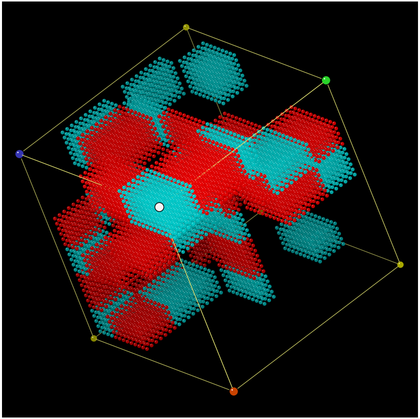
Clustering in multiway networks

Vladimir Batagelj
IMFM Ljubljana and IAM UP Koper

Compstat 2024
27-30 August 2024, Giessen

Outline



Vladimir Batagelj: vladimir.batagelj@fmf.uni-lj.si
Current version of slides (August 28, 2024 at 01 : 35): slides PDF
`https://github.com/bavla/ibm3m`

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 $w : \mathcal{L} \to S_w$.

We will illustrate the proposed approach by analyzing selected data from ESS - European Social Survey 2023. The approach is supported by the R package MWnets.

ESS10 – Media and social trust

ESS10 (2020) Democracy, Digital social contacts / Media and social trust

- `cntry` - Country
- `ppltrst` - Most people can be trusted (10) or you can't be too careful (0)
- `pplfair` - Most people try to take advantage of you (0), or try to be fair (10)
- `pplhlp` - Most of the time people helpful (10) or mostly looking out for themselves (0)

77 Refusal; 88 Don't know; 99 No answer

ESS10 – Media and social trust

```
> str(MM)
List of 6
 $ format: chr "MWnets"
 $ info  :List of 4
  ..$ network: chr "ESSmedia-ppl"
  ..$ title  : chr "ESS media 2023"
  ..$ by     : chr "DF2MWN"
  ..$ date   : chr "Sat Aug 17 02:10:10 2024"
 $ ways  :List of 4
  ..$ cntry  : chr "cntry"
  ..$ ppltrst: chr "ppltrst"
  ..$ pplfair: chr "pplfair"
  ..$ pplhlp : chr "pplhlp"
 $ nodes :List of 4
  ..$ cntry  :'data.frame':     22 obs. of  1 variable:
  .. ..$ ID: chr [1:22] "BE" "BG" "CH" "CZ" ...
  ..$ ppltrst:'data.frame':     14 obs. of  1 variable:
  .. ..$ ID: chr [1:14] "0" "1" "2" "3" ...
  ..$ pplfair:'data.frame':     14 obs. of  1 variable:
  .. ..$ ID: chr [1:14] "0" "1" "2" "3" ...
  ..$ pplhlp :'data.frame':     14 obs. of  1 variable:
  .. ..$ ID: chr [1:14] "0" "1" "2" "3" ...
 $ links :'data.frame': 37611 obs. of  5 variables:
  ..$ one    : num [1:37611] 1 1 1 1 1 1 1 1 1 1 ...
  ..$ cntry  : int [1:37611] 1 1 1 1 1 1 1 1 1 1 ...
  ..$ ppltrst: int [1:37611] 7 4 7 8 4 5 2 7 8 9 ...
  ..$ pplfair: int [1:37611] 8 5 9 6 9 5 6 5 9 9 ...
  ..$ pplhlp : int [1:37611] 5 4 6 6 9 5 4 4 3 6 ...
 $ data  : list()
> table(M$cntry)
  BE   BG   CH   CZ   EE   FI   FR   GB   GR   HR   HU
1341 2718 1523 2476 1542 1577 1977 1149 2799 1592 1849
  IE   IS   IT   LT   ME   MK   NL   NO   PT   SI   SK
1770  903 2640 1659 1278 1429 1470 1411 1838 1252 1418
>
```

