Salton index

Eurovision 2022

Multiway networks

European airports

References

References

Salton cosine index in network analysis

Vladimir Batagelj
IMFM Ljubljana, IAM & FAMNIT UP Koper, FMF UL Ljubljana

INSNA Sunbelt 2025
June 23-29, 2025, Paris, France

Outline

1 Salton index
2 Eurovision 2022
3 Multiway networks
4 European airports
5 References

Vladimir Batagelj: vladimir.batagelj@fmf.uni-lj.si
Current version of slides (June 28, 2025 at 13 : 02): slides PDF
https://github.com/bavla/wNets

Inner product

Salton index

Eurovision 2022

Multiway networks

European airports

References

References

The *inner product* of vectors $x, y \in \mathbb{R}^n$ is defined as

$$\langle x, y \rangle = \sum_{i=1}^{n} x_i \cdot y_i$$

The following four properties hold for all $x, y, z \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$:

1. $\langle x, x \rangle \geq 0$ and $\langle x, x \rangle = 0$ if and only if $x = 0$,
2. $\langle x, y + z \rangle = \langle x, y \rangle + \langle x, z \rangle$,
3. $\langle x, \alpha y \rangle = \alpha \langle x, y \rangle$,
4. $\langle x, y \rangle = \langle y, x \rangle$.

An inner product $\langle ., . \rangle$ induces the *norm* of x

$$\|x\| = \sqrt{\langle x, x \rangle}$$

Cauchy-Schwarz inequality and Salton index

Cauchy-Schwarz inequality

$$| \langle x, y \rangle | \leq \|x\| \cdot \|y\|$$

Salton or cosine index

$$S(x, y) = \frac{\langle x, y \rangle}{\|x\| \cdot \|y\|} \in [-1, 1]$$

The Salton index has the following properties

1. $S(x, y) \in [-1, 1]$
2. $S(x, y) = S(y, x)$
3. $S(x, x) = 1$
4. $x, y \in (\mathbb{R}_0^+)^n \Rightarrow S(x, y) \in [0, 1]$
5. $S(\alpha x, \beta y) = S(x, y), \quad \alpha, \beta > 0$
6. $S(\alpha x, x) = 1, \quad \alpha > 0$

Salton index
V. Batagelj
Salton index
Eurovision 2022
Multiway networks
European airports
References
References

Salton index

Property 5 says that the value of the Salton index doesn't depend on the size of vectors but on their "shape". It is very useful when analyzing data with units of different "sizes" – for example, trade between world countries.

The Salton index is a measure of similarity.

It is usually transformed into a dissimilarity d using the transformation

$$d(x, y) = \frac{1 - S(x, y)}{2}$$

or the angular distance θ wp

$$\theta(x, y) = \frac{\arccos(S(x, y))}{\pi}$$

Matrix based (dis)similarities

Salton index

V. Batagelj

Salton index

Eurovision 2022

Multiway networks

European airports

References

References

Let $\mathbf{W} = [w[u, v]]$ be a matrix describing (one/two-mode) (weighted) network $\mathcal{N}$. For clustering units (nodes) we need a dissimilarity matrix $\mathbf{D}$. For two-mode networks, we can get it from the Salton matrix $\mathbf{S} = [s[u, v]]$ where

$$s[u, v] = S(w[u, \cdot], w[v, \cdot])$$

If the matrix $\mathbf{W}$ is one-mode we have a problem

$$w[u, \cdot] = [w[u, 1], \dots, w[u, i], \dots, w[u, u], \dots, w[u, v], \dots, w[u, k]]$$

$$w[v, \cdot] = [w[v, 1], \dots, w[v, i], \dots, w[v, u], \dots, w[v, v], \dots, w[v, k]]$$

In traditional (dis)similarities, comparing $w[u, i]$ and $w[v, i]$ we are comparing how u relates to i with how v relates to i. The problem arises for $i = u$ and $i = v$. We would need to compare $w[u, u]$ with $w[v, v]$ and $w[u, v]$ with $w[v, u]$. This leads to *corrected* (dis)similarities.

Corrected Salton index

Corrected Salton index of the link $(u, v) \in \mathcal{L}$

$$S'(u, v) = \frac{\langle w[u, \cdot], w[v, \cdot]\rangle + (w[u, u] - w[u, v]) \cdot (w[v, v] - w[v, u])}{\|w[u, \cdot]\| \cdot \|w[v, \cdot]\|}$$

It preserves the usual Salton index properties 1-6.

