



# On projections of a binary two-mode network

**Vladimir Batagelj**

IMFM Ljubljana, IAM UP Koper, and NRU HSE Moscow

**Yasin 2022**

on Zoom, April 4-8, 2022



# Outline

2-mode  
network  
projections

V. Batagelj

Two-mode  
networks

Binary  
projections

Fractional  
approach

Binary  
similarity  
measures

References

- 1 Two-mode networks
- 2 Binary projections
- 3 Fractional approach
- 4 Binary similarity measures
- 5 References

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

**Current version of slides (April 6, 2022 at 05:03):** [slides PDF](#)

<https://github.com/bavla/NormNet/tree/main/docs>

# Two-mode networks

2-mode  
network  
projections

V. Batagelj

Two-mode  
networks

Binary  
projections

Fractional  
approach

Binary  
similarity  
measures

References

A simple directed two-mode network  $\mathcal{N} = ((U, V), L, a)$  links the set of nodes  $U$  to the set of nodes  $V$  with the arcs from the set of links  $L$ . The mapping  $a: L \rightarrow \mathbb{R}^+$  assigns to each arc  $(u, v)$  its weight  $a(u, v)$ . The network  $\mathcal{N}$  can be represented with the corresponding matrix  $\mathbf{A} = [A[u, v]]_{u \in U, v \in V}$

$$A[u, v] = \begin{cases} a(u, v) & (u, v) \in L \\ 0 & \text{otherwise} \end{cases}$$

The set  $N(u)$  of *(out-)neighbors* (successors) of the node  $u \in U$

$$N(u) = \{v \in V : (u, v) \in L\}$$

and the set  $N^-(u)$  of *in-neighbors* (predecessors) of the node  $v \in V$

$$N^-(v) = \{u \in U : (u, v) \in L\}$$

In traditional two-mode networks we usually assume that  $U \cap V = \emptyset$ . In the case  $U = V$  we get an ordinary one-mode simple directed network.



# Multiplication of networks

2-mode  
network  
projections

V. Batagelj

Two-mode  
networks

Binary  
projections

Fractional  
approach

Binary  
similarity  
measures

References

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

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

(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  $L_C = \{(i, j) : c[i, j] \neq 0\}$  and the weight  $c$  is determined by the matrix  $\mathbf{C}$ ,  $c(i, j) = C[i, j]$ .

# Multiplication of networks

2-mode  
network  
projections

V. Batagelj

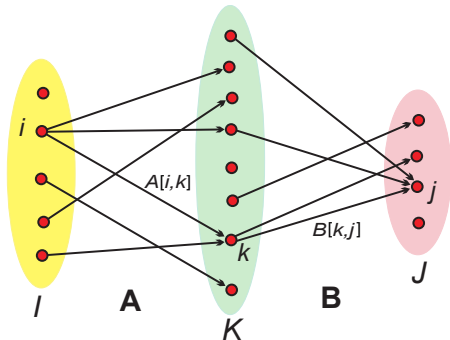
Two-mode  
networks

Binary  
projections

Fractional  
approach

Binary  
similarity  
measures

References



If all weights in networks  $\mathcal{N}_A$  and  $\mathcal{N}_B$  are equal to 1, 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  $K$ ,  $C[i, j] = |N_A(i) \cap N_B^-(j)|$ .

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



# Outer product decomposition

2-mode  
network  
projections

V. Batagelj

Two-mode  
networks

Binary  
projections

Fractional  
approach

Binary  
similarity  
measures

References

For vectors  $x = [x_1, x_2, \dots, x_n]$  and  $y = [y_1, y_2, \dots, y_m]$  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*

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

$\mathbf{A}[:, k]$  is the  $k$ -th column of matrix  $\mathbf{A}$ , and  $\mathbf{B}[k, :]$  is the  $k$ -th row of matrix  $\mathbf{B}$ .



# Projections

2-mode  
network  
projections

V. Batagelj

Two-mode  
networks

Binary  
projections

Fractional  
approach

Binary  
similarity  
measures

References

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

*row projection* to  $U$ :  $\text{row}(\mathbf{A}) = \mathbf{A} \cdot \mathbf{A}^T$ , or

*column projection* to  $V$ :  $\text{col}(\mathbf{A}) = \mathbf{A}^T \cdot \mathbf{A}$

and analyze the obtained weighted network.

$$\text{col}(\mathbf{A}) = \mathbf{A}^T \cdot \mathbf{A} = \mathbf{A}^T \cdot (\mathbf{A}^T)^T = \text{row}(\mathbf{A}^T), \quad \text{row}(\mathbf{A}) = \text{col}(\mathbf{A}^T)$$



# Binary projections

2-mode  
network  
projections

V. Batagelj

Two-mode  
networks

Binary  
projections

Fractional  
approach

Binary  
similarity  
measures

References

In the following, we will consider a two-mode network  $\mathcal{N} = ((W, A), L, wa)$  described with the corresponding binary matrix **WA**. To support our intuition, we can interpret it as an authorship network linking a work  $w \in W$  to its authors from  $A$ .

Its column projection **Co** =  $col(\mathbf{WA}) = \mathbf{WA}^T \cdot \mathbf{WA}$  has entries

$Co[a, b] = |N(a) \cap N(b)| = Co[b, a] = \#$  of works that authors  $a$  and  $b$  co-authored

**Co** is a *co-appearance* / co-authorship matrix.

$Co[a, a] = |N(a)| = \deg(a) = \#$  of works (co-)authored by the author  $a$ .





# Example: SNA18 / edge cut at level 15

$|W| = 70792$ ,  $|A| = 93011$

2-mode  
network  
projections

V. Batagelj

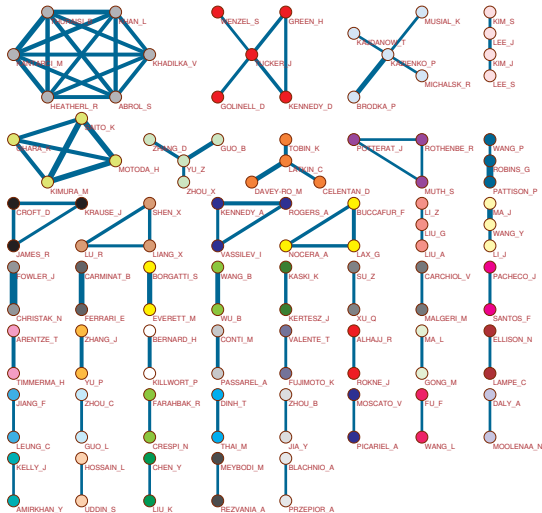
Two-mode  
networks

Binary  
projections

Fractional  
approach

Binary  
similarity  
measures

References





## ... nodes

2-mode  
network  
projections

V. Batagelj

Two-mode  
networks

Binary  
projections

Fractional  
approach

Binary  
similarity  
measures

References

A real-life network  $\mathcal{N}$  can contain nodes  $w \in W$  of degree 0 (works with no author) and 1 (single author works). Works with no author do not contribute to the matrix **Co**. Single author works contribute only to the author's diagonal entry in the matrix **Co**.

We denote  $W_{[d]} = \{w \in W : \deg(w) \geq d\}$  and  $\mathcal{N}_{[d]} = ((W_{[d]}, A), L(W_{[d]}), wa/W_{[d]})$ .

Using the entries of the matrix **Co** we can express the *Salton's cosine similarity* between nodes  $a_i$  and  $a_j$  as

$$\text{Cos}[a_i, a_j] = \frac{\text{Co}[a_i, a_j]}{\sqrt{\text{Co}[a_i, a_i] \cdot \text{Co}[a_j, a_j]}}$$



# Total weight

2-mode  
network  
projections

V. Batagelj

Two-mode  
networks

Binary  
projections

Fractional  
approach

Binary  
similarity  
measures

References

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

$$T(\mathcal{N}) = \sum_{(u,v) \in L} w(u, v)$$

On the basis of outer product decomposition we have

$$T(\mathbf{C}) = T\left(\sum_k \mathbf{H}_k\right) = \sum_k T(\mathbf{H}_k)$$

$$T(\mathbf{H}_k) = \left(\sum_i A[i, k]\right) \cdot \left(\sum_j B[k, j]\right)$$

Therefore for  $\mathbf{Co}$  we have  $\mathbf{H}_w = \mathbf{K}_{N(w)}$  and

$$T(\mathbf{Co}) = \sum_{w \in W} \deg(w)^2$$



# Structure of projection

2-mode  
network  
projections

V. Batagelj

Two-mode  
networks

Binary  
projections

Fractional  
approach

Binary  
similarity  
measures

References

In words

- a derived network is a sum of complete subgraphs;
- the contribution to the total of a node  $w \in W$  from the other set is  $\deg(w)^2$ .