Blockmodeling

Clustering in multiway networks

V. Batagelj

Multiway networks

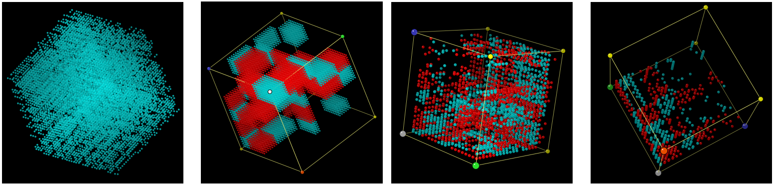
ESS10

Clustering

Example

Conclusions

References



I already worked on the indirect approach to blockmodeling of binary 3-way networks (Github bavla/ibm3m) [Batagelj etal.(2007)]. Lazega, Krackhardt [Borgatti and Everett(1992)]

We were dealing with 3-way networks also at the INSNA 2009 Viszards session analyzing the Bibsonomy data.

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 $\mathbf{C} = [c[u, t]] = \text{row}(\mathbf{W}) = \mathbf{W}\mathbf{W}^T$; and to the set $\mathcal{V}$ by the matrix $\text{col}(\mathbf{W}) = \mathbf{W}^T\mathbf{W}$.

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

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

Projection to a selected way

To make notation simple, we 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

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

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

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
$p(u,u) = p(t,t) = 0 \Rightarrow p(u,t) = 0$.

Salton index

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 $\mathcal{V}_1$.

The Salton index has the following properties

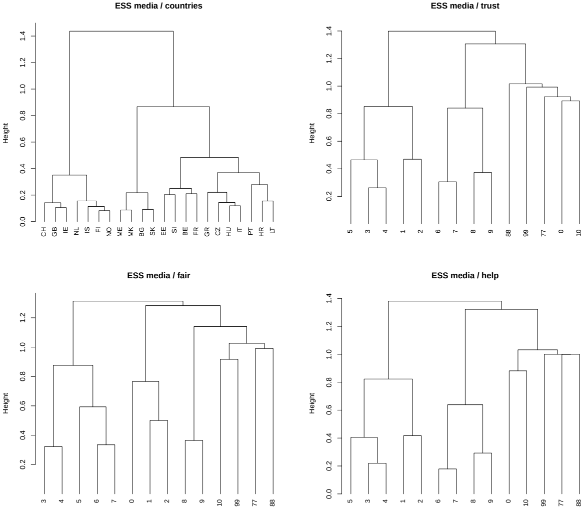
1. $S(u,t) \in [-1,1]$
2. $S(u,t) = S(t,u)$
3. $S(u,u) = 1$
4. $w : L \to \mathbb{R}_0^+ \Rightarrow S(u,t) \in [0,1]$
5. $S(\alpha u, \beta t) = S(u,t), \quad \alpha, \beta > 0$
6. $S(\alpha u, u) = 1, \quad \alpha > 0$

Clusterings of 3-way networks

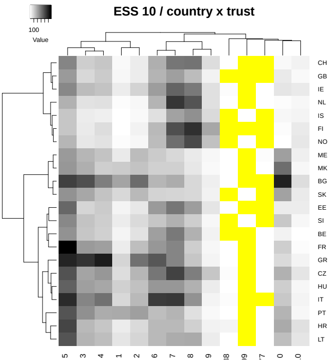
- Italian students mobility - entire network
- Complete European Airports network; Core
- Summer Olympics till 2016 sum core

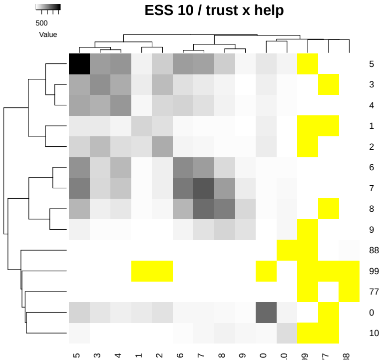
How do we deal with clusterings of a larger number of ways?

Clusterings



Views






```
MCT <- flatten(MN,"one",c("cntry","ppltrst"))
MTH <- flatten(MN,"one",c("ppltrst","pplhlp"))
```

yellow cells have value NA

Partitions

From dendrograms we obtained partitions in 6, 5, 5, 5 clusters

```
Country:
BE BG CH CZ EE FI FR GB GR HR HU IE IS IT LT ME MK NL NO PT SI SK
 1  2  3  4  1  5  1  3  4  6  4  3  5  4  6  2  2  5  5  6  1  2
Trust:
 0  1  2  3  4  5  6  7  8  9 10 77 88 99
 1  4  4  2  2  2  3  3  5  5  1  1  1  1
Fair:
 0  1  2  3  4  5  6  7  8  9 10 77 88 99
 1  1  1  5  5  2  2  2  3  3  4  4  4  4
Help:
 0  1  2  3  4  5  6  7  8  9 10 77 88 99
 1  5  5  2  2  2  3  3  3  3  1  4  4  4
```

C_{ij} is the j-th cluster in the partition $\mathcal{C}_i$ of the way $\mathcal{V}_i$.

$$C_{13} = \{CH, GB, IE\}$$

Block density

$n_i = |\mathcal{V}_i|$ is the size of the i-th way $\mathcal{V}_i$.
$m = |\mathcal{L}|$ is the number of links $\mathcal{L}$.

$\mathbf{j} = (j_1, j_2, \ldots, j_k)$ is a *selection* of clusters from partitions of all ways. We define a *block* of the selection $\mathbf{j}$ as $B(\mathbf{j}) = \times_i C_{ij_i}$ and its *volume* $\mathrm{vol}(B(\mathbf{j})) = \prod_i |C_{ij_i}|$.

$B = \times_i \mathcal{V}_i$ and its *volume* $\mathrm{vol}(B) = \prod_i |\mathcal{V}_i| = \prod_i n_i$.
$m(\mathbf{j}) = |\mathcal{L} \cap B(\mathbf{j})|$ is the number of links in the block $B(\mathbf{j})$.

The *density* $D(\mathbf{j})$ of a block $B(\mathbf{j})$ is then $m(\mathbf{j}) = D(\mathbf{j}) \frac{m}{\mathrm{vol}(B)} \mathrm{vol}(B(\mathbf{j}))$ or

$$D(\mathbf{j}) = \frac{m(\mathbf{j}) \cdot \mathrm{vol}(B)}{m \cdot \mathrm{vol}(B(\mathbf{j}))}$$

An alternative considering the weight w of links is to use $m = \sum_{e \in \mathcal{L}} w(e)$ and $m(\mathbf{j}) = \sum_{e \in \mathcal{L} \cap B(\mathbf{j})} w(e)$.

We inspect the blocks with the largest density.

