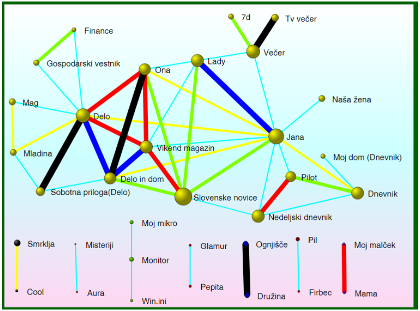
Projections of weighted two-mode networks

Vladimir Batagelj
IMFM Ljubljana, IAM UP Koper, and NRU HSE Moscow

1322. sredin seminar
Ljubljana, 5. oktober 2022

Outline



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

Current version of slides (October 5, 2022 at 16 : 23): slides PDF

https://github.com/bavla/NormNet/blob/main/docs/

Two-mode networks

In a *two-mode* (affiliation or bipartite) network $\mathcal{N} = ((U, V), L, w)$ the set of nodes is split into two disjoint sets (*modes*) U and V. Each link $e \in L$ has one end-node in the set U and the other end-node in the set V. The function $w : L \to \mathbb{R}$ assigns to each link its weight.

In general, the weight can be measured on different measurement scales (counts, ratio, interval, ordinal, nominal, binary, TQ, etc.).

Names of Participants of Group I	Code Numbers and Dates of Social Events Reported in *Old City Herald*													
	(1) 6/27	(2) 3/2	(3) 4/12	(4) 9/26	(5) 2/25	(6) 5/19	(7) 3/15	(8) 9/16	(9) 4/8	(10) 6/10	(11) 2/23	(12) 4/7	(13) 11/21	(14) 8/3
1. Mrs. Evelyn Jefferson	×	×	×	×	×	×		×	×					
2. Miss Laura Mandeville	×	×	×		×	×	×	×						
3. Miss Theresa Anderson		×	×	×	×	×	×	×	×					
4. Miss Brenda Rogers	×		×	×	×	×	×	×						
5. Miss Charlotte McDowd			×	×	×		×							
6. Miss Frances Anderson			×		×	×		×						
7. Miss Eleanor Nye					×	×	×	×						
8. Miss Pearl Oglethorpe						×		×	×					
9. Miss Ruth DeSand					×		×	×	×					
10. Miss Verne Sanderson							×	×	×			×		
11. Miss Myra Liddell								×	×	×		×		
12. Miss Katherine Rogers								×	×	×		×	×	×
13. Mrs. Sylvia Avondale							×	×	×	×		×	×	×
14. Mrs. Nora Fayette						×	×		×	×	×	×	×	×
15. Mrs. Helen Lloyd							×	×		×	×	×		
16. Mrs. Dorothy Murchison								×	×					
17. Mrs. Olivia Carleton									×		×			
18. Mrs. Flora Price									×		×			

Davis: Southern women, 1941 [9]

Two-mode network matrix and some notions

The network *matrix* **UV** of a two-mode network $\mathcal{UV}$ is defined as

$$UV[u,v] = \begin{cases} w(u,v) & (u,v) \in L \\ \square & \text{otherwise} \end{cases}$$

We represent number 0 with two symbols, 0 (weight 0) and $\square$ (no link) where $\square = 0$ with rules $\square + a = a$ and $\square \cdot a = \square$.

The function $\delta : \{\textbf{false}, \textbf{true}\} \to \{0,1\}$ is determined by $\delta(\textbf{false}) = 0$ and $\delta(\textbf{true}) = 1$. We will also use some additional functions:

out/in-degree $N(u)$ is the set of neighbors of node u
$\mathrm{od}_{UV}(u) = \sum_{v \in V} \delta((u,v) \in L) = |N(u)|$ and
$\mathrm{id}_{UV}(v) = \sum_{u \in U} \delta((u,v) \in L) = |N(v)|$

weighted out/in-degree (row/column sums)
$\mathrm{wod}_{UV}(u) = \sum_{v \in V} UV[u,v]$, $\quad \mathrm{wid}_{UV}(v) = \sum_{u \in U} UV[u,v]$ and
$\mathrm{wod}_{UV}(u/t) = \sum_{v \in N(u) \cap N(t)} UV[u,v]$.

Some notions

It holds $N(u) \cap N(t) \neq \emptyset \Rightarrow \mathrm{wod}_{UV}(u/t) \neq \square$.

We denote $U_{[d]} = \{u \in U : \mathrm{od}(u) \geq d\}$ and
$\mathcal{UV}_{[d]} = ((U_{[d]}, V), L(U_{[d]}), w|U_{[d]})$.
$\hat{U} = \{u \in U : \mathrm{wod}(u) \neq 0\}$

The *total weight* of links in the network $\mathcal{N} = (V, L, w)$

$$T(\mathbf{N}) = \sum_{(u,v) \in L} w(u, v) = \sum_{u,v} N[u, v] = \sum_{u} wod_N(u) = \sum_{v} wid_N(v)$$

Approaches

There are three main approaches to the analysis of two-mode networks:

1. treat the two-mode network as an ordinary one-mode network (degrees, components, etc.) considering a bipartition to sets U and V.
2. apply special methods developed for the analysis of two-mode networks (two-mode hubs and authorities, two-mode cores, 4-ring weights, blockmodeling, etc.).
3. transform (project) the two-mode network to a corresponding one-mode (weighted) network and use the usual methods (link cuts, cores, islands, skeletons, clustering, etc.) to analyze it.

In this talk, we will discuss the last option and limit our attention to numerical and binary scales.

Projections

$\mathcal{UV} = ((U, V), L, w)$ – two-mode network with a network matrix **UV**.

$p : \mathbf{UV} \to \mathbf{VV}$ *projection*, $\mathbf{VV} = [p(v, z)]$; and
$\mathcal{VV}$ the corresponding (ordinary, one-mode) network

1. undirected projection: $p(v, z) = p(z, v)$, resemblance
   - similarity: $p(v, z) \leq \min(p(v, v), p(z, z))$
   - dissimilarity: $p(v, z) \geq \max(p(v, v), p(z, z))$
2. directed projection: $\exists v, z : p(v, z) \neq p(z, v)$ ([2], [11])

Many projections are based on the multiplication of networks.

Multiplication of networks

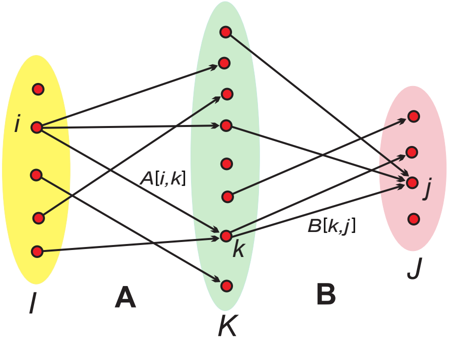
The *product* $\mathbf{C} = \mathbf{A} \cdot \mathbf{B}$ of two compatible matrices $\mathbf{A}_{I \times K}$ and $\mathbf{B}_{K \times J}$ is defined in the standard way

$$C[i,j] = \sum_{k \in K} A[i,k] \cdot B[k,j]$$

(it can be extended to semirings !!!)

The product of two compatible networks $\mathcal{N}_A = ((I,K), L_A, a)$ and $\mathcal{N}_B = ((K,J), L_B, b)$ is the network $\mathcal{N}_C = ((I,J), L_C, c)$ where $L_C = \{(i,j) : c[i,j] \neq \square\}$ and the weight c is determined by the matrix $\mathbf{C}$, $c(i,j) = C[i,j]$.

