Analysis of weighted networks

3. Multiplication of networks

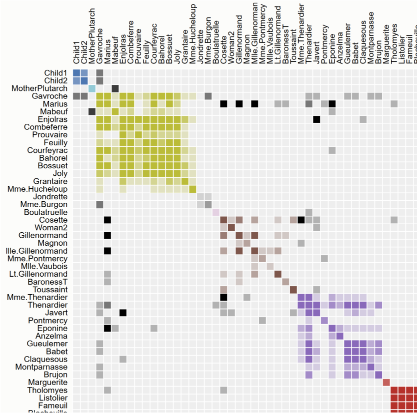
Vladimir Batagelj

IMFM Ljubljana, UP IAM Koper, UL FMF Ljubljana

INSNA Sunbelt 2025 workshop

June 23-29, 2025, Paris, France

Outline



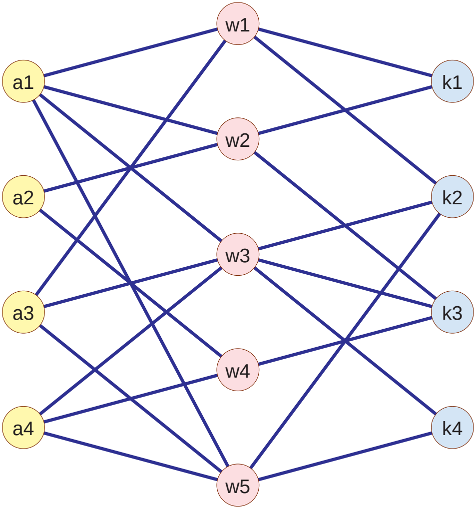
Les Misérables

Vladimir Batagelj: vladimir.batagelj@fmf.uni-lj.si
Current version of slides (June 21, 2025 at 07 : 44): slides PDF
https://github.com/bavla/Nets

Work in progress

Two mode networks

Two mode networks



Two mode networks

```
> library(jsonlite); library(httr)
> # AW <- read_graph("AW.net",format="pajek")
> AW <- read_graph(paste0(nWdir,"/data/AW.net"),format="pajek")
> AW
> V(AW)[[]]
> E(AW)[[]]
> is_bipartite(AW)
> AW <- delete_vertex_attr(AW,"type")
> AW
> is_bipartite(AW)
> V(AW)$type <- bipartite_mapping(AW)$type
> is_bipartite(AW)
> AW$name <- "Authors-Works"
> AW$date <- date()
> AW$twomode <- TRUE
> AW
> plot(AW,main=AW$name)

> WK <- read_graph(WKfile,format="pajek")
> WK <- read_graph(paste0(nWdir,"/data/WK.net"),format="pajek")
> WK
```

Two mode networks

```
> davis <- read.csv(file.choose(), header=FALSE)
> D <- graph_from_data_frame(davis, directed=FALSE)
> V(D)$type <- bipartite_mapping(D)$type
> V(D)$name[19:32] <- paste("event-",1:14,sep="")
> is_bipartite(D)
> D$name <- "Davis"; D$twomode <- TRUE
```

NetsJSON


```
> saveRDS(AW,file="AW.rds")
> AW1 <- readRDS(file="AW.rds")
> AW1
> write_graph_netsJSON(AW,file="AW.json")
> AW2 <- netsJSON_to_graph(fromJSON("AW.json"),directed=TRUE)
> AW2
> AW3 <- netsJSON_to_graph(fromJSON("AW.json"),directed=FALSE)
> AW3
> V(AW3)[[]]
> E(AW3)[[]]
> graph_attr(AW3)
```

Matrices

Matrix – sparse matrices

```
as_sparse_matrix <- function(N,weight="weight"){
  if(is_bipartite(N)) return(as_biadjacency_matrix(N,
    attr=weight,sparse=TRUE))
  return(as_adjacency_matrix(N,attr=weight,sparse=TRUE))
}
```