Projections of two-mode networks

Let $\mathcal{N} = ((\mathcal{U}, \mathcal{V}), \mathcal{L}, w)$ be a weighted two-mode network with a matrix $\mathbf{W} = [w[u, v]]$. A usual approach to its analysis is to *project* it to the set $\mathcal{U}$ or to the set $\mathcal{V}$ and analyze the so obtained ordinary (one-mode) weighted network.

The projection to the set $\mathcal{U}$ is determined by the matrix $\mathbf{C} = [c[u, t]] = \mathbf{W}\mathbf{W}^T$; and to the set $\mathcal{V}$ by the matrix $\mathbf{W}^T\mathbf{W}$.

$\mathcal{N}_C = (\mathcal{U}, \mathcal{L}_C, c)$, where $\mathcal{L}_C = \{(u, t) : c[u, t] \neq 0\}$ and for $(u, t) \in \mathcal{L}_C : c(u, t) = c[u, t]$.

$$c[u, t] = \sum_{v \in \mathcal{V}} w[u, v] \cdot w^T[v, t] = \sum_{v \in \mathcal{V}} w[u, v] \cdot w[t, v] = \langle w[u, \cdot], w[t, \cdot] \rangle$$

from which we can get the corresponding Salton matrix

$$s[u, v] = \frac{c[u, v]}{\sqrt{c[u, u] \cdot c[v, v]}}$$

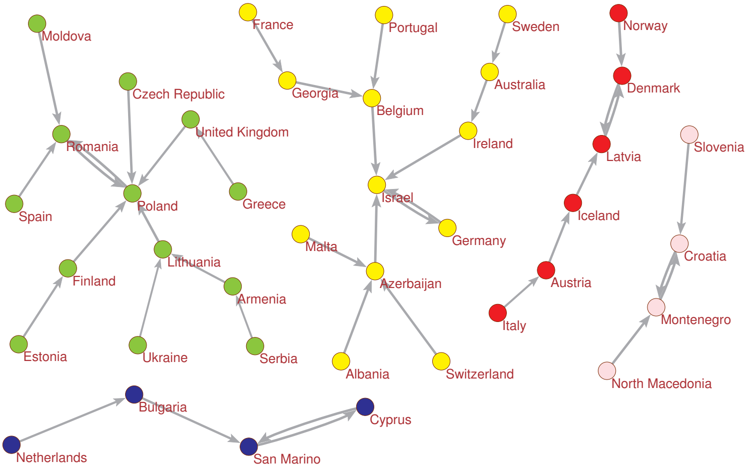
Eurovision 2022 data

EUROVISION 2022 SCOREBOARD

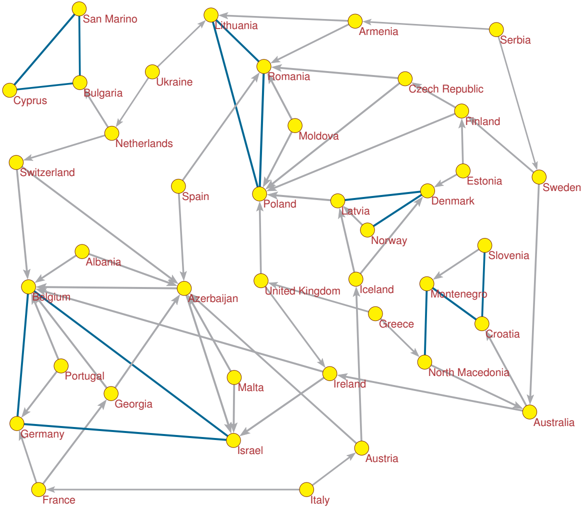
Total Tele Jury

1	Ukraine	**631**	17	16	19	12	18	18	12	18	19	16	17	12	12	22	18	22	10	22	15	19	12	**24**	**24**	8	**24**	16	12	8	15	**24**	20	22	14	7	18	12	12	13		12
2	United Kingdom	**466**	14	1	7	20	20	15	13		9	14	12	9	14	14	16	18	5	14	14	20	12	11	14	20	13	5	12	8	12	11	16	11	14	1	2	11	14	11	19	
3	Spain	**459**	12	20	18	3	15	16	16	13	14	13	5	1	3	12	11	10	17	6	17	11		4	11	19	11	16	12	22	8	7	22	9	22	14	5		14	14	1	15
4	Sweden	**438**	8	12	15	13	12	12		9	5	15	17	22	20		8	12	4	20	13	17	5	16	13	15	13	8	3	10	16	20	7	16	8	6	5	13		10	4	16
5	Serbia	**312**	3	6	16	10	6	6	10	**24**	11	10			14	8	7	7	8	1	6	7	11	2	1	6		**24**	3	22	3	5	8	7	8		18	8	12	12		2
