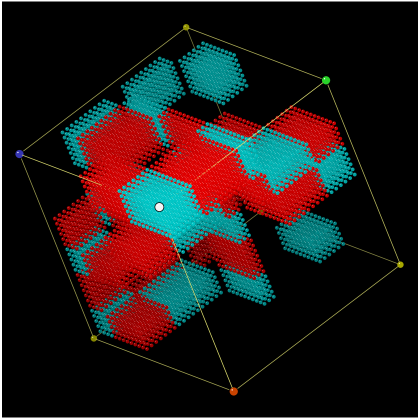
Multiway network analysis

Vladimir Batagelj
IMFM Ljubljana and IAM & FAMNIT UP Koper

NetSlo 2025
January 23, 2025, Ljubljana

Outline



Vladimir Batagelj: vladimir.batagelj@fmf.uni-lj.si
Current version of slides (January 23, 2025 at 05 : 18): slides PDF
https://github.com/bavla/ibm3m

Introduction

In the literature on social network analysis, we find some well-known multi-relational networks, such as CKM physicians innovation (1957), Kapferer tailor shop (1972), Krackhardt office CSS (1987), Lazega law firm (2001), etc. Multi-relational networks are a special case of multi-directional networks. They have been re-examined by physicists studying complex networks, for example, Manlio De Domenico (2015). They call them multiplex networks. Recently, their attention has shifted to networks based on hypergraphs, in which a single link can link several nodes.

In 1992, Borgatti and Everett extended block modeling to general k-directional binary networks, following Baker (1986).

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

For a subset of links $\mathcal{L}' \subseteq \mathcal{L}$ we denote

$$\mathcal{V}_i(\mathcal{L}') = \{e(i) : e \in \mathcal{L}'\}$$

the set of nodes from the way $\mathcal{V}_i$ that are an end-node of a link from $\mathcal{L}'$.

Multiway networks

A multiway network $\mathcal{N}$ is *simple* if and only if all its link tuples are mutually different – for all $e, f \in \mathcal{L}$ it holds $e \neq f \Rightarrow I(e) \neq I(f)$.
A multiway network $\mathcal{N}' = (\mathcal{V}', \mathcal{L}', \mathcal{P}', \mathcal{W}')$ is a *subnetwork* of the multiway network $\mathcal{N} = (\mathcal{V}, \mathcal{L}, \mathcal{P}, \mathcal{W})$ if and only if it can be obtained from $\mathcal{N}$ by removing:

1. some nodes; if a node $v \in \mathcal{V}_i$ is removed then also all links $e \in \mathcal{L}$ that contain it, $e(i) = v$, are removed.
2. some links.
3. some ways.
4. some properties from $\mathcal{P}$ or $\mathcal{W}$.

We could also introduce the notion of homomorphism/ isomorphism of multiway networks.

For a node $u \in \mathcal{V}_i$ we call a *star* in the node u, the set of links

$$S(u) = \{e \in \mathcal{L} : e(i) = u\}.$$

Multiway networks and multi-relational networks

A special kind of multiway network is the multi-relational (multiplex) network. Manlio De Domenico Datasets

In a *multi-relational* network

$$\mathcal{N} = (\mathcal{V}, (\mathcal{L}_1, \mathcal{L}_2, \dots, \mathcal{L}_k), \mathcal{P}, \mathcal{W})$$

the set of links $\mathcal{L}$ is partitioned into different relations (sets of links) $(\mathcal{L}_i)_{i \in I}$ over the same set of nodes. Also the weights from $\mathcal{W}$ are defined on different relations or their union. A default weight is $w(e) = 1$ for all $e \in \mathcal{L}$.

The corresponding multiway network $\mathcal{M} = (\mathcal{V}_M, \mathcal{L}_M, \mathcal{P}_M, \mathcal{W}_M)$ is constructed as follows:

- $\mathcal{V}_M = (\mathcal{V}, \mathcal{V}, (L_1, L_2, \dots, L_k))$, L_i are labels of relations
- $\mathcal{P}_M = \mathcal{P}$
- the link $(u, v) \in \mathcal{L}_i$ with a weight $w_i(u, v)$ produces a link $e = (u, v, i) \in \mathcal{V}_M$ and $w_M(e) = w_i(u, v)$

MWnets

The multiway network analysis is supported by the R package MWnets.