Multiplication of networks



In binary networks $\mathcal{N}_A$ and $\mathcal{N}_B$, the value of $C[i,j]$ of $\mathbf{C} = \mathbf{A} \cdot \mathbf{B}$ counts the number of ways we can go from the node $i \in I$ to the node $j \in J$ passing through K, $C[i,j] = |N_A(i) \cap N_B(j)|$.

$$C[i,j] = \sum_{k \in N_A(i) \cap N_B(j)} A[i,k] \cdot B[k,j]$$

Standard projections

A standard approach to the analysis of a two-mode network $\mathcal{UV}$ is to transform it into the corresponding one-mode networks determined by:

row projection to U: $\mathbf{UU} = \mathrm{row}(\mathbf{UV}) = \mathbf{UV} \cdot \mathbf{UV}^T$, or
column projection to V: $\mathbf{VV} = \mathrm{col}(\mathbf{UV}) = \mathbf{UV}^T \cdot \mathbf{UV}$

and analyze the obtained weighted network.

$$\mathrm{col}(\mathbf{UV}) = \mathbf{UV}^T \cdot \mathbf{UV} = \mathbf{UV}^T \cdot (\mathbf{UV}^T)^T = \mathrm{row}(\mathbf{UV}^T)$$

$$\mathrm{row}(\mathbf{UV}) = \mathrm{col}(\mathbf{UV}^T)$$

We will limit our discussion to column projections.

Outer product decomposition

For vectors $x = [x_1, x_2, \ldots, x_n]$ and $y = [y_1, y_2, \ldots, y_m]$ their *outer product* $x \circ y$ is defined as a matrix

$$x \circ y = [x_i \cdot y_j]_{n \times m}$$

then we can express the product $\mathbf{C}$ of two compatible matrices $\mathbf{A}$ and $\mathbf{B}$ as the *outer product decomposition*

$$\mathbf{C} = \mathbf{A} \cdot \mathbf{B} = \sum_k \mathbf{H}_k \quad \text{where} \quad \mathbf{H}_k = \mathbf{A}[\cdot, k] \circ \mathbf{B}[k, \cdot],$$

$\mathbf{A}[\cdot, k]$ is the k-th column of matrix $\mathbf{A}$, and $\mathbf{B}[k, \cdot]$ is the k-th row of matrix $\mathbf{B}$.
On the basis of outer product decomposition we have

$$T(\mathbf{C}) = T(\sum_k \mathbf{H}_k) = \sum_k T(\mathbf{H}_k) \quad \text{and} \quad T(\mathbf{H}_k) = \mathrm{wid}_A(k) \cdot \mathrm{wod}_B(k)$$

Structure of projection

In words

1. the product network is a sum of complete subgraphs;
2. the contribution of a node $k \in K$ to the total T is $T(\mathbf{H}_k)$.

This means that the nodes with different weighted degrees in K are not equally represented in the projection.

For a column projection $\mathbf{VV} = \mathrm{col}(\mathbf{UV})$, a real-life network $\mathcal{UV}$ can contain nodes $u \in U$ of degree 0 (in WA, works with no author) and 1 (in WA, single author works). Nodes from U of degree 0 do not contribute to the matrix $\mathbf{VV}$, and nodes of degree 1 contribute only to its neighbor's diagonal entry.

Fractional approach / "stochastic" normalization

$n(\mathbf{UV}) = [n(UV)[u, v]]$

$$n(UV)[u, v] = \begin{cases} \dfrac{UV[u, v]}{\mathrm{wod}_{UV}(u)} & u \in \hat{U} \\ 0 & u \notin \hat{U} \end{cases}$$

$$T(n(\mathbf{UV})) = \sum_{u \in U} \mathrm{wod}_{n(UV)}(u) = \sum_{u \in \hat{U}} 1 = |\hat{U}|$$

Interpretation: probabilistic co-linkage.

Fractional co-appearance: $\mathbf{Cn} = n(\mathbf{UV})^T \cdot n(\mathbf{UV})$,

$Cn[v, z] = \sum_{u \in \hat{U}} \frac{UV[u,v] \cdot UV[u,z]}{\mathrm{wod}(u)^2}$

$Cn[v, z] = Cn[z, v], \quad T(u) = T(\mathbf{H}_u) = \mathrm{wod}_{n(UV)}(u)^2$

$$T(\mathbf{Cn}) = \sum_{u \in U} \mathrm{wod}_{n(UV)}(u)^2 = \sum_{u \in \hat{U}} 1 = |\hat{U}|$$