6	Italy	**268**	20	10	4	6	15	12	3	10	6			10	2	3	10	3	8	6	6	11			7	17		16		13		4	3	8	8	4	18	15		10		
7	Moldova	**253**	1	6	4	7		8	7	8		8	5	6	5	10	8	10	9	6	7	6	15	7	5			7	12	7	4	7	8	12	4	12	7	7		4	6	8
8	Greece	**215**	12	5		4	6		18	3	22	3	12	6	6	3	4						10	7		1			12		19	2	4	2	17		12	2		12	7	4
9	Portugal	**207**	1	12	6	10	7	6	4	10		5	8	7		7	7		3	5	4		1	6	15		2	9	8		10	8		7		7	3	4		12	10	3
10	Norway	**182**			10	10	7		3	5		5	8	4	10			4	6	18	4	4	4	10	2	4	6	2	8	2		4	2	2	2	8	3	3	13	1	2	6
11	Netherlands	**171**	12	4		8	4	10		2			10	4	5	4	4	8	4	9		3	14	4	7			3			5	3	11	3			8		1	10	11	
12	Poland	**151**			3	4	3	7			3	7	4			12	2	8	10	4	10	7	3			6		4	6	2	5				4			1	5	1	12	18
13	Estonia	**141**	5	15	1					2		12	3		10									8	8	5	12	3	1	3	2	8		5		16			10		5	7
14	Lithuania	**128**		2	2					10	2		10	9	2		5			3	8			11			4	2		1	12	2	6			5	12		3		10	7
15	Australia	**125**					2	1			8	1		8	8			5	8		2	5	2	2		6	6		7	5		6		6	6			8	10	4	3	6
16	Azerbaijan	**106**	3		2			3	3		10			3	3		3	6	12							2	1	1		4					7	12		12	7		5	7
17	Switzerland	**78**				1			7	1		6	2					2					3	3	7		7		10	1	1	5	2	5	1		5	1			6	2
18	Romania	**65**							4		4					1		1	9		2		8			3	10								1	5		14				3
19	Belgium	**64**	2		5				6		2					6						4	7				5		10	3						8		5	1			
20	Armenia	**61**	4			6			10		1				1	12	10						8								4		1						2		2	
21	Finland	**38**	2									1		13			2						1	1		1										6			7		4	
22	Czech Republic	**38**			4			2								2			2	2	7	2				3				5						3			5		1	
23	Iceland	**20**											3								1				2								6								8	
24	France	**17**		7			1			1					1		1		1								2							1		2						
25	Germany	**6**				2								2																										2		

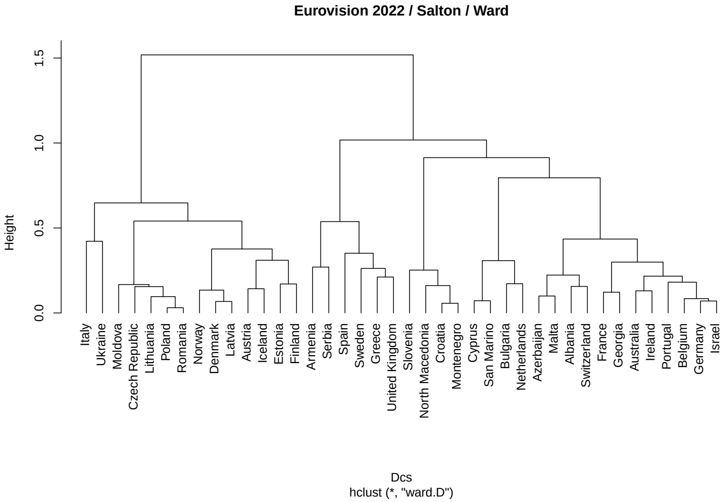
Corrected Salton 1-neighbors



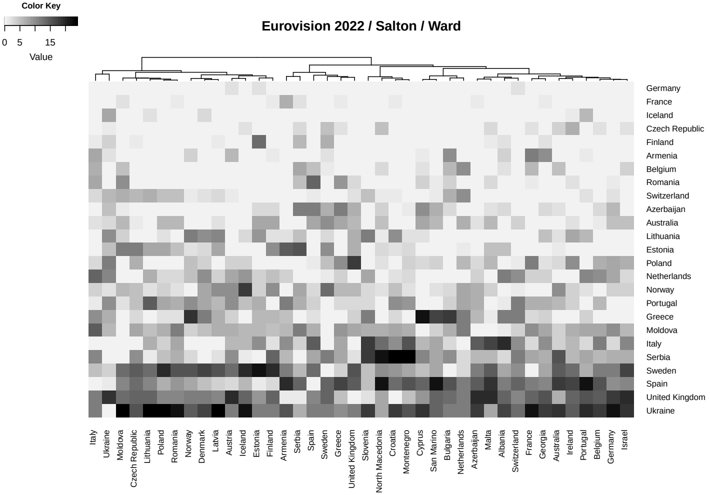
Corrected Salton 2-neighbors



Corrected Salton dendrogram



Matrix network display



Multiway networks

A *weighted multiway network* $\mathcal{N} = (\mathcal{V}, \mathcal{L}, w)$ is based on *nodes* from k finite sets (ways or dimensions) $\mathcal{V} = (\mathcal{V}_1, \mathcal{V}_2, \ldots, \mathcal{V}_k)$, the set of *links* $\mathcal{L}$, and the *weight* $w : \mathcal{L} \to \mathbb{R}$. The incidence function $I : \mathcal{L} \to \mathcal{V}_1 \times \mathcal{V}_2 \times \cdots \times \mathcal{V}_k$ assigns to each link $e \in \mathcal{L}$ a k-tuple of its nodes $I(e) = (e(1), e(2), \ldots, e(i), \ldots, e(k))$, $e(i) \in \mathcal{V}_i$. If for $i \neq j$, $\mathcal{V}_i = \mathcal{V}_j$, we say that $\mathcal{V}_i$ and $\mathcal{V}_j$ are of the same *mode*.

In a general multiway network, different additional data can be known for nodes and/or links $\mathcal{N} = (\mathcal{V}, \mathcal{L}, \mathcal{P}, \mathcal{W})$, where $\mathcal{P}$ is a set of node properties $p : \mathcal{V}_i \to S_p$, and $\mathcal{W}$ is a set of link weights $w : \mathcal{L} \to S_w$.

MWnets

To support the analysis of multi-way networks, I started developing an R package **`MWnets`**. It is available at GitHub/Bavla

https://github.com/bavla/ibm3m

```
> source("https://raw.githubusercontent.com/bavla/ibm3m/master/
multiway/MWnets.R")
```

Multiway network data sets

https://github.com/bavla/ibm3m/tree/master/data

For user-friendly inspection of a JSON file, you can use the Firefox browser or application JSON in Chrome.

Projection to a selected way

Because of the reordering option, we can assume that we selected the way $\mathcal{V}_1$. A projection to a selected way is a generalization of the projection of two-mode networks. The projection creates an ordinary weighted network $(\mathcal{V}_1, \mathcal{A}, p)$, $\mathcal{A} \subseteq \mathcal{V}_1 \times \mathcal{V}_1$ and $p : \mathcal{A} \to \mathbb{R}$. Let $u, t \in \mathcal{V}_1$ then

$$p(u,t) = \sum_{(v_2,\ldots,v_k)\in\mathcal{V}_2\times\cdots\times\mathcal{V}_k} w(u, v_2, \ldots, v_k) \cdot w(t, v_2, \ldots, v_k)$$

Note that the projection network is symmetric $p(u,t) = p(t,u)$ and considering that the right side in the definition of $p(u,t)$ is a inner (scalar) product we get $p(u,t) \leq \sqrt{p(u,u) \cdot p(t,t)}$ – Cauchy-Schwarz inequality. From it it follows $p(u,u) = p(t,t) = 0 \Rightarrow p(u,t) = 0$.