Matrices

Matrix – sparse matrices

Multiplication of networks

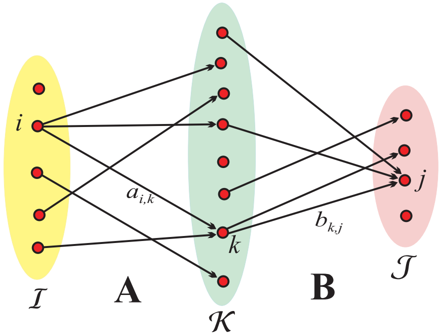
To a simple (no parallel arcs) two-mode *network* $\mathcal{N} = (\mathcal{I}, \mathcal{J}, \mathcal{A}, w)$; where $\mathcal{I}$ and $\mathcal{J}$ are sets of *nodes*, $\mathcal{A}$ is a set of *arcs* linking $\mathcal{I}$ and $\mathcal{J}$, and $w : \mathcal{A} \to \mathbb{R}$ (or some other semiring) is a *weight*; we can assign a *network matrix* $\mathbf{W} = [w_{i,j}]$ with elements: $w_{i,j} = w(i,j)$ for $(i,j) \in \mathcal{A}$ and $w_{i,j} = 0$ otherwise.

Given a pair of compatible networks $\mathcal{N}_A = (\mathcal{I}, \mathcal{K}, \mathcal{A}_A, w_A)$ and $\mathcal{N}_B = (\mathcal{K}, \mathcal{J}, \mathcal{A}_B, w_B)$ with corresponding matrices $\mathbf{A}_{\mathcal{I}\times\mathcal{K}}$ and $\mathbf{B}_{\mathcal{K}\times\mathcal{J}}$ we call a *product of networks* $\mathcal{N}_A$ and $\mathcal{N}_B$ a network $\mathcal{N}_C = (\mathcal{I}, \mathcal{J}, \mathcal{A}_C, w_C)$, where $\mathcal{A}_C = \{(i,j) : i \in \mathcal{I}, j \in \mathcal{J}, c_{i,j} \neq 0\}$ and $w_C(i,j) = c_{i,j}$ for $(i,j) \in \mathcal{A}_C$. The product matrix $\mathbf{C} = [c_{i,j}]_{\mathcal{I}\times\mathcal{J}} = \mathbf{A} * \mathbf{B}$ is defined in the standard way

$$c_{i,j} = \sum_{k\in\mathcal{K}} a_{i,k} \cdot b_{k,j}$$

In the case when $\mathcal{I} = \mathcal{K} = \mathcal{J}$ we are dealing with ordinary one-mode networks (with square matrices).

Multiplication of networks



$$c_{i,j} = \sum_{k \in N_A(i) \cap N_B^-(j)} a_{i,k} \cdot b_{k,j}$$

If all weights in networks $\mathcal{N}_A$ and $\mathcal{N}_B$ are equal to 1 the value of $c_{i,j}$ counts the number of ways we can go from $i \in \mathcal{I}$ to $j \in \mathcal{J}$ passing through $\mathcal{K}$, $c_{i,j} = |N_A(i) \cap N_B^-(j)|$.

Multiplication of networks

The standard matrix multiplication has the complexity $O(|\mathcal{I}| \cdot |\mathcal{K}| \cdot |\mathcal{J}|)$ – it is too slow to be used for large networks. For sparse large networks we can multiply much faster considering only nonzero elements.

for k **in** $\mathcal{K}$ **do**
  for (i, j) **in** $N_A^-(k) \times N_B(k)$ **do**
    if $\exists c_{i,j}$ **then** $c_{i,j} := c_{i,j} + a_{i,k} \cdot b_{k,j}$
    else new $c_{i,j} := a_{i,k} \cdot b_{k,j}$

`Networks/Multiply Networks`

In general the multiplication of large sparse networks is a 'dangerous' operation since the result can 'explode' – it is not sparse.

Multiplication of networks

From the network multiplication algorithm we see that each intermediate node $k \in \mathcal{K}$ adds to a product network a complete two-mode subgraph $K_{N_A^-(k),N_B(k)}$ (or, in the case $\mathcal{I} = \mathcal{J}$, a complete subgraph $K_{N(k)}$). If both degrees $\deg_A(k) = |N_A^-(k)|$ and $\deg_B(k) = |N_B(k)|$ are large then already the computation of this complete subgraph has a quadratic (time and space) complexity – the result 'explodes'.

If at least one of the sparse networks $\mathcal{N}_A$ and $\mathcal{N}_B$ has small maximal degree on $\mathcal{K}$ then also the resulting product network $\mathcal{N}_C$ is sparse.