Total weight preserving normalization

$$s(UV)[u,v] = \begin{cases} \dfrac{UV[u,v]}{\sqrt{\mathrm{wod}_{UV}(u)}} & u \in \hat{U} \\ 0 & u \notin \hat{U} \end{cases}$$

$$\mathrm{wod}_{n(UV)}(u) = \begin{cases} 1 & u \in \hat{U} \\ 0 & u \notin \hat{U} \end{cases}, \quad wod_{s(UV)}(u) = \begin{cases} \sqrt{\mathrm{wod}_{UV}(u)} & u \in \hat{U} \\ 0 & u \notin \hat{U} \end{cases}$$

Total weight preserving projection: $\mathbf{Cs} = s(\mathbf{UV})^T \cdot s(\mathbf{UV})$

$$Cs[v,z] = \sum_{u \in \hat{U}} \frac{UV[u,v] \cdot UV[u,z]}{\mathrm{wod}(u)}$$

$$Cs[v,z] = Cs[z,v], \; T(u) = \mathrm{wod}_{s(UV)}(u)^2 = \mathrm{wod}_{UV}(u)$$

$$T(\mathbf{Cs}) = \sum_{u \in U} \mathrm{wod}_{s(UV)}(u)^2 = \sum_{u \in \hat{U}} \mathrm{wod}_{UV}(u) = T(\mathbf{UV})$$

Embedding primary node values

Node values $c : U \to \mathbb{R}_0^+$ – impact factor, number of citations,...

$$x(UV)[u,v] = \begin{cases} \frac{\sqrt{c(u)}}{\mathrm{wod}_{UV}(u)} UV[u,v] & u \in \hat{U} \\ 0 & u \notin \hat{U} \end{cases}$$

$$\mathrm{wod}_{x(UV)}(u) = \begin{cases} \sqrt{c(u)} & u \in \hat{U} \\ 0 & u \notin \hat{U} \end{cases}$$

Embedded primary node values: $\mathbf{Cx} = x(\mathbf{UV})^T \cdot x(\mathbf{UV})$

$$Cx[v,z] = \sum_{u \in \hat{U}} \frac{c(u)}{\mathrm{wod}(u)^2} UV[u,v] \cdot UV[u,z], \quad Cx[v,z] = Cx[z,v]$$

$$T(u) = \sum_{v \in V} \sum_{z \in V} Cx[v,z] = \frac{c(u)}{\mathrm{wod}(u)^2} \sum_{v \in V} UV[u,v] \cdot \sum_{z \in V} UV[u,z] = c(u)$$

$$T(\mathbf{Cx}) = \sum_{u \in U} \mathrm{wod}_{x(UV)}(u)^2 = \sum_{u \in \hat{U}} c(u)$$

Binarization and left and right (fractional) contribution

Binarization: $b(\mathbf{UV})$: $b(UV)[u,v] = \delta(UV[u,v] \neq \square)$

Left contribution: $L(\mathbf{UV}) = \mathbf{UV} \cdot b(\mathbf{UV})^T$

$$L(UV)[u,t] = \sum_{v \in V} UV[u,v] \cdot b(UV)[t,v] = \mathrm{wod}_{UV}(u/t)$$

Left fractional contribution: $\ell(\mathbf{UV}) = n(\mathbf{UV}) \cdot b(\mathbf{UV})^T$

$$\ell(UV)[u,t] = \frac{1}{\mathrm{wod}_{UV}(u)} \sum_{v \in V} UV[u,v] \cdot b(UV)[t,v] = \frac{\mathrm{wod}_{UV}(u/t)}{\mathrm{wod}_{UV}(u)} \leq 1$$

Right fractional contribution: $r(\mathbf{UV}) = b(\mathbf{UV}) \cdot n(\mathbf{UV})^T$

$$r(UV)[u,t] = \frac{1}{\mathrm{wod}_{UV}(t)} \sum_{v \in V} b(UV)[u,v] \cdot UV[t,v] = \frac{\mathrm{wod}_{UV}(t/u)}{\mathrm{wod}_{UV}(t)}$$

$$r(UV)[u,t] = \ell(UV)[t,u]$$

Mean value similarities

$$\begin{aligned} VV_X[u,t] &= meanX(\ell(UV)[u,t], r(UV)[u,t]) \\ &= meanX(\ell(UV)[u,t], \ell(UV)[t,u]) \end{aligned}$$

