Cores in multiway networks

Vladimir Batagelj

EUSN 2023

The 7th European Conference on Social Networks

Faculty of Social Sciences at the University of Ljubljana

4–8 September 2023

Outline

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

Current version of slides (September 7, 2023 at 03 : 57): slides PDF

https://github.com/bavla/ibm3m/

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

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 $\mathcal{L}'$.

Simple networks and subnetworks

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 network representation in R

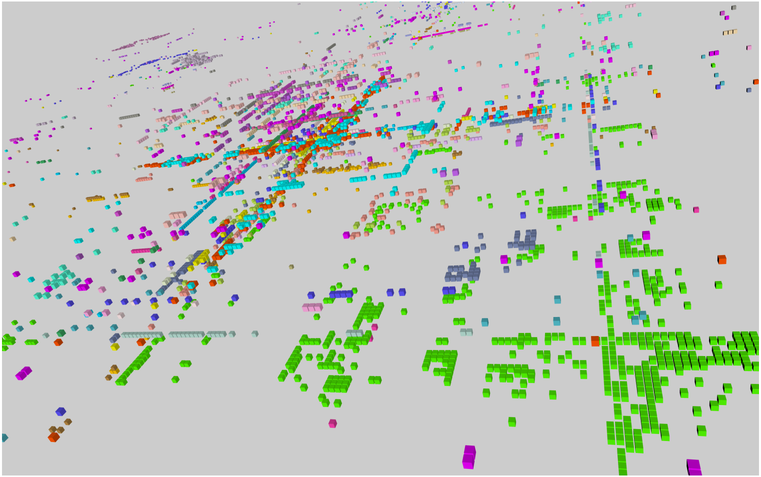
We decided to use for multiway networks the following representation in R

```
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
)
```

Dataset

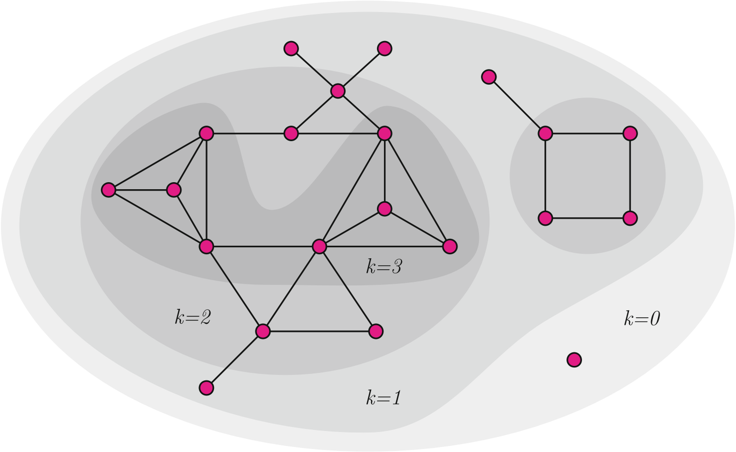
```
List of 6
 $ format: chr "MWnets"
 $ info  :List of 7
  ..$ network: chr "AirEu2013"
  ..$ title  : chr "Air Transportation Multiplex"
  ..$ ref    : chr "Cardillo A. et al. Emergence of network features from multiplexity, Scientif
  ..$ href   : chr "http://complex.unizar.es/~atnmultiplex/"
 $ ways  :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
```

Screenshot of a part of the 3D layout of AirEu2013



3D layout: The green links in the bottom right corner are provided by Ryan air – it is mostly linking airports that are not linked by other companies.

Cores in networks



Extended to other node properties, two-mode networks, and temporal networks [4, 7].

Node properties

A list of subways $\mathcal{C} = (C_1, C_2, \ldots, C_r)$, $C_i \subseteq \mathcal{V}_{h(i)}$ we will call a *selection*. Note that the selection $\mathcal{C}$ defines a function $h : 1..r \to 1..k$. For a link $e \in \mathcal{L}$ we introduce abbreviations

$$h(e) \equiv (e(h(1)), e(h(2)), \ldots, e(h(r)))$$

for a *selected part* of the link e; and

$$h(e) \in \mathcal{C} \equiv \forall i \in 1..r : e(h(i)) \in C_i$$

for a statement that all nodes of the selected part of the link e belong to selected subways.

For a node $u \in \mathcal{V}_i$ and a selection $\mathcal{C}$, we call a *star* in the node u for the selection $\mathcal{C}$, the set

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

... Node properties

For a node $u \in \mathcal{V}_i$ we call the set of *neigbors* of the node u in the way $\mathcal{V}_j$ for the selection $\mathcal{C}$, the set of nodes $N(u, \mathcal{C}, \mathcal{V}_j) = \mathcal{V}_j(S(u, \mathcal{C}))$.

Using stars and sets of neighbors we can define some node property functions on multiway networks. For a node $u \in \mathcal{V}_i$ and a selection $\mathcal{C}$

degree: $p_d(u, \mathcal{C}) = \text{card}(S(u, \mathcal{C}))$

weighted degree for the weight w: $p_S(u, \mathcal{C}; w) = \sum_{e \in S(u,\mathcal{C})} w(e)$

diversity of the way $\mathcal{V}_j$: $p_\delta(u, \mathcal{C}, \mathcal{V}_j) = \text{card}(N(u, \mathcal{C}, \mathcal{V}_j))$

We say that the node property function $p(v, \mathcal{C})$:

- is *local* iff: $p(v, \mathcal{C}) = p(v, N(v, \mathcal{C}))$ for all $v \in \mathcal{V}$.
- is *monotonic* iff: $\mathcal{C}_1 \subset \mathcal{C}_2 \Rightarrow \forall v \in \mathcal{V} : p(v, \mathcal{C}_1) \le p(v, \mathcal{C}_2)$.

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

Multiway cores are partially nested

$$\mathbf{t}' < \mathbf{t} \Rightarrow \mathit{Core}(\mathbf{p}, \mathbf{t}) \subseteq \mathit{Core}(\mathbf{p}, \mathbf{t}')$$

In an elaboration of this algorithm, we have many options. Again the result does not depend on the node elimination order.

In our R library, we included a simple algorithm `MWcore` that eliminates nodes way-wise. The core conditions are given in a list P. Each entry of the list is a triple (p, t, args) where p is a node property function, t is the corresponding level, and args is a list of ways on which the function p is defined. It is assumed that the node u belongs to the way $\mathit{args}[1]$.

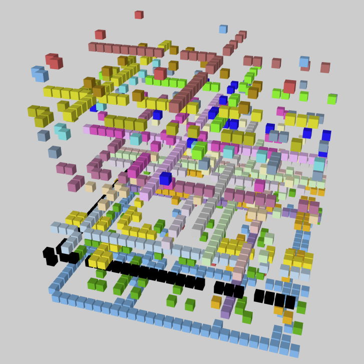
Airports cores

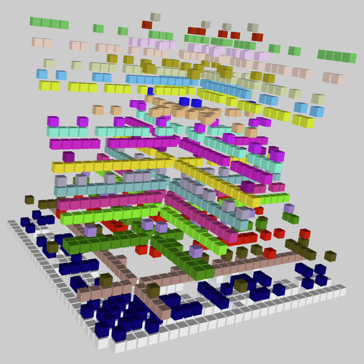
```
> library(jsonlite); library(Polychrome)
> source("https://raw.githubusercontent.com/bavla/ibm3m/master/multiway/MWnets.R")
> MN <- fromJSON("https://raw.githubusercontent.com/bavla/ibm3m/master/data/AirEu2013Ext.json")
> cw <- c("airA","airB")
> ci <- unname(sapply(cw,\(x) which(names(MN$ways)==x)))
> pDIV <- function(MN,v,cip,C,...) pDiv(MN,v,cip,C,way="line")
> P <- list(
+   p1 = list(p = pDIV, t = 10, cwp = c("airA","airB"), cip = NULL),
+   p2 = list(p = pDIV, t = 10, cwp = c("airB","airA"), cip = NULL),
+   cways = list(cw=cw,ci=ci)  )
> for(i in 1:(length(P)-1)) P[[i]]$cip <- unname(sapply(P[[i]]$cwp,\(x) which(cw==x)))
> str(P)
> P[[1]]$t <- 13; P[[2]]$t <- 13
> CC <- MWcore(MN,P)
> listCore(MN,CC,P)
> Ap <- MN$nodes$airA$long[CC[[1]]]

> w1 <- CC$airA; w2 <- CC$airB
> Score <- extract(MN,c("airA","airB"),c("w1","w2"))
> act <- as.integer(names(table(Score$links$line)))
> Rcore <- extract(Score,"line","act")
> str(Rcore)
> c27 <- glasbey.colors(27); CC <- col2rgb(c27)/255
> Col <- cbind(CC[1,Rcore$links$line],CC[2,Rcore$links$line],CC[3,Rcore$links$line])
> ts <- c(1,20,10,26,27,19,24,9,7,8,21,13,18,4,2,15,11,23,6,22,5,16,3,12,14,17,25,28)
> t <- inv(ts)
> qs <- c(1,5,10,19,3,7,12,4,16,13,21,8,20,6,11,18,17,2,23,9,15,24,25,14,22,27,26)
> qq <- inv(qs)
> mwnX3D(Rcore,"airA","airB","line","w",pu=t,pv=t,pz=qq,lu="long",lv="long",maxsize=0.85,
+   col=Col,file="EuAirCore13.x3d")

> percents(MN,Rcore,"airA","airB","line","w")
```

Core 3D layout

3D layout of EUair diversity core at level 13

initial (left) and reordered by crosses (right)



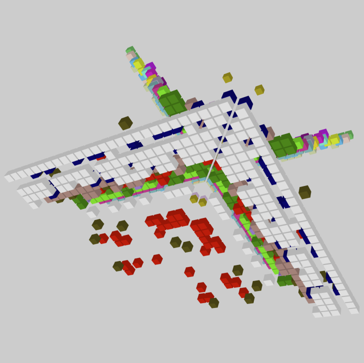


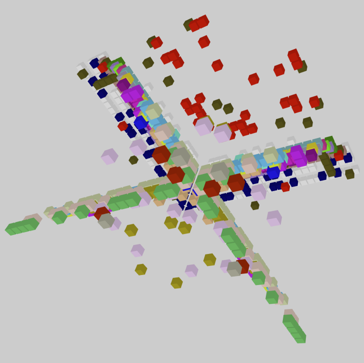
(1) Frankfurt Ap, (2) Berlin Tegel Ap, (3) Munich Ap, (4) Dusseldorf Ap, (5) Hamburg Ap, (6) Zurich Ap, (7) Geneva Ap, (8) Milan-Malpensa Ap, (9) Copenhagen Ap, (10) Stockholm Arlanda Ap, (11) Heathrow Ap, (12) Warsaw Chopin Ap, (13) Vienna IAp, (14) Amsterdam Ap Schiphol, (15) Charles de Gaulle Ap (Roissy 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.

Reordered 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.

Some observations

- Lufthansa and Air Berlin are serving lines from German airports Frankfurt, Berlin, Munich, Dusseldorf, Hamburg, and Milan in Italy
- Swiss IAL and Netjets are serving lines from Swiss airports Zurich and Geneva
- SAS and Norwegian AS are serving lines from Copenhagen and Stockholm
- Spanish airports Madrid, Barcelona, and Malaga are served by companies Iberia, Air Nostrum, Vueling, and Ryanair
- companies serving lines from their base airport: (British A, Heathrow), (LOT, Warsaw), (Austrian A, Niki; Vienna), (KLM, Transavia; Amsterdam), (Air France, CDG), (Czech A, Prague), (Alitalia, Fiumicino), (Brussels A, European AT; Brussels), (Malev, Budapest), (Aegean, Olympic Air; Athens); the only "irregular" link is by Iberia between Barcelona and Budapest

. . . Some observations

- Companies with dispersed services inside the core are Easyjet, Netjets, Wizz Air, European AT, and TNT Airways
  - Easyjet is serving lines from Milan and also from CDG, Amsterdam, Fiumicino, Madrid
  - Netjets is serving lines also from Barcelona, Venice, Vienna, and Nice
  - Wizz Air is serving lines from Budapest and Fiumicino
  - European AT has also some lines from Heathrow, Milan, Venice, Barcelona, and Madrid
  - TