdija (2017)

Potential Method[6]



Preference flow $\mathcal{F}$

$$\mathcal{F}_\alpha + \mathcal{F}_\beta - \mathcal{F}_\gamma = 0$$

$$\mathcal{F}_\epsilon + \mathcal{F}_\delta + \mathcal{F}_\beta = 7$$

$\mathcal{F}$ cycle DBCD is not consistent!

Incidence matrix $A \in \mathbb{R}^{m \times n}$

	nodes$_n$				flow
arcs$_m$	A	B	C	D	$\mathcal{F}$
α	−1	1	0	0	1
β	0	−1	1	0	3
γ	−1	0	1	0	4
δ	0	1	0	−1	2
ϵ	0	0	−1	1	2

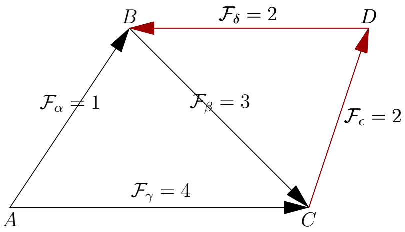
$$N(A^T) \oplus R(A) = \mathbb{R}^m$$

$$c \oplus \mathcal{F}_o = \mathcal{F}$$

$\mathcal{F}$ is consistent iff $\mathcal{F} \in R(A)$

[6] Čaklović (2012); Čaklović and Kurdija (2017)

Potential Method[6]



Preference flow $\mathcal{F}$

$$\mathcal{F}_\alpha + \mathcal{F}_\beta - \mathcal{F}_\gamma = 0$$

$$\mathcal{F}_\epsilon + \mathcal{F}_\delta + \mathcal{F}_\beta = 7$$

$\mathcal{F}$ cycle DBCD is not consistent!

Incidence matrix $A \in \mathbb{R}^{m \times n}$

	nodes$_n$				flow
arcs$_m$	A	B	C	D	$\mathcal{F}$
α	−1	1	0	0	1
β	0	−1	1	0	3
γ	−1	0	1	0	4
δ	0	1	0	−1	2
ϵ	0	0	−1	1	2

$$N(A^T) \oplus R(A) = \mathbb{R}^m$$

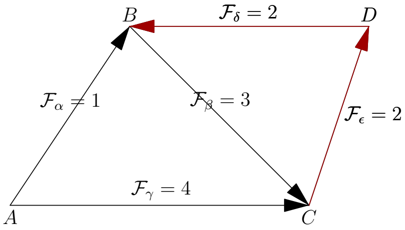
$$c \oplus \mathcal{F}_o = \mathcal{F}$$

$\mathcal{F}$ is consistent iff $\mathcal{F} \in R(A)$

$\mathcal{F}$ je consistent iff $AX = \mathcal{F}$

[6] Čaklović (2012); Čaklović and Kurdija (2017)

Potential Method[6]



Preference flow $\mathcal{F}$

$$\mathcal{F}_\alpha + \mathcal{F}_\beta - \mathcal{F}_\gamma = 0$$

$$\mathcal{F}_\epsilon + \mathcal{F}_\delta + \mathcal{F}_\beta = 7$$

$\mathcal{F}$ cycle DBCD is not consistent!

Incidence matrix $A \in \mathbb{R}^{m\times n}$

	nodes$_n$				flow
arcs$_m$	A	B	C	D	$\mathcal{F}$
α	-1	1	0	0	1
β	0	-1	1	0	3
γ	-1	0	1	0	4
δ	0	1	0	-1	2
ϵ	0	0	-1	1	2

$$N(A^\tau) \oplus R(A) = \mathbb{R}^m$$

$$c \oplus \mathcal{F}_o = \mathcal{F}$$

$\mathcal{F}$ is consistent iff $\mathcal{F} \in R(A)$

$\mathcal{F}$ je consistent iff $AX = \mathcal{F}$

$\mathcal{F}$ je consistent iff $c \perp \mathcal{F}, \forall c$

$c \in N(A^\tau)$ cycle

[6] Čaklović (2012); Čaklović and Kurdija (2017)

Potential of preference graph

A — incidence matrix, $n = \#$Vertices, $m = \#$Arcs.
$\mathcal{F}$ — preference flow.

Ranking of the vertices is given by *potential* X:

$$A^{\tau} A X = A^{\tau} \mathcal{F}.$$

$A^{\tau} \mathcal{F}$ — flow gain in vertices
$L = A^{\tau} A$ — *Laplace matrix* of the graph.

Potential of preference graph

A — incidence matrix, $n = \#\text{Vertices}$, $m = \#\text{Arcs}$.
$\mathcal{F}$ — preference flow.

Ranking of the vertices is given by *potential* X:

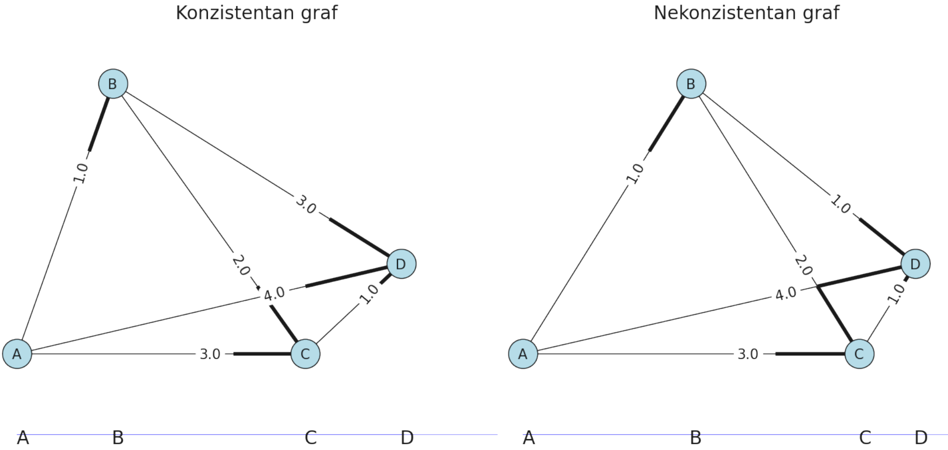
$$A^\tau AX = A^\tau \mathcal{F}.$$

$A^\tau \mathcal{F}$ — flow gain in vertices
$L = A^\tau A$ — *Laplace matrix* of the graph.

For connected graph, the matrix A has range $n-1$, the kernel is generated by the vector of ones $\mathbb{1} = [1, 1, \dots, 1]^\tau$. For uniqueness of X we put the condition

$$\sum_{i=1}^{n} x_i = 0.$$

Konsistency (bis)



Bibliography

Čaklović, L. (2012). Measure of Inconsistency for the Potential Method. In Torra, V., Narukawa, Y., López, B., and Villaret, M., editors, *MDAI*, volume 7647 of *Lecture Notes in Computer Science*, pages 102–114. Springer.

Čaklović, L. and Kurdija, A. S. (2017). A universal voting system based on the Potential Method. *European Journal of Operational Research*, 259:677–688.