$VV_A[u,t] = \frac{1}{2}(\ell(UV)[u,t] + \ell(UV)[t,u])$ – arithmetic mean
$VV_m[u,t] = \min(\ell(UV)[u,t], \ell(UV)[t,u])$ – minimum
$VV_M[u,t] = \max(\ell(UV)[u,t], \ell(UV)[t,u])$ – maximum
$VV_G[u,t] = \sqrt{\ell(UV)[u,t] \cdot \ell(UV)[t,u]}$ – geometric, Salton
$VV_H[u,t] = 2(\ell(UV)[u,t]^{-1} + \ell(UV)[t,u]^{-1})^{-1}$ – harmonic, Dice
$VV_J[u,t] = (\ell(UV)[u,t]^{-1} + \ell(UV)[t,u]^{-1} - 1)^{-1}$ – Jaccard

Note: $\ell(\mathbf{UV})$ can be computed from $L(\mathbf{UV})$
$L(UV)[u,u] = \mathrm{wod}_{UV}(u/u) = \mathrm{wod}_{UV}(u)$.

It holds: $VV_X[u,t] = VV_X[t,u]$, $VV_X[u,t] \in [0,1]$ and $VV_J[u,t] \leq VV_m[u,t] \leq VV_H[u,t] \leq VV_G[u,t] \leq VV_A[u,t] \leq VV_M[u,t]$.

Inner product

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

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

Using the inner product we can write $C[i,j] = \left\langle A^T[i,\cdot], B[\cdot,j] \right\rangle$.

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

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

Inner product and measurement scales

1. binary $x, y \in \{0, 1\}^n$: $\langle x, y \rangle = |X \cap Y|$, where $X = \{i : x_i = 1\}$
2. integer $x, y \in \mathbb{N}^n$: number of paths – basic rules of combinatorics
3. positive or nonnegative real numbers – similarity measure $x \leq y \Rightarrow \langle x, z \rangle \leq \langle y, z \rangle$
4. positive and negative real numbers – similarity measure

Some inner product inequalities

Cauchy-Schwarz inequality

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

Salton index, cosine

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

$\mathrm{incr}(x)$ = vector of elements of vector x ordered in increasing order
$\mathrm{decr}(x)$ = vector of elements of vector x ordered in decreasing order

$$m(x, y) = \langle \mathrm{incr}(x), \mathrm{decr}(y)\rangle \leq \langle x, y\rangle \leq \langle \mathrm{incr}(x), \mathrm{incr}(y)\rangle = M(x, y)$$

$$N(x, y) = \frac{\langle x, y\rangle - m(x, y)}{M(x, y) - m(x, y)} \in [0, 1]$$

$N(x, x) = 1$, $N(x, 0) = 1$, $N(x, y) = N(y, x)$, $N(\alpha x, y) = N(x, y)$, $\alpha > 0$, $N(e, x) = 1$

Salton and ordering

From the column projection matrix $\mathbf{VV} = \mathrm{col}(\mathbf{VV})$ we can compute the corresponding Salton similarity matrix $S(\mathbf{UV})$

$$S(\mathbf{UV})[v,z] = \frac{VV[v,z]}{\sqrt{VV[v,v] \cdot VV[z,z]}}$$

For computing the ordering similarity matrix $N(\mathbf{UV})$ we additionally need matrices $m(\mathbf{UV})$ and $M(\mathbf{UV})$

$$m(\mathbf{UV})[v,z] = m(UV[\cdot,v], UV[\cdot,z])$$

$$M(\mathbf{UV})[v,z] = M(UV[\cdot,v], UV[\cdot,z])$$

Then

$$N(\mathbf{UV})[v,z] = \frac{VV[v,z] - m(UV)[v,z]}{M(UV)[v,z] - m(UV)[v,z]}$$

(Dis)similarity based projections

$$VV[v,z] = r(UV[\cdot,v], UV[\cdot,z])$$

where r is a selected resemblance ((dis)similarity) measure compatible with the weight measurement scale [10].