In MWnets the following representation of multiway networks in R is used

```
MW <- list(
   format="MWnets",
   info= metadata,
   ways= list of ways' names,
   nodes= list of ways' dataframes,
   links= dataframe of links and weights,
   data= list of additional data
)
```

Multiway networks are saved/loaded in JSON format. A collection of example networks is available at GitHub/bavla.

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 $(from, to, line) \in \mathcal{L}$ 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.

Dataset

```
List of 6
 $ format: chr "MWnets"
 $ info  :List of 4
  ..$ network: chr "AirEu2013"
  ..$ title  : chr "Air Transportation Multiplex"
  ..$ ref    : chr "Cardillo A. et al. Emergence of network features from multiplexity, Scientif
  ..$ 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" "Cologne Bonn
  .. ..$ 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" "Cologne Bonn Air
  ..$ 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" "Eastern Euro
```

Transformations of multiway networks

In analysis of multiway networks, some transformations could prove to be useful:

- slicing
- reordering of ways
- joining the ways
- flattening of a way
- projection to a selected way
- aggregation by a way partition (blockmodeling)
- normalization
- re-coding (binarization)

MWnets

Derived networks

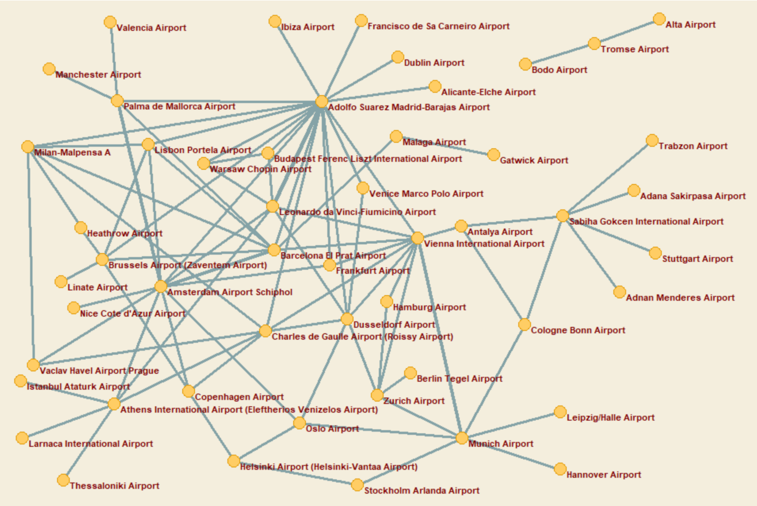
Using transformations of multiway networks we can create ordinary networks and analyze them using network analysis tools.

```
> wdir <- "C:/Users/vlado/DL/data/multi/cores/air"; setwd(wdir)
> library(jsonlite)
> source("https://raw.githubusercontent.com/bavla/Rnet/master/R/Pajek.R")
> source("https://raw.githubusercontent.com/bavla/ibm3m/master/multiway/MWnets.R")
> MN <- fromJSON("https://raw.githubusercontent.com/bavla/ibm3m/master/data/AirEu2013Ext.json")
> str(MN)
> Net <- flatten(MN,"w",c("airA","airB"))
> head(Net$links)
> mwn2net(Net,"airA","airB",ID1="long",ID2="long",w="w",twomode=FALSE,Net="lines.net")
```

We obtained a network of flights between airports with the weight $w(a, b)$ counting the number of companies flying from the airport a to the airport b.

We analyze it in Pajek: draw, transform to undirected (max), info weights, link-cut at level 3, draw; skeletons, cores, islands.

Link-cut at level 3



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* $S(u,t)$ [?]

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

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

To get some insight into the structure of the multiway network we can inspect their 3D layouts in X3D format. They are created using the MWnets function `mwnX3D`.

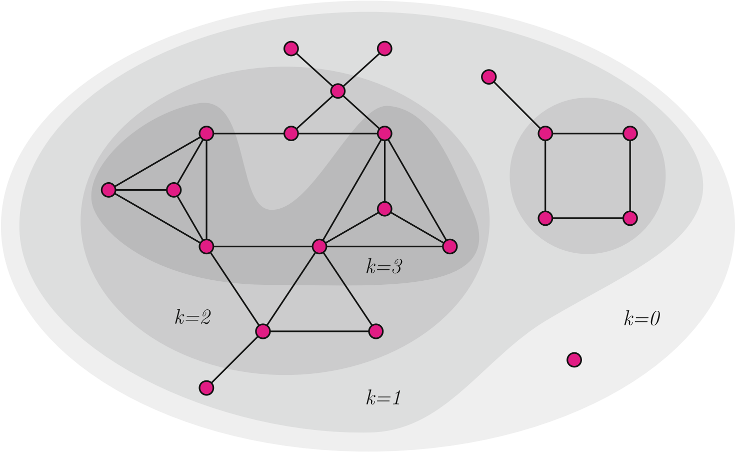
As in the matrix representation of ordinary networks, we can detect internal structure using clustering.