If for the sparse networks $\mathcal{N}_A$ and $\mathcal{N}_B$ there are in $\mathcal{K}$ only few nodes with large degree and no one among them with large degree in both networks then also the resulting product network $\mathcal{N}_C$ is sparse.

Projections of two-mode networks

Often we transform a two-mode network $\mathcal{N} = (\mathcal{U}, \mathcal{V}, \mathcal{E}, w)$ into an ordinary (one-mode) network $\mathcal{N}_1 = (\mathcal{U}, \mathcal{E}_1, w_1)$ or/and $\mathcal{N}_2 = (\mathcal{V}, \mathcal{E}_2, w_2)$, where $\mathcal{E}_1$ and w_1 are determined by the matrix $\mathbf{W}^{(1)} = \mathbf{W}\mathbf{W}^T$,

$$w^{(1)}_{uv} = \sum_{z \in \mathcal{V}} w_{uz} \cdot w^T_{zv} = \sum_{z \in \mathcal{V}} w_{uz} \cdot w_{vz}.$$

Evidently $w^{(1)}_{uv} = w^{(1)}_{vu}$. There is an edge $(u : v) \in \mathcal{E}_1$ in $\mathcal{N}_1$ iff $N(u) \cap N(v) \neq \emptyset$. Its weight is $w_1(u, v) = w^{(1)}_{uv}$.

The network $\mathcal{N}_2$ is determined in a similar way by the matrix $\mathbf{W}^{(2)} = \mathbf{W}^T\mathbf{W}$.
The networks $\mathcal{N}_1$ and $\mathcal{N}_2$ are analyzed using standard methods.

`Network/2-Mode Network/2-Mode to 1-Mode/Rows`

Derived networks

WA – works $\times$ authors – authorship network
WK – works $\times$ keywords
Ci – works $\times$ works – citation network
AK $= \mathbf{WA}^T * \mathbf{WK}$ – authors $\times$ keywords
Co $= \mathbf{WA}^T * \mathbf{WA}$ – coauthorship
ACi $= \mathbf{WA}^T * \mathbf{Ci} * \mathbf{WA}$ – citations between authors

co_{ij} = the number of works that authors i and j wrote together
It holds: $co_{ij} = co_{ji}$. Using the weights co_{ij} we can determine the Salton's cosine similarity or Ochiai coefficient between authors i and j as

$$S(i,j) = \cos(i,j) = \frac{co_{ij}}{\sqrt{co_{ii}\, co_{jj}}}$$

Properties of Salton index

The Salton index has the following properties

1. $S(i,j) \in [-1,1]$
2. $S(i,j) = S(j,i)$
3. $S(i,i) = 1$
4. $wa_{pi} \in \mathbb{R}_0^+ \Rightarrow S(u,t) \in [0,1]$
5. $S(\alpha i, \beta j) = S(i,j), \quad \alpha, \beta > 0$
6. $S(\alpha i, i) = 1, \quad \alpha > 0$

Outer product decomposition

WA^T · WK

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 previous observation about the structure of product network as the *outer product decomposition*

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

For binary (weights) networks we have $\mathbf{H}_k = K_{N_A^-(k), N_B(k)}$.

Example A

As an example let us take the binary network matrices **WA** and **WK**:

$$\mathbf{WA} = \begin{array}{c c} & \begin{array}{cccc} a_1 & a_2 & a_3 & a_4 \end{array} \\ \begin{array}{c} w_1 \\ w_2 \\ w_3 \\ w_4 \\ w_5 \end{array} & \left[\begin{array}{cccc} 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \end{array}\right] \end{array}, \quad \mathbf{WK} = \begin{array}{c c} & \begin{array}{cccc} k_1 & k_2 & k_3 & k_4 \end{array} \\ \begin{array}{c} w_1 \\ w_2 \\ w_3 \\ w_4 \\ w_5 \end{array} & \left[\begin{array}{cccc} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{array}\right] \end{array}$$

and compute the product $\mathbf{H} = \mathbf{WA}^T \cdot \mathbf{WK}$. We get a network matrix $\mathbf{H}$ which can be decomposed as