Often the matrix **UV** is first normalized in an appropriate way. Not needed for S and N for ratio scales because
$S(\alpha v, \beta z) = S(v,z)$ and $N(\alpha v, \beta z) = N(v,z)$

Projection's skeleton

For a projection $C = (V, L_C, w_C)$ the graph $S_C = (V, L_C)$ is called a skeleton of C.

Most of the projections of the two-mode network $((U, V), L, w)$ to V have the same skeleton.

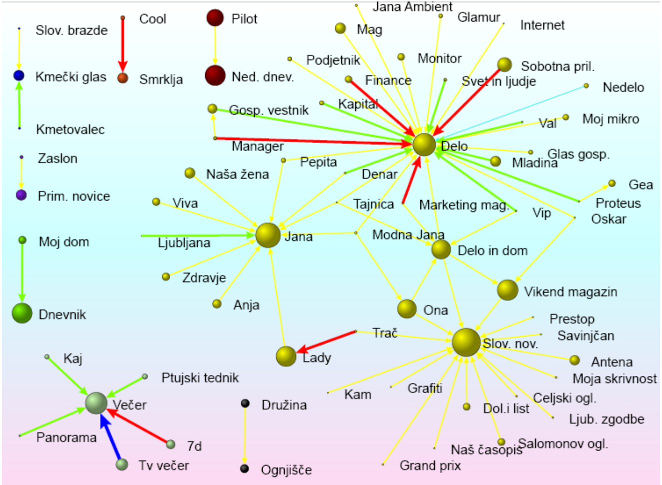
Asymmetric projections

The column projection matrix **VV** can be further transformed into an asymmetric matrix/network For example [2, 4, p. 94]

$$\text{MinDir}[v,z] = \begin{cases} \frac{VV[v,z]}{VV[v,v]} & VV[v,v] \leq VV[z,z] \\ \square & \textit{otherwise} \end{cases}$$

$$\text{MaxDir}[v,z] = \begin{cases} \frac{VV[v,z]}{VV[z,z]} & VV[v,v] \leq VV[z,z] \\ \square & \textit{otherwise} \end{cases}$$

MinDir of Slovenian journals 2000



Over 100000 people were asked in the years 1999 and 2000 about the journals they read. They mentioned 124 different journals. (source Cati)

Eurovision 2022 data

EUROVISION 2022 SCOREBOARD

Total Tele Jury

#	Country	Total																																								
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 Euclidean distance and Salton index

$UV[v, v] = UV[z, z] = 0$ – a country doesn't vote on its song

Corrected Euclidean distance

$d(v, z) = \sqrt{\sum_u (UV[u, v] - UV[u, z])^2}$

For $v, z \in U$

$d_c(v, z)^2 = d(v, z)^2 - (UV[v, v] - UV[v, z])^2 - (UV[z, v] - UV[z, z])^2 + (UV[z, v] - UV[v, z])^2 = d(v, z)^2 - 2 \cdot UV[z, v] \cdot UV[v, z]$

$d_c(v, z) = \sqrt{\sum_u (UV[u, v] - UV[u, z])^2 - 2 \cdot UV[z, v] \cdot UV[v, z]}$

Salton index

$\langle v, z \rangle = \sum_u UV[u, v] \cdot UV[u, z]$

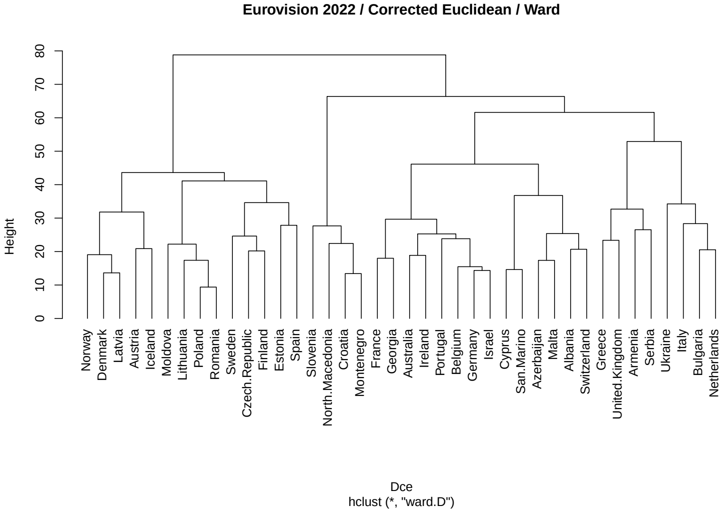
For $v, z \in U$

$\langle v, z \rangle_c = \sum_u UV[u, v] \cdot UV[u, z] + UV[z, v] \cdot UV[v, z]$