In computing the projection, we consider only existing links

$$p(u,t) = \sum_{(v_2,\ldots,v_k):((u,v_2,\ldots,v_k)\in\mathcal{L}\wedge(t,v_2,\ldots,v_k)\in\mathcal{L})} w(u, v_2, \ldots, v_k) \cdot w(t, v_2, \ldots, v_k)$$

Salton index
V. Batagelj
Salton index
Eurovision 2022
Multiway networks
European airports
References
References

$p(u,u) \geq 0$.

$(\exists z \in \{z : (u,z) \in \mathcal{L}\} : w(u,z) \neq 0) \rightarrow p(u,u) > 0$.

This network can be analyzed using traditional methods for analyzing weighted networks. Sometimes it is more appropriate to apply projection(s) to a normalized version of the original multi-way network.

From the projection p we can get the corresponding measure of similarity – *Salton index* $S(u,t)$ (Batagelj and Cerinšek (2013))

$$S(u,t) = \frac{p(u,t)}{\sqrt{p(u,u) \cdot p(t,t)}}$$

that can be used for clustering the set $\mathcal{V}_1$. In the case $p(u,u) = p(t,t) = 0$ we set $S(u,t) = 1$.

European airports and flight companies dataset

The *European airports and flight companies* dataset was collected by A. Cardillo and his collaborators for the paper

> *A. Cardillo, J. Gómez-Gardeñes, M. Zanin, M. Romance, D. Papo, F. del Pozo, S. Boccaletti: Emergence of Network Features from Multiplexity, Scientific Reports 3, 1344 (2013).*

It is an example of a multiway network. It contains 450 airports making two ways `airA` (departure airports) and `airB` (arrival airports), the third way is `line` (37 flight companies). There are 7176 links $(\mathit{from}, \mathit{to}, \mathit{line}) \in \mathcal{L}$ telling that there is a flight provided by the company *line* from the airport *from* to the airport *to*. Each link has the weight $w = 1$.

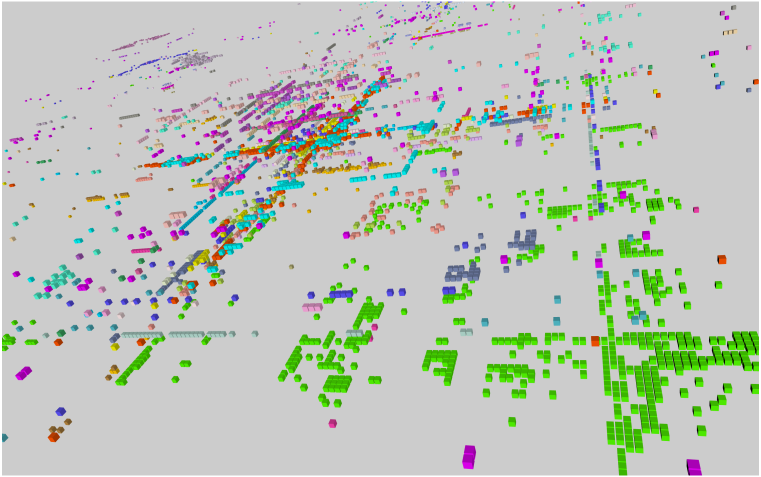
The multiway network `AirEu2013` represented in JSON format and an R package MWnets for analysis of multiway networks are available at GitHub/Bavla.