Top 40 blocks

```
 1 30.22031 1205          11 13.21204 889          21 8.777685 175           31 6.95527 208
1 cntry FI+IS+NL+NO      1 cntry BE+EE+FR+SI      1 cntry BG+ME+MK+SK       1 cntry BG+ME+MK+SK
2 ppltrst 8+9            2 ppltrst 3+4+5          2 ppltrst 1+2             2 ppltrst 1+2
3 pplfair 8+9            3 pplfair 5+6+7          3 pplfair 3+4             3 pplfair 3+4
4 pplhlp 6+7+8+9         4 pplhlp 3+4+5           4 pplhlp 1+2              4 pplhlp 3+4+5

 2 25.61412 1149          12 13.10801 784          22 8.530609 287           32 6.925547 233
1 cntry CZ+GR+HU+IT      1 cntry FI+IS+NL+NO      1 cntry HR+LT+PT          1 cntry CH+GB+IE
2 ppltrst 3+4+5          2 ppltrst 6+7            2 ppltrst 3+4+5           2 ppltrst 3+4+5
3 pplfair 3+4            3 pplfair 5+6+7          3 pplfair 3+4             3 pplfair 3+4
4 pplhlp 3+4+5           4 pplhlp 6+7+8+9         4 pplhlp 3+4+5            4 pplhlp 3+4+5

 3 22.75324 1531          13 12.46153 559          23 8.322546 420           33 6.696119 267
1 cntry CZ+GR+HU+IT      1 cntry BE+EE+FR+SI      1 cntry CH+GB+IE          1 cntry BE+EE+FR+SI
2 ppltrst 3+4+5          2 ppltrst 6+7            2 ppltrst 3+4+5           2 ppltrst 8+9
3 pplfair 5+6+7          3 pplfair 5+6+7          3 pplfair 5+6+7           3 pplfair 8+9
4 pplhlp 3+4+5           4 pplhlp 3+4+5           4 pplhlp 3+4+5            4 pplhlp 6+7+8+9

 4 20.83237 1246          14 12.08813 482          24 8.292822 248           34 6.568866 442
1 cntry CZ+GR+HU+IT      1 cntry CZ+GR+HU+IT      1 cntry CH+GB+IE          1 cntry CH+GB+IE
2 ppltrst 6+7            2 ppltrst 8+9            2 ppltrst 6+7             2 ppltrst 3+4+5
3 pplfair 5+6+7          3 pplfair 8+9            3 pplfair 8+9             3 pplfair 5+6+7
4 pplhlp 6+7+8+9         4 pplhlp 6+7+8+9         4 pplhlp 6+7+8+9          4 pplhlp 6+7+8+9

 5 17.82288 533           15 11.23544 336          25 7.985185 597           35 6.453688 193
1 cntry CH+GB+IE         1 cntry BG+ME+MK+SK      1 cntry BG+ME+MK+SK       1 cntry CZ+GR+HU+IT
2 ppltrst 8+9            2 ppltrst 1+2            2 ppltrst 0+10+77+88+99   2 ppltrst 6+7
3 pplfair 8+9            3 pplfair 0+1+2          3 pplfair 0+1+2           3 pplfair 3+4
4 pplhlp 6+7+8+9         4 pplhlp 1+2             4 pplhlp 0+10             4 pplhlp 3+4+5
```

. . . Top 40 blocks

```
 6 17.54422 787
1 cntry CZ+GR+HU+IT
2 ppltrst 6+7
3 pplfair 5+6+7
4 pplhlp 3+4+5

 7 16.09025 812
1 cntry HR+LT+PT
2 ppltrst 3+4+5
3 pplfair 5+6+7
4 pplhlp 3+4+5

 8 16.07292 721
1 cntry CH+GB+IE
2 ppltrst 6+7
3 pplfair 5+6+7
4 pplhlp 6+7+8+9

 9 14.06659 631
1 cntry BG+ME+MK+SK
2 ppltrst 3+4+5
3 pplfair 3+4
4 pplhlp 3+4+5

 10 13.98485 941
1 cntry BG+ME+MK+SK
2 ppltrst 3+4+5
3 pplfair 5+6+7
4 pplhlp 3+4+5

 16 11.18528 446
1 cntry FI+IS+NL+NO
2 ppltrst 6+7
3 pplfair 8+9
4 pplhlp 6+7+8+9

 17 10.85089 649
1 cntry BE+EE+FR+SI
2 ppltrst 6+7
3 pplfair 5+6+7
4 pplhlp 6+7+8+9

 18 9.897884 296
1 cntry CZ+GR+HU+IT
2 ppltrst 1+2
3 pplfair 0+1+2
4 pplhlp 1+2

 19 9.749268 328
1 cntry HR+LT+PT
2 ppltrst 6+7
3 pplfair 5+6+7
4 pplhlp 3+4+5

 20 9.318279 418
1 cntry FI+IS+NL+NO
2 ppltrst 6+7
3 pplfair 5+6+7
4 pplhlp 3+4+5

 26 7.824679 234
1 cntry CZ+GR+HU+IT
2 ppltrst 1+2
3 pplfair 3+4
4 pplhlp 3+4+5

 27 7.79124 233
1 cntry CZ+GR+HU+IT
2 ppltrst 3+4+5
3 pplfair 3+4
4 pplhlp 1+2

 28 7.520014 253
1 cntry CH+GB+IE
2 ppltrst 6+7
3 pplfair 5+6+7
4 pplhlp 3+4+5

 29 7.423413 222
1 cntry BG+ME+MK+SK
2 ppltrst 3+4+5
3 pplfair 3+4
4 pplhlp 1+2

 30 7.272939 145
1 cntry CZ+GR+HU+IT
2 ppltrst 1+2
3 pplfair 3+4
4 pplhlp 1+2

 36 6.397957 287
1 cntry BG+ME+MK+SK
2 ppltrst 6+7
3 pplfair 5+6+7
4 pplhlp 3+4+5

 37 6.397957 287
1 cntry HR+LT+PT
2 ppltrst 6+7
3 pplfair 5+6+7
4 pplhlp 6+7+8+9

 38 6.375664 286
1 cntry BE+EE+FR+SI
2 ppltrst 3+4+5
3 pplfair 3+4
4 pplhlp 3+4+5

 39 6.319933 189
1 cntry FI+IS+NL+NO
2 ppltrst 8+9
3 pplfair 8+9
4 pplhlp 3+4+5

 40 6.202897 371
1 cntry FI+IS+NL+NO
2 ppltrst 8+9
3 pplfair 5+6+7
4 pplhlp 6+7+8+9
```