- Italian students mobility - entire network
- Airports dendrogram, Complete European Airports network.
- Summer Olympics till 2016 sum core

Clustering in R

```
> setwd("C:/test/cores")
> library(jsonlite)
> library(Polychrome)
> source("https://raw.githubusercontent.com/bavla/ibm3m/master/multiway/MWnets.R")
> MN <- fromJSON("https://raw.githubusercontent.com/bavla/ibm3m/master/data/AirEu2013Ext.json")
> str(MN)
> MN$nodes$line$short <- substr(MN$nodes$line$ID,1,4)
> Cou <- projection(MN,"airA","w")
> Sau <- salton(Cou); Du <- as.dist(1-Sau); Du[is.na(Du)] <- 1
> tu <- hclust(Du,method="ward")
> plot(tu,hang=-1,cex=0.2,main="EU airports - airports / Ward")
> Coz <- projection(MN,"line","w")
> Saz <- salton(Coz); Dz <- as.dist(1-Saz); Dz[is.na(Dz)] <- 1
> tz <- hclust(Dz,method="ward")
> plot(tz,hang=-1,cex=0.8,main="EU airports - companies / Ward")
> I <- inv(tu$order); K <- inv(tz$order)
> CC <- col2rgb(createPalette(37,c("#ff0000","#00ff00","#0000ff")))/255
> Col <- cbind(CC[1,MN$links$line],CC[2,MN$links$line],CC[3,MN$links$line])
> mwnX3D(MN,"airA","airB","line","w",maxsize=0.85,pu=I,pv=I,pz=K,lu="long",lv="long",
+   col=Col,file="EUair.x3d")
> row.names(Cou) <- MN$nodes$airA$long
> svg("airports.svg",width=18,height=8)
> plot(tu,hang=-1,cex=0.2,main="EU airports - airports / Ward")
> dev.off()
```

Cores in networks



Extended to other node properties, two-mode networks, and temporal networks [2, 3].

Cores in multiway networks

Let $\mathcal{C} = (C_1, C_2, \ldots, C_k) \subseteq \mathcal{V}$ be a selection, $\mathbf{p} = (p_i)$ be a list of monotonic node property functions over the selected ways, and $\mathbf{t} = (t_i)$ a list of the corresponding thresholds.

The multiway subnetwork $Core(\mathbf{p}, \mathbf{t}) = (\mathcal{C}, \mathcal{L}(\mathcal{C}))$ induced by the selection $\mathcal{C}$ in a multiway network $\mathcal{N}$ is a *generalized multiway core* for node property functions $\mathbf{p}$ at levels $\mathbf{t}$ if and only if it holds that for all $v \in \mathcal{C}_i : p_i(v, \mathcal{C}) \geq t_i$ and $\mathcal{C}$ is the maximal such selection.

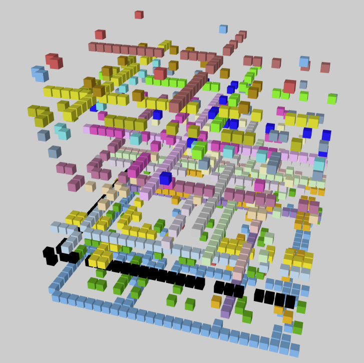
Then for monotonic node property functions $\mathbf{p}$ the generalized core can be determined by the following algorithm

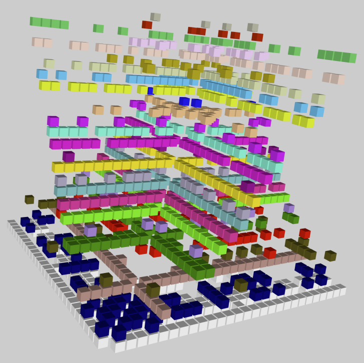
function $\textsc{CoreMW}(\mathcal{N}, \mathbf{p}, \mathbf{t})$
  $\mathcal{C} \leftarrow (\mathcal{V}_{h(1)}, \mathcal{V}_{h(2)}, \ldots, \mathcal{V}_{h(r)})$
  while $\exists u \in C_i : p_i(u, \mathcal{C}) < t_i$ **do** $C_i \leftarrow C_i \setminus \{u\}$
  return $\mathcal{C}$
end function