Conclusions

$$
\begin{array}{c|cccc}
\mathbf{H} & k_1 & k_2 & k_3 & k_4 \\
\hline
a_1 & & & & \\
a_2 & & & & \\
a_3 & & & & \\
a_4 & & & &
\end{array}
\begin{bmatrix} 2 & 3 & 2 & 2 \\ 1 & 0 & 2 & 0 \\ 1 & 3 & 1 & 2 \\ 0 & 2 & 2 & 2 \end{bmatrix}
=
\begin{array}{c} \mathbf{H}_1 \\ a_1 \\ a_2 \\ a_3 \\ a_4 \end{array}
\begin{array}{c} k_1 \quad k_2 \quad k_3 \quad k_4 \\ \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \end{array}
+
\begin{array}{c} \mathbf{H}_2 \\ a_1 \\ a_2 \\ a_3 \\ a_4 \end{array}
\begin{array}{c} k_1 \quad k_2 \quad k_3 \quad k_4 \\ \begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \end{array}
$$

$$
\begin{array}{c} \mathbf{H}_3 \\ a_1 \\ a_2 \\ a_3 \\ a_4 \end{array}
\begin{array}{c} k_1 \quad k_2 \quad k_3 \quad k_4 \\ \begin{bmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{bmatrix} \end{array}
+
\begin{array}{c} \mathbf{H}_4 \\ a_1 \\ a_2 \\ a_3 \\ a_4 \end{array}
\begin{array}{c} k_1 \quad k_2 \quad k_3 \quad k_4 \\ \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \end{array}
+
\begin{array}{c} \mathbf{H}_5 \\ a_1 \\ a_2 \\ a_3 \\ a_4 \end{array}
\begin{array}{c} k_1 \quad k_2 \quad k_3 \quad k_4 \\ \begin{bmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix} \end{array}
$$

Derived networks

We can use the multiplication to obtain new networks from existing *compatible* two-mode networks. For example, from basic bibliographic networks **WA** and **WK** we get

$$\mathbf{AK} = \mathbf{WA}^T \cdot \mathbf{WK}$$

a network relating authors to keywords used in their works, and

$$\mathbf{Ca} = \mathbf{WA}^T \cdot \mathbf{Ci} \cdot \mathbf{WA}$$

is a network of citations between authors.
Networks obtained from existing networks using some operations are called *derived* networks. They are very important in analysis of collections of *linked* networks.
What is the meaning of the product network? In general we could consider weights, addition and multiplication over a selected semiring [**?**]. In this paper we will limit our attention to the traditional addition and multiplication of real numbers.
The weight $\mathbf{AK}[a, k]$ is equal to the number of times the author a used the keyword k in his/her works.

Conclusions

Using network multiplication we can also transform a given two-mode network, for example $\mathbf{WA}$, into corresponding ordinary one-mode networks (*projections*)

$$\mathbf{WW} = \mathbf{WA} \cdot \mathbf{WA}^T \qquad \text{and} \qquad \mathbf{AA} = \mathbf{WA}^T \cdot \mathbf{WA}$$

The obtained projections can be analyzed using standard network analysis methods. This is a traditional recipe how to analyze two-mode networks. Often the weights are not considered in the analysis; and when they are considered we have to be very careful about their meaning.

The weight $\mathbf{WW}[p, q]$ is equal to the number of common authors of works p and q.

The weight $\mathbf{AA}[a, b]$ is equal to the number of works that author a and b coauthored. In a special case when $a = b$ it is equal to the number of works that the author a wrote. The network $\mathbf{AA}$ is describing the *coauthorship* (collaboration) between authors and is also denoted as $\mathbf{Co}$ – the "first" coauthorship network.

Conclusions

In the paper [?] it was shown that there can be problems with the network **Co** when we try to use it for identifying the most collaborative authors. By the outer product decomposition the coauthorship network **Co** is composed of complete subgraphs on the set of work's coauthors. Works with many authors produce large complete subgraphs, thus bluring the collaboration structure, and are over-represented by its total weight. To see this, let $S_x = \sum_i x_i$ and $S_y = \sum_j y_j$ then the *contribution* of the outer product $x \circ y$ is equal

$$T = \sum_{i,j} (x \circ y)_{ij} = \sum_i \sum_j x_i \cdot y_j = \sum_i x_i \cdot \sum_j y_j = S_x \cdot S_y$$

In general each term $\mathbf{H}_w$ in the outer product decomposition of the product $\mathbf{C}$ has different total weight $T(\mathbf{H}_w) = \sum_{a,k} (\mathbf{H}_w)_{ak}$ leading to over-representation of works with large values. In the case of coautorship network **Co** we have $S(\mathbf{WA}[w,.]) = \text{outdeg}_{\mathbf{WA}}(w)$ and therefore $T(\mathbf{H}_w) = \text{outdeg}_{\mathbf{WA}}(w)^2$. To resolve the problem we apply the fractional approach.