This means that the nodes with a large degree in  $W$  are over-represented in the projection.



# Fractional approach

2-mode  
network  
projections

V. Batagelj

Two-mode  
networks

Binary  
projections

Fractional  
approach

Binary  
similarity  
measures

References

To make in a projection the contributions of all nodes from the set  $W$  equal we apply the *fractional approach* by normalizing the weights [2].

The normalized 2-mode network  $n(\mathbf{WA})$  has weights

$$n(\mathbf{WA})[w, a] = \frac{\mathbf{WA}[w, a]}{\max(1, \deg(w))}$$

*Normalized projection*     $\mathbf{Cn} = n(\mathbf{WA})^T \cdot n(\mathbf{WA})$

$$T(\mathbf{Cn}) = |W_{[1]}|$$



# Newman's normalization

2-mode  
network  
projections

V. Batagelj

Two-mode  
networks

Binary  
projections

Fractional  
approach

Binary  
similarity  
measures

References

Mark Newman proposed an alternative normalization that considers only co-authorship between different authors – single-author works and self co-authorship are excluded.

The *Newman's normalized* 2-mode network  $n'(\mathbf{WA})$  has weights

$$n'(\mathbf{WA})[w, a] = \begin{cases} \frac{\mathbf{WA}[w, a]}{\deg(w) - 1} & w \in W_{[2]} \\ 0 & \text{otherwise} \end{cases}$$

*Newman's projection*  $\mathbf{Cn}' = n(\mathbf{WA})^T \cdot n'(\mathbf{WA})$ .

Network: symmetrize with sum and remove loops.

# Example: SNA18 / selected Newman islands 10-30

$|W| = 70792$ ,  $|A| = 93011$

2-mode  
network  
projections

V. Batagelj

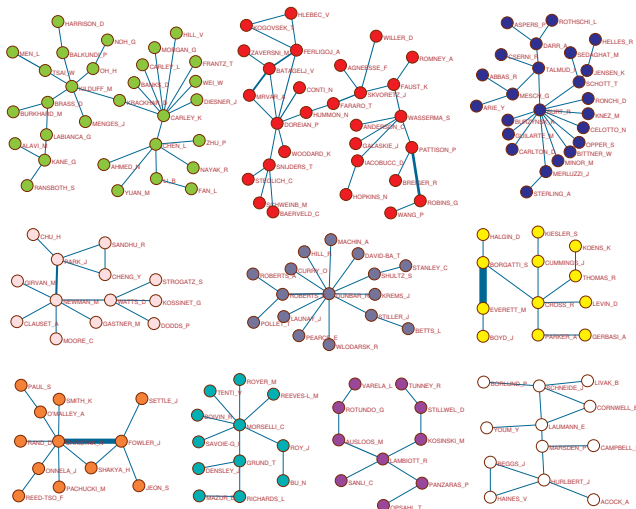
Two-mode  
networks

Binary  
projections

Fractional  
approach

Binary  
similarity  
measures

References





# Binary similarity measures

2-mode  
network  
projections

V. Batagelj

Two-mode  
networks

Binary  
projections

Fractional  
approach

Binary  
similarity  
measures

References

For analysis of two-mode networks, besides Salton's cosine similarity, the use of some other similarities between binary vectors was already proposed in SNA literature [5, p.420-424]. In principle, we could consider any similarity measure between binary vectors [1, 6].