[1] / arXiv

3D layout of EUair diversity core at level 13

initial (left) and reordered by crosses (right)

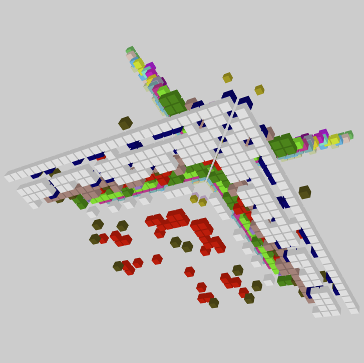


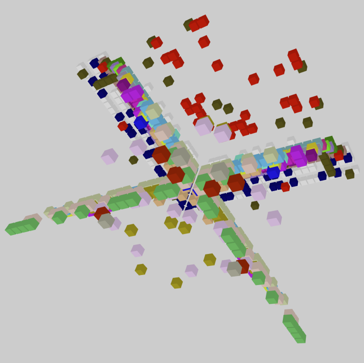


(1) Frankfurt Ap, (2) Berlin Tegel Ap, (3) Munich Ap, (4) Dusseldorf Ap, (5) Hamburg Ap, (6) Zurich Ap, (7) Geneva Ap, (8) Milan-Malpensa Ap, (9) Copenhagen Ap, (10) Stockholm Arlanda Ap, (11) Heathrow Ap, (12) Warsaw Chopin Ap, (13) Vienna IAp, (14) Amsterdam Ap Schiphol, (15) Charles de Gaulle Ap (Roissy Ap), (16) Adolfo Suarez Madrid-Barajas Ap, (17) Barcelona El Prat Ap, (18) Malaga Ap, (19) Vaclav Havel Ap Prague, (20) Leonardo da Vinci-Fiumicino Ap, (21) Brussels Ap (Zaventem Ap), (22) Budapest Ferenc Liszt IAp, (23) Athens IAp (Eleftherios Venizelos Ap), (24) Ben Gurion Ap, (25) Henri Coanda IAp, (26) Venice Marco Polo Ap, (27) Sofia Ap, (28) Nice Cote d'Azur Ap. Core 3D layout

3D layout of EUair diversity core at level 13

two views exposing companies based on multiple airports





(1) Lufthansa, (2) Air Berlin, (3) Swiss IAL, (4) Netjets, (5) Easyjet, (6) SAS, (7) Norwegian AS, (8) British A, (9) LOT Polish A, (10) Austrian A, (11) Niki, (12) KLM, (13) Transavia H, (14) Air France, (15) Iberia, (16) Air Nostrum, (17) Vueling A, (18) Ryanair, (19) Czech A, (20) Alitalia, (21) Brussels A, (22) European AT, (23) Malev HA, (24) Wizz Air, (25) Aegean A (26) Olympic Air, (27) TNT Airways.

The companies that were not providing any line between core airports are Turkish A, Flybe, TAP Portugal, Finnair, Air Lingus, Germanwings, Pegasus A, SunExpress, Air Baltic, and Wideroe.

Acknowledgments

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 partly supported by the Slovenian Research Agency ARIS (research program P1-0294, research program CogniCom (0013103) at the University of Primorska, and research projects J1-2481, J5-2557, and J5-4596), and prepared within the framework of the COST action CA21163 (HiTEc).

References I

- Batagelj, V. Cores in multiway networks. Soc. Netw. Anal. Min. 14, 122 (2024). Springer.
- 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.
- 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.
- Borgatti, SP, Everett, MG (1992) Regular blockmodels of multiway, multimode matrices. Social Networks 14: 1-2, 91–120.
- Cerinšek, M, Batagelj, V (2015) Generalized two-mode cores. Social Networks 42, 80–87.
- Coleman, J, Katz, E, Menzel, H (1957) The Diffusion of an Innovation Among Physicians. Sociometry 20:253–270.
- de Domenico, M (2023) Datasets Released for Reproducibility. `https://manliodedomenico.com/data.php`

References II

- Genova, VG, Giordano, G, Ragozini, G, Vitale, MP (2022) Clustering student mobility data in 3-way networks. IFCS 2022, Porto. Slides.
- Kapferer B (1972) Strategy and transaction in an African factory. Manchester University Press.
- 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.
- Seidman, SB (1983). Network structure and minimum degree. *Social Networks*, 5:269–287.
- x3dom – instant 3D the HTML way! (2023) `https://www.x3dom.org/`