Fractional approach

To make the contributions of all works equal we can apply the *fractional* approach by normalizing the weights: setting $x' = x/S_x$ and $y' = y/S_y$ we get $S_{x'} = S_{y'} = 1$ and therefore $T(\mathbf{H}'_w) = 1$ for all works w.

In the case of two-mode networks **WA** and **WK** we denote

$$S_w^{\mathbf{WA}} = \begin{cases} \sum_a \mathbf{WA}[w,a] & \text{outdeg}_{\mathbf{WA}}(w) > 0 \\ 1 & \text{outdeg}_{\mathbf{WA}}(w) = 0 \end{cases}$$

(and similarly $S_w^{\mathbf{WK}}$) and define the *normalized* matrices

$$\mathbf{WAn} = \text{diag}(\frac{1}{S_w^{\mathbf{WA}}}) \cdot \mathbf{WA}, \quad \mathbf{WKn} = \text{diag}(\frac{1}{S_w^{\mathbf{WK}}}) \cdot \mathbf{WK}$$

In real life networks **WA** (or **WK**) it can happen that some work has no author. In such a case $S_w^{\mathbf{WA}} = \sum_a \mathbf{WA}[w,a] = 0$ which makes problems in the definition of the normalized network **WAn**. We can bypass the problem by setting $S_w^{\mathbf{WA}} = 1$, as we did in the above definition.

Normalized product

Then the *normalized product* matrix is $\mathbf{AKt} = \mathbf{WAn}^T \cdot \mathbf{WKn}$.

Denoting $\mathbf{F}_w = \frac{1}{S_w^{\mathbf{WA}} S_w^{\mathbf{WK}}} \mathbf{H}_w$ the outer product decomposition gets form

$$\mathbf{AKt} = \sum_w \mathbf{F}_w$$

Since

$$T(\mathbf{F}_w) = \begin{cases} 1 & (\text{outdeg}_{\mathbf{WA}}(w) > 0) \wedge (\text{outdeg}_{\mathbf{WK}}(w) > 0) \\ 0 & \text{otherwise} \end{cases}$$

we have further

$$\sum_{a,k} \mathbf{F}[a,k] = \sum_{a,k} \sum_w \mathbf{F}_w[a,k] = \sum_w T(\mathbf{F}_w) = |W^+|$$

where $W^+ = \{w \in W : (\text{outdeg}_{\mathbf{WA}}(w) > 0) \wedge (\text{outdeg}_{\mathbf{WK}}(w) > 0)\}$. In the network **AKt**, the contribution of each work to the bibliography is 1. These contributions are redistributed to arcs from authors to keywords.

Normalizations

in netsWeight

```
normalize_matrix_Markov <- function(M)
normalize_matrix_Newman <- function(M)
normalize_matrix_Balassa <- function(M)
normalize_matrix_activity <- function(M)
```

Example B

For matrices from Example A we get the corresponding diagonal normalization matrices

$$\mathrm{diag}\left(\frac{1}{S_w^{\mathbf{WA}}}\right) = \begin{array}{c c} & \begin{array}{ccccc} w_1 & w_2 & w_3 & w_4 & w_5 \end{array} \\ \begin{array}{c} w_1 \\ w_2 \\ w_3 \\ w_4 \\ w_5 \end{array} & \left[\begin{array}{ccccc} 1/2 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 1/3 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 1/3 \end{array}\right] \end{array}$$

$$\mathrm{diag}\left(\frac{1}{S_w^{\mathbf{WK}}}\right) = \begin{array}{c c} & \begin{array}{ccccc} w_1 & w_2 & w_3 & w_4 & w_5 \end{array} \\ \begin{array}{c} w_1 \\ w_2 \\ w_3 \\ w_4 \\ w_5 \end{array} & \left[\begin{array}{ccccc} 1/2 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 1/3 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1/2 \end{array}\right] \end{array}$$