Blocks / Country clusters

	1	12	16	20	39	40
	30.22031	13.10801	11.18528	9.318279	6.319933	6.202897
	1205	784	446	418	189	371
1 cntry	FI+IS+NL+NO					
2 ppltrst	8+9	6+7	6+7	6+7		
3 pplfair	8+9	5+6+7		5+6+7		5+6+7
4 pplhlp	6+7+8+9			3+4+5	3+4+5	

	5	8	23	24	28	32	34
	17.82288	16.07292	8.322546	8.292822	7.520014	6.925547	6.568866
	533	721	420	248	253	233	442
1 cntry	CH+GB+IE						
2 ppltrst	8+9	6+7	3+4+5	6+7	6+7	3+4+5	3+4+5
3 pplfair	8+9	5+6+7	5+6+7		5+6+7	3+4	5+6+7
4 pplhlp	6+7+8+9		3+4+5		3+4+5	3+4+5	

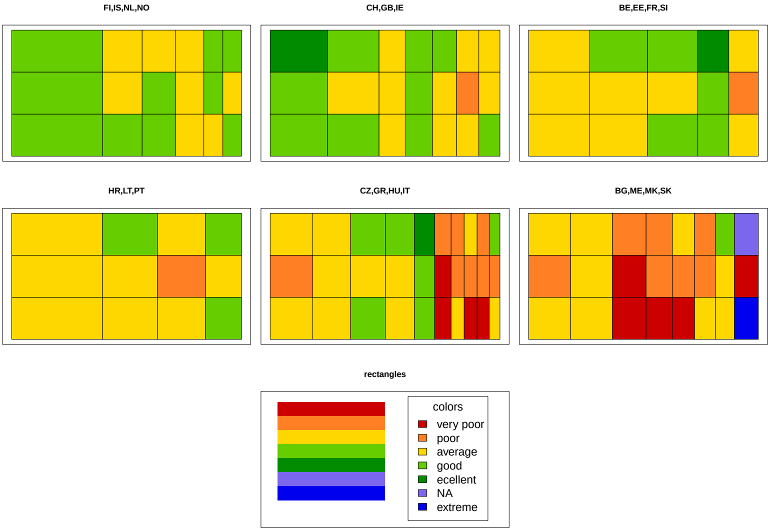
	11	13	17	33	38
	13.21204	12.46153	10.85089	6.696119	6.375664
	889	559	649	267	286
1 cntry	BE+EE+FR+SI				
2 ppltrst	3+4+5	6+7	6+7	8+9	
3 pplfair	5+6+7			8+9	3+4
4 pplhlp	3+4+5		6+7+8+9	6+7+8+9	