European airports and flight companies dataset

```
List of 6
 $ format: chr "MWnets"
 $ info  :List of 7
  ..$ network: chr "AirEu2013"
  ..$ title  : chr "Air Transportation Multiplex"
  ..$ ref    : chr "Cardillo A. et al. Emergence of network features from multiplexit
  ..$ href   : chr "http://complex.unizar.es/~atnmultiplex/"
 $ ways  :List of 3
  ..$ airA: chr "first airport"
  ..$ airB: chr "second airport"
  ..$ line: chr "airline"
 $ nodes :List of 3
  ..$ airA:'data.frame':        450 obs. of  7 variables:
  .. ..$ ID     : chr [1:450] "LCLK" "EDDF" "EDDK" "EGNX" ...
  .. ..$ lon    : num [1:450] 33.63 8.57 7.14 -1.33 -3.41 ...
  .. ..$ lat    : num [1:450] 34.9 50 50.9 52.8 50.7 ...
  .. ..$ iata   : chr [1:450] "LCA" "FRA" "CGN" "EMA" ...
  .. ..$ long   : chr [1:450] "Larnaca International Airport" "Frankfurt Airport" "Co
  .. ..$ region : chr [1:450] "Larnaka" "Hessen" "Nordrhein-Westfalen" "England" ...
  .. ..$ country: chr [1:450] "CY" "DE" "DE" "GB" ...
  ..$ airB:'data.frame':        450 obs. of  2 variables:
  .. ..$ ID  : chr [1:450] "LCLK" "EDDF" "EDDK" "EGNX" ...
  .. ..$ long: chr [1:450] "Larnaca International Airport" "Frankfurt Airport" "Colog
  ..$ line:'data.frame':        37 obs. of  1 variable:
  .. ..$ ID: chr [1:37] "Lufthansa" "Ryanair" "Easyjet" "British A" ...
 $ links :'data.frame': 7176 obs. of  4 variables:
  ..$ airA: int [1:7176] 1 1 2 2 2 2 2 2 2 2 ...
  ..$ airB: int [1:7176] 2 38 1 7 8 10 14 15 17 18 ...
  ..$ line: int [1:7176] 1 1 1 1 1 1 1 1 1 1 ...
  ..$ w   : int [1:7176] 1 1 1 1 1 1 1 1 1 1 ...
 $ data  :List of 1
  ..$ Eu:'data.frame':  40 obs. of  6 variables:
  .. ..$ alpha2   : chr [1:40] "AT" "BA" "BE" "BG" ...
  .. ..$ alpha3   : chr [1:40] "AUT" "BIH" "BEL" "BGR" ...
  .. ..$ Ccode    : int [1:40] 40 70 56 100 756 196 203 276 208 233 ...
  .. ..$ long     : chr [1:40] "Austria" "Bosnia and Herzegovina" "Belgium" "Bulgaria
  .. ..$ region   : chr [1:40] "Europe" "Europe" "Europe" "Europe" ...
  .. ..$ subregion: chr [1:40] "Western Europe" "Southern Europe" "Western Europe" "E
```

3D layout of `AirEu2013`

Airports dendrogram, 3D layout



The green links in the bottom right corner are provided by Ryan air – it is mostly linking airports that are not linked by other companies.

Conclusions

- An inner product can be defined for some other descriptions of (weighted) networks. It can be used to define the corresponding network (matrix) multiplication $\mathbf{C} = \mathbf{A} \cdot \mathbf{B}$

$$c[u, v] = \langle a[u, \cdot], b[\cdot, v] \rangle$$

- Using the "co-appearance" matrix $\mathbf{C}$ we can get also the *Jaccard* index

$$J[u, v] = \frac{c[u, v]}{c[u, u] + c[v, v] - c[u, v]}$$

and others.

- The notion of the network projection can be generalized in different ways (Batagelj (2020, 2021)).

Acknowledgments

The computational work reported in this presentation was performed in R (using packages igraph, netsWeight, and MwNets) and using the program Pajek. The code and data are available at https://github.com/bavla.

This work is partly supported by the Slovenian Research Agency ARIS (research program P1-0294 and research project J5-4596), and prepared within the framework of the COST action CA21163 (HiTEc).

References I

Batagelj, V. (2016). Network visualization based on JSON and D3.js (slides). *Second European Conference on Social Networks. June 14-17, 2016, Paris*. https://github.com/bavla/netsJSON/blob/master/doc/netVis.pdf.

Batagelj, V, Cerinšek, M (2013) On bibliographic networks. Scientometrics 96.3: 845-864.

Batagelj, V.: On fractional approach to the analysis of linked networks. Scientometrics 123 (2020), pages 621–633. Springer

Batagelj, V. (2021). Analysis of the Southern women network using the fractional approach. Social Networks 68, 229–236.

Batagelj, V. Cores in multiway networks. Soc. Netw. Anal. Min. 14, 122 (2024). Springer.

x3dom – instant 3D the HTML way! (2023) https://www.x3dom.org/