compute the normalized matrices

Conclusions

outer products such as

$$\mathbf{F}_1 = \begin{array}{c} \\ a_1 \\ a_2 \\ a_3 \\ a_4 \end{array} \begin{array}{c} \begin{array}{cccc} k_1 & k_2 & k_3 & k_4 \end{array} \\ \begin{bmatrix} 1/4 & 1/4 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1/4 & 1/4 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \end{array} \qquad \mathbf{F}_5 = \begin{array}{c} \\ a_1 \\ a_2 \\ a_3 \\ a_4 \end{array} \begin{array}{c} \begin{array}{cccc} k_1 & k_2 & k_3 & k_4 \end{array} \\ \begin{bmatrix} 0 & 1/6 & 0 & 1/6 \\ 0 & 0 & 0 & 0 \\ 0 & 1/6 & 0 & 1/6 \\ 0 & 1/6 & 0 & 1/6 \end{bmatrix} \end{array}$$

and finally the product matrix $\mathbf{AKt} = \mathbf{WAn}^T \cdot \mathbf{WKn} =$

$$\sum_{w=1}^{5} \mathbf{F}_w = \begin{array}{c} \\ a_1 \\ a_2 \\ a_3 \\ a_4 \end{array} \begin{array}{c} \begin{array}{cccc} k_1 & k_2 & k_3 & k_4 \end{array} \\ \begin{bmatrix} 0.50000 & 0.52778 & 0.36111 & 0.27778 \\ 0.25000 & 0.00000 & 0.75000 & 0.00000 \\ 0.25000 & 0.52778 & 0.11111 & 0.27778 \\ 0.00000 & 0.27778 & 0.61111 & 0.27778 \end{bmatrix} \end{array}$$

Conclusions

Normalized co-authorship network

Let $\mathbf{N}$ be the normalized version $\forall p \in W : \sum_{i \in A} n_{pi} \in \{0, 1\}$ obtained from $\mathbf{WA}$ by $n_{pi} = wa_{pi} / \max(1, \text{outdeg}_{WA}(p))$, or by some other rule determining the author's contribution – the *fractional* approach. Then the *normalized co-authorship network* is

$$\mathbf{Cn} = \mathbf{N}^T \cdot \mathbf{N}$$

$cn_{ij} =$ the total contribution of 'collaboration' of authors i and j to works.

It holds $\quad cn_{ij} = cn_{ji} \quad$ and $\quad \sum_{i \in A} \sum_{j \in A} n_{pi} n_{pj} = 1$.

The total contribution of a complete subgraph corresponding to the authors of a work p is 1.

$$\sum_{i \in A} \sum_{j \in A} cn_{ij} = |W|$$

Normalized strict co-authorship network

Newman defined a *strict normalization* $\mathbf{N}'$ obtained from $\mathbf{WA}$ by $n'_{pi} = wa_{pi} / \max(1, \mathrm{outdeg}_{WA}(p) - 1)$. Then the *normalized strict co-authorship network* is

$$\mathbf{Cn}' = \mathbf{N}^T \cdot \mathbf{N}'$$

The diagonal (loops) of the so-obtained network $\mathbf{Cn}'$ is set to 0.

The network $\mathbf{Cn}'$ doesn't consider the contribution of single-author works.