	7	19	22	37
	16.09025	9.749268	8.530609	6.397957
	812	328	287	287
1 cntry	HR+LT+PT			
2 ppltrst	3+4+5	6+7		6+7
3 pplfair	5+6+7		3+4	
4 pplhlp	3+4+5			6+7+8+9

	2	3	4	6	14	18	26	27	30
	25.61412	22.75324	20.83237	17.54422	12.08813	9.897884	7.824679	7.79124	7.27
	1149	1531	1246	787	482	296	234	233	145
1 cntry	CZ+GR+HU+IT								
2 ppltrst	3+4+5		6+7	6+7	8+9	1+2	1+2		1+2
3 pplfair	3+4	5+6+7	5+6+7	5+6+7	8+9	0+1+2			
4 pplhlp	3+4+5		6+7+8+9		6+7+8+9	1+2		1+2	1+2

	9	10	15	21	29	31	36	25
	14.06659	13.98485	11.23544	8.777685	7.423413	6.95527	6.397957	7.985185
	631	941	336	175	222	208	287	597
1 cntry	BG+ME+MK+SK							
2 ppltrst	3+4+5		1+2	1+2		1+2	6+7	0+10+77+88+99
3 pplfair	3+4	5+6+7	0+1+2				5+6+7	0+1+2
4 pplhlp	3+4+5		1+2	1+2	1+2			0+10

Flags


FI,IS,NL,NO
CH,GB,IE
BE,EE,FR,SI
HR,LT,PT
CZ,GR,HU,IT
BG,ME,MK,SK
rectangles
colors
very poor
poor
average
good
ecellent
NA
extreme

Conclusions

Work in progress.

- how to consider the weights in ESS?
- more readable presentation/visualization of results
- sparse networks with many ways
- *Jaccard index*

$$J(u,t) = \frac{p(u,t)}{p(u,u) + p(t,t) - p(u,t)}$$

Acknowledgments

Clustering in multiway networks

V. Batagelj

The computational work reported in this presentation was performed in R using the multiway network analysis library *MwNets*. The code and data are available at https://github.com/bavla/ibm3m.

This work is supported in part by the Slovenian Research Agency (research program P1-0294 and research projects J5-2557, J1-2481, and J5-4596), and prepared within the framework of the COST action CA21163 (HiTEc).

References I

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

Batagelj, V, Ferligoj, A, Doreian, P (2007) Indirect Blockmodeling of 3-Way Networks. In: Brito, P, Cucumel, G, Bertrand, P, de Carvalho, F (eds) Selected Contributions in Data Analysis and Classification pp 151–159, Springer.

Borgatti, SP, Everett, MG (1992) Regular blockmodels of multiway, multimode matrices. Social Networks 14: 1-2, 91–120.

Krackhardt, D (1987) Cognitive social structures. Social Networks, 9, 104-134. data.

Lazega, E (2001) The Collegial Phenomenon: The Social Mechanisms of Cooperation Among Peers in a Corporate Law Partnership. Oxford University Press.

Partial projections

In multiway networks with many ways, the probability of co-apperance can be very small - almost all non-diagonal projections $p(u, t)$ are zero.

An option is to use for a way a "partial projections" to "planes" $(\mathcal{V}_a, \mathcal{V}_b), b \in I_a$ where $I_a = 1 : k \setminus \{a\}$

$$\mathbf{M}_{ab} = \mathit{flatten}(M, w, (\mathcal{V}_a, \mathcal{V}_b))$$

we define a projection

$$\mathbf{Q}_a = \sum_{b \in I_a} \mathbf{M}_{ab} \cdot \mathbf{M}_{ab}^T$$

or collecting the planes in a matrix $\mathbf{K}_a = [\mathbf{M}_{a1}, \mathbf{M}_{a2}, \dots, \mathbf{M}_{ak}]$ we get

$$\mathbf{Q}_a = \mathbf{K}_a \cdot \mathbf{K}_a^T$$

In the package MwNets it is implemented with functions `projector2` and `projector`.