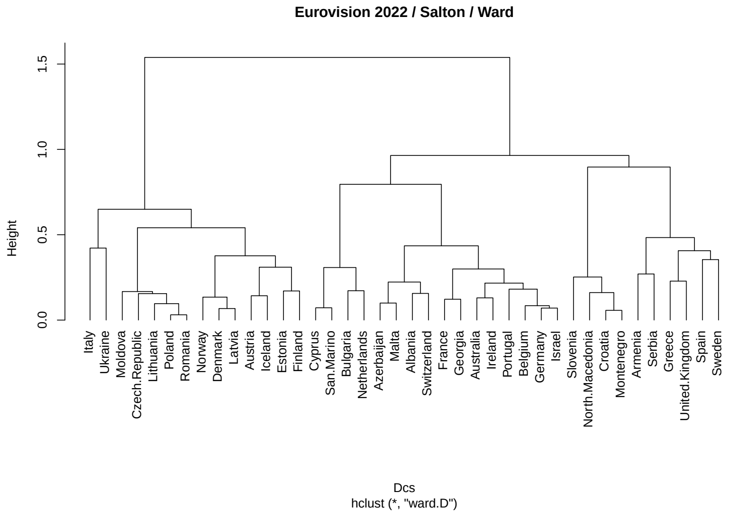
R code


```
> wdir <- "C:/Users/vlado/docs/papers/2022/sreda/1322/data"; setwd(wdir)
> R <- read.delim("Eurovision2022.csv",skip=1,row.names=1)
> dim(R)
> SC <- as.matrix(R[,2:41])
> SC[is.na(SC)] <- 0; m <- ncol(SC)
> rn <- rownames(SC); cn <- colnames(SC)
> # Corrected Euclidean distance
> Ce <- matrix(0,nrow=m,ncol=m)
> rownames(Ce) <- colnames(Ce) <- cn
> for(v in 1:(m-1)) for(z in (v+1):m) {
+    ss <- sum((SC[,v]-SC[,z])**2)
+    if((cn[v]%in%rn)&&(cn[z] %in% rn)) ss <- ss - 2*SC[cn[z],v]*SC[cn[v],z]
+    Ce[v,z] <- Ce[z,v] <- sqrt(ss)
+ }
> Dce <- as.dist(Ce)
> te <- hclust(Dce,method="ward.D")
> plot(te,hang=-1,cex=1,main="Eurovision 2022 / Corrected Euclidean / Ward")
> # Corrected Salton
> Co <- crossprod(SC)
> for(v in 1:(m-1)) for(z in (v+1):m) {
+    if((cn[v]%in%rn)&&(cn[z] %in% rn)) Co[v,z] <- Co[z,v] <- Co[z,v] + SC[cn[z],v]*SC[cn[v],z]
+ }
> S <- Co; diag(S) <- 1
> for(v in 1:(m-1)) for(z in (v+1):m) S[v,z] <- S[z,v] <- Co[v,z]/sqrt(Co[v,v]*Co[z,z])
> Dcs <- as.dist(1-S)
> ts <- hclust(Dcs,method="ward.D")
> plot(ts,hang=-1,cex=1,main="Eurovision 2022 / Salton / Ward")
> # export Salton to Pajek
> source("https://raw.githubusercontent.com/bavla/Rnet/master/R/Pajek.R")
> matrix2net(S,Net="Salton.net")
```