For all these similarities the corresponding matrices can be computed from the events co-affiliation or co-appearances matrix (ordinary column projection)  $\mathbf{Co}$ . The similarities between vectors  $v = WA[:, a_i]$  and  $z = WA[:, a_j]$  are expressed in terms of the quantities  $a$ ,  $b$ ,  $c$ , and  $d$

		$z$		
		1	0	
$v$	1	$a$	$b$	$a + b$
	0	$c$	$d$	$c + d$
		$a + c$	$b + d$	$ W $

The quantity  $a$  counts the number of cases (indices) for which both vectors  $v$  and  $z$  have value 1, etc.





# Association coefficients

2-mode  
network  
projections

V. Batagelj

Two-mode  
networks

Binary  
projections

Fractional  
approach

Binary  
similarity  
measures

References

Association Coefficients

measure		definition	range	class
Russel and Rao (1940)	$s_1$	$\frac{a}{m}$	[1, 0]	
Kendall, Sokal-Michener (1958)	$s_2$	$\frac{a+d}{m}$	[1, 0]	S
Rogers and Tanimoto (1960)	$s_3$	$\frac{a+d}{m+b+c}$	[1, 0]	S
Hamann (1961)	$s_4$	$\frac{a+d-b-c}{m}$	[1, -1]	S
Sokal & Sneath (1963), $un_3^{-1}$ , $S$	$s_5$	$\frac{b+c}{a+d}$	[0, $\infty$ ]	S
Jaccard (1900)	$s_6$	$\frac{a}{a+b+c}$	[1, 0]	T
Kulczynski (1927), $T^{-1}$	$s_7$	$\frac{a}{b+c}$	[ $\infty$ , 0]	T
Dice (1945), Czekanowski (1913)	$s_8$	$\frac{a}{a+\frac{1}{2}(b+c)}$	[1, 0]	T
Sokal and Sneath	$s_9$	$\frac{a}{a+2(b+c)}$	[1, 0]	T
Kulczynski	$s_{10}$	$\frac{1}{2}(\frac{a}{a+b} + \frac{a}{a+c})$	[1, 0]	
Sokal & Sneath (1963), $un_4$	$s_{11}$	$\frac{1}{4}(\frac{a}{a+b} + \frac{a}{a+c} + \frac{d}{d+b} + \frac{d}{d+c})$	[1, 0]	
$Q_0$	$s_{12}$	$\frac{bc}{ad}$	[0, $\infty$ ]	Q
Yule (1912), $\omega$	$s_{13}$	$\frac{\sqrt{ad}-\sqrt{bc}}{\sqrt{ad}+\sqrt{bc}}$	[1, -1]	Q
Yule (1927), $Q$	$s_{14}$	$\frac{ad-bc}{ad+bc}$	[1, -1]	Q
$-bc -$	$s_{15}$	$\frac{4bc}{m^2}$	[0, 1]	
Driver & Kroeber (1932), Ochiai (1957)	$s_{16}$	$\frac{a}{\sqrt{(a+b)(a+c)}}$	[1, 0]	
Sokal & Sneath (1963), $un_5$	$s_{17}$	$\frac{ad}{\sqrt{(a+b)(a+c)(d+b)(d+c)}}$	[1, 0]	
Pearson, $\phi$	$s_{18}$	$\frac{ad-bc}{\sqrt{(a+b)(a+c)(d+b)(d+c)}}$	[1, -1]	
Baroni-Urbani, Buser (1976), $S^{**}$	$s_{19}$	$\frac{a+\sqrt{ad}}{a+b+c+\sqrt{ad}}$	[1, 0]	
Braun-Blanquet (1932)	$s_{20}$	$\frac{a}{\max(a+b, a+c)}$	[1, 0]	
Simpson (1943)	$s_{21}$	$\frac{a}{\min(a+b, a+c)}$	[1, 0]	
Michael (1920)	$s_{22}$	$\frac{4(ad-bc)}{(a+d)^2+(b+c)^2}$	[1, -1]	

# Association coefficients and co-appearance

For example, the *Jaccard similarity*

$$J = \frac{|N(v) \cap N(z)|}{|N(v) \cup N(z)|} = \frac{a}{a + b + c}.$$

The following equalities hold

$$a = |N(v) \cap N(z)| = Co[v, z]$$

$$a + b = |N(v)| = \deg(v) = Co[v, v]$$

$$a + c = |N(z)| = \deg(z) = Co[z, z]$$

$$a + b + c + d = |W|.$$

From them we get

$$b = |N(v) \setminus N(z)| = Co[v, v] - Co[v, z]$$

$$c = |N(z) \setminus N(v)| = Co[z, z] - Co[v, z]$$

$$d = |W| + Co[v, z] - Co[v, v] - Co[z, z]$$

$$a + b + c = |N(v) \cup N(z)| = Co[v, v] + Co[z, z] - Co[v, z]$$

For Jaccard similarity we get [3]

$$Jcol[v, z] = \frac{Co[v, z]}{Co[v, v] + Co[z, z] - Co[v, z]}$$

In this sense, the similarity measures (matrices) can be seen as a kind of compatibility normalization of the weights obtained with the standard projection [4].

The corresponding Jaccard network  $\mathcal{J} = (V, L_J, Jcol)$  is undirected with loops removed.

The Jaccard dissimilarity

$$d_J(A, B) = 1 - J(A, B) = \frac{|A \oplus B|}{|A \cup B|}$$

is also a distance. Note:  $b + c = |N(v) \oplus N(z)|$ .



# Example: SNA18 / some Jaccard islands 30-50

$|W| = 70792$ ,  $|A| = 93011$

2-mode  
network  
projections

V. Batagelj

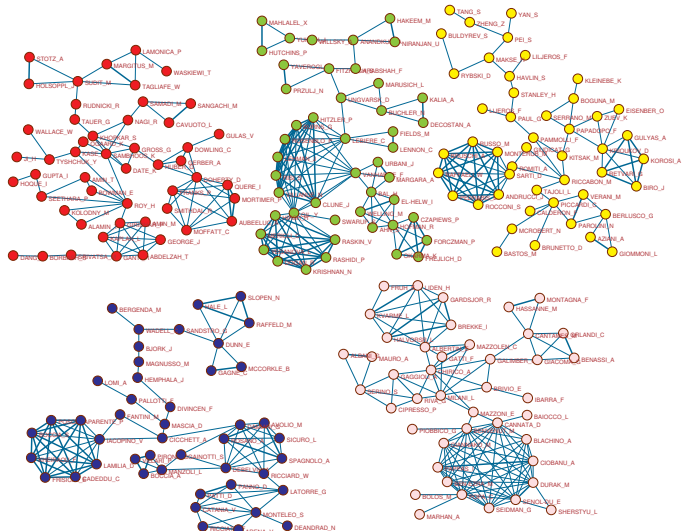
Two-mode  
networks

Binary  
projections

Fractional  
approach

Binary  
similarity  
measures

References



V. Batagelj

2-mode network projections



# Acknowledgments

2-mode  
network  
projections

V. Batagelj

Two-mode  
networks

Binary  
projections






Fractional  
approach

Binary  
similarity  
measures

References

The computational work reported on in this presentation was performed using R and Pajek. The code and data are available at <https://github.com/bavla/NormNet/>

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.

-  Batagelj, V., Bren, M. (1995) Comparing resemblance measures. J Classif 12 (1): 73-90.
-  Batagelj, V.: On fractional approach to analysis of linked networks. Scientometrics 123 (2020) 2: 621-633
-  Batagelj, V.: Analysis of the Southern women network using fractional approach. Social Networks 68(2022), 229-236
-  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](#)
-  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.



# References II

2-mode  
network  
projections

V. Batagelj




Two-mode  
networks

Binary  
projections

Fractional  
approach

Binary  
similarity  
measures

References

-  Cibulková, J., Šulc, Z., Řezanková, H., Sirota, S. (2020). Associations among similarity and distance measures for binary data in cluster analysis. Metodološki zvezki, Vol. 17, No. 1, 33–54. [PDF](#)
-  Deza, M.M., Deza, E. (2013) Encyclopedia of Distances. Springer.
-  Newman, M.E.J. (2004). Coauthorship networks and patterns of scientific collaboration. Proceedings of the National Academy of Sciences of the United States of America 101(Suppl1), 5200–5205.