OpenAlex

```
> setwd(wdir <- "C:/Users/vlado/docs/papers/2025/AS/AMC")
> library(httr); library(jsonlite)
> source("https://raw.githubusercontent.com/bavla/Rnet/master/R/Pajek.R"
> source("https://raw.githubusercontent.com/bavla/OpenAlex/main/code/Ope
> sID <- "S61442588"
> R <- OpenAlexSources(sID,step=250)
OpenAlex2Pajek / Sources Mon May 26 05:59:58 2025
...
865 source S61442588 works collected Mon May 26 06:00:00 2025
..
 5231 citing works collected Mon May 26 06:05:46 2025
...
 12530 cited works collected Mon May 26 06:05:54 2025
12758 different works Mon May 26 06:05:54 2025
> csv <- file("worksAMC.csv","w",encoding="UTF-8")
> write(R,sep="\n",file=csv)
> close(csv)
```

```
> OpenAlex2PajekAll(NULL,name="AMC",listF="worksAMC.csv")
OpenAlex2Pajek / All - Start Mon May 26 06:10:05 2025
*** OpenAlex2Pajek / All - Process Mon May 26 06:10:05 2025
Mon May 26 06:13:38 2025  n = 500
Mon May 26 06:16:25 2025  n = 1000
...
Mon May 26 07:19:31 2025  n = 12500
*** OpenAlex2Pajek / All - Data Collected Mon May 26 07:20:44 2025
hits: 12758 works: 137751 authors: 10849 anon: 185 sources: 1192
>>> Citation Cite
>>> publication year
>>> type of publication
>>> language of publication
>>> cited by count
>>> countries distinct count
>>> referenced works
>>> Authorship WA
>>> Sources WJ
>>> Countries WC
>>> Keywords WK
*** OpenAlex2Pajek / All - Stop Mon May 26 07:21:54 2025
```

Analysis

We first clean the networks **Ci**, **WA**, **WJ**, ..., removing multiple links and loops.

The product $\mathbf{u} = \mathbf{A} \cdot \mathbf{v}$ of the network $\mathbf{A}$ with the vector $\mathbf{v}$ is defined as

$$u_i = \sum_{j:(i,j)\in L} A_{ij} \cdot v_j$$

We need the index j of the node representing MZ in the set of journals J. We can get it from $\mathbf{JW} = \mathbf{WJ}^T$

We apply the command "Info/Vertex label -¿ Vertex number [S4210169332]" on the network **JW**. We get $j = 142$ - the index of the node representing MZ.

We start with the set W_j of all works published by the journal j.

$$W_j = \{w : WJ[w, j] > 0\}$$

Let $\mathbf{w}_j$ be its characteristic vector. Then $\mathbf{w}_j = \mathbf{WJ} \cdot [j]$ where $[j]$ is a vector over J having 1 at the jth place. We create the vector $[j]$

Analysis

Next, for the set W_j, we determine the set W_I of citing works and the set W_O of cited works.

$$W_I = \{w : \exists z \in W_j : Ci[w, z] > 0\} \quad \text{and} \quad W_O = \{w : \exists z \in W_j : Ci[z,$$

The vectors $\mathbf{d}_I = \mathbf{Ci} \cdot \mathbf{w}_j$ and $\mathbf{d}_O = \mathbf{Ci}^T \cdot \mathbf{w}_j$

$$d_I(i) = \sum_k Ci[i, k] \cdot w_j(k) \quad \text{and} \quad d_O(i) = \sum_k Ci^T[i, k] \cdot w_j(k) = \sum_k Ci[k$$

count: $d_I(i)$ - how many works from W_j are cited by the work i; and $d_O(i)$ - how many works from W_j are citing the work i.
Inspect the vector dI. We list the largest 20 nodes [+20]. Extract the selected top lines and copy them in a text file in TextPad. Remove the Rank and Vertex columns. We get
Save it to a CSV file "dI.csv". We will collect in R from OpenAlex the additional information about the selected works.
It turns out that the authors' names are not directly accessible as a data field - they are contained inside the field "authorships". To extract them, we use the function authors.

Conclusions

Now we are ready to get the information about the selected works. Some data ("authors" and "title") can be very long. To get a readable report we truncate them.

```
> LI <- read.table("dI.csv",head=FALSE,sep="")
> selW <- paste0("id,language,countries_distinct_count,cited_by_count,",
+    "relevance_score,publication_year,title,authorships")
> RW <- unitsInfo(IDs=LI$V2,units="works",select=selW,order="input")
> rep <- data.frame(id=RW$id,cdc=RW$countries_distinct_count,cby=RW$cite
+ year=RW$publication_year,authors=substr(authors(RW),1,35),title=substr
> rep
```

Some improvements: add source; in names, use only the last name, or initials + last name; ...

We can check selected works - for example

Using the same approach as for $\mathbf{d}_I$ we get

Citations between journals

$$\mathbf{JJ} = \mathbf{WJ}^T \cdot \mathbf{Ci} \cdot \mathbf{WJ}$$

$JJ[i, k] = \#$ of citations of a work from journal i to a work from journal $k \equiv \#$ of times journal i cites journal k.
Figure
$n_{JJ} = 1776$, $m^A_{JJ} = 8382$, and 141 loops.
Using "Network/Info/Line values" we select the threshold $t = 30$.
We make a link cut at level t.
We save the line-cut network in a file "JJlc30.net", using TextPad extract its node names to a file "JJlc30.csv" and apply the "unitsInfo" approach. It has a problem - "unitsInfo" doesn't like the source "Sunknown". We replace it with resource 1 (duplicated). Now it works, but we must manually correct the info about the source "Sunknown". Finally, we replace IDs in the file "JJlc30.net" with journal names and save it as "JJlc30Names.net". We read it into Pajek and draw it.

Conclusions

There are other options to produce an "interesting" subnetwork * Extract the subset of nodes of the link-cut * Determine the set of "interesting" nodes as a weighted degree core (Ps-core)

Citations between authors

$$\mathbf{ACiA} = \mathbf{WA}^T \cdot \mathbf{Ci} \cdot \mathbf{WA}$$

$ACiA[a,b] = \#$ of citations of a work of author a to a work of author $b \equiv \#$ of times author a cites author b.
$n_{ACiA} = 10268$, $m^A_{ACiA} = 119301$, and 701 loops.
Using "Network/Info/Line values" we select the threshold $t = 15$.
We make a link cut at level t.
To get a more readable network visualization, we have to replace IDs with the corresponding author names. The procedure is as in the case of journals.
AMC authors

$$\mathbf{a}_j = \mathbf{WA}^T \cdot \mathbf{w}_j$$

$a_j(a) = \#$ of works in the journal j co-authored by the author a.
Select the network **ACiA** as the First network (to provide labels) and inspect (Info button) the vector $\mathbf{a}_j$. Again we need to replace IDs in the report by author names.

Conclusions

Conclusions

1. Why is Spain the most attractive country?
2. How can the blue between less and high developed countries be reduced?
3. This is exploratory network analysis. Collect and use additional data (neighbors relation, population size, GDP, etc.).
4. Temporal version of the network.

Vladimir Batagelj, Analysis of the Southern women network using fractional approach, Social Networks, Volume 68, 2022, Pages 229-236, ISSN 0378-8733, https://doi.org/10.1016/j.socnet.2021.08.001. socnet

Acknowledgments

The computational work reported in this presentation was performed using R and Pajek. The code and data are available at GitHub/Vlado.

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

References I

- Batagelj, V, Doreian, P, Ferligoj, A, Kejžar, N (2014). Understanding Large Temporal Networks and Spatial Networks: Exploration, Pattern Searching, Visualization and Network Evolution. Wiley. ISBN: 978-0-470-71452-2
- Batagelj, V (2011). Large-Scale Network Analysis. in John Scott, Peter J. Carrington eds. The SAGE Handbook of Social Network Analysis SAGE Publications.
- Batagelj, V, & Mrvar, A (2003). Density based approaches to network analysis; Analysis of Reuters terror news network. Workshop on link analysis for detecting complex behavior (LinkKDD2003). Retrieved August 27, 2003, from http://www.cs.cmu.edu/dunja/LinkKDD2003/papers/Batagelj.pdf
- Batagelj, V, Zaveršnik, M (2011) Fast algorithms for determining (generalized) core groups in social networks. Advances in Data Analysis and Classification, Volume 5, Number 2, 129–145.
- Matveeva, N, Batagelj, V, Ferligoj, A (2023). Scientific collaboration of post-Soviet countries: the effects of different network normalizations Scientometrics, Volume 128, issue 8, Pages: 4219 – 4242

References II

- Schvaneveldt, RW (Editor) (1990). Pathfinder associative networks: Studies in knowledge organization. Norwood, NJ: Ablex. book
- Schvaneveldt, RW, Dearholt, DW, Durso, FT (1988). Graph theoretic foundations of Pathfinder networks. Computers and Mathematics with Applications, 15, 337-345.
- Vavpetič, A, Batagelj, V, Podpečan, V (2009). "An implementation of the Pathfinder algorithm for sparse networks and its application on text network." In 12th International Multiconference Information Society, vol. A, pp. 236-239.