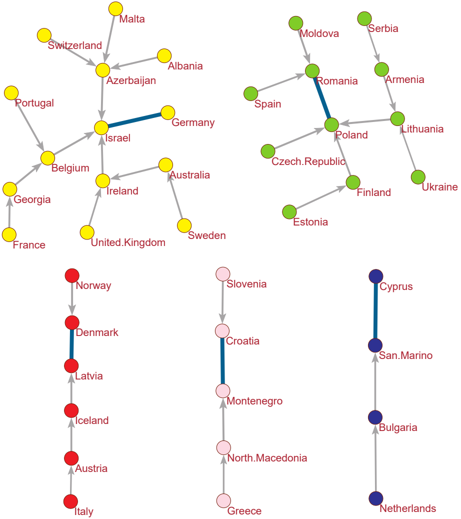
Corrected Euclidean dendrogram



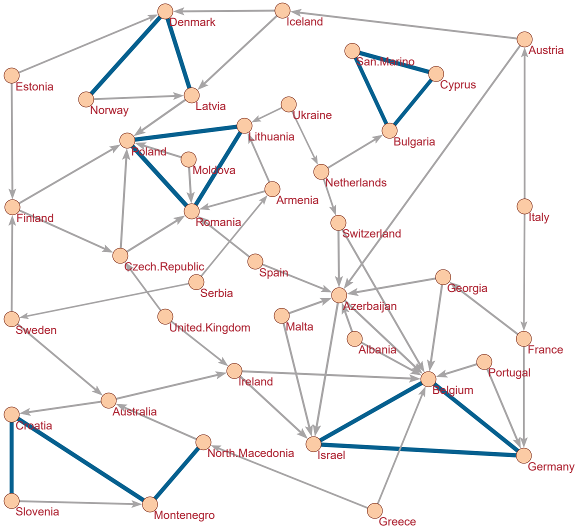
Corrected Salton dendrogram



Corrected Salton 1-neighbors



Corrected Salton 2-neighbors



Conclusions

- in principle we can base a projection on any resemblance measure between rows/columns (related to our question(s))
- problem of nodes with large degree – network multiplication with a threshold
- R package, Pajek macros
- testing: collection of datasets; examples GitHub
- (partial) extension to semirings
- https://github.com/bavla/NormNet/

Acknowledgments

This work is supported in part by the Slovenian Research Agency (research program P1-0294 and research projects J1-9187 and J5-2557), and prepared within the framework of the HSE University Basic Research Program.

References I

Batagelj, V.: Pajek / Example – Slovenian magazines and journals. Dagstuhl seminar, 2001. WWW

Batagelj, V., Mrvar, A. (2003). Density-based approaches to network analysis; Analysis of Reuter's terror news network. Workshop on Link Analysis for Detecting Complex Behavior (LinkKDD2003), August 27, 2003. PDF

Batagelj, V, Cerinšek, M.: On bibliographic networks. Scientometrics 96 (2013), pages 845–864. Springer

Batagelj, V., Doreian, P., Ferligoj, A., Kejžar, N.: Understanding Large Temporal Networks and Spatial Networks. Wiley 2014.

Batagelj, V.: On fractional approach to the analysis of linked networks. Scientometrics 123 (2020), pages621–633. Springer

Batagelj, V., Maltseva, D.: Temporal bibliographic networks. Journal of Informetrics, 14 (2020) 1, 101006.

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

References II

- Borgatti, S.P., Halgin, D.S. (2014) Analyzing Affiliation Networks. Chapter 28 in The SAGE Handbook of Social Network Analysis. John Scott, Peter J. Carrington (Eds.), Sage.
- Davis, A., Gardner, BB., Gardner, MR. (1941) Deep South. University of Chicago Press, Chicago, IL.
- Deza, M.M., Deza, E. (2013) Encyclopedia of Distances. Springer.
- Jana Krejčí: Multi-Criteria Decision Making with a new Fuzzy Approach; Pairwise Comparison Matrices and their Fuzzy Extension. Springer 2018.
- Leydesdorff, L. (2008). On the normalization and visualization of author co-citation data: Salton's Cosine versus the Jaccard index. Journal of the American Society for Information Science and Technology, 59(1), 77-85.
- Pajek datasets: Journals / Slovenian magazines and journals 1999 and 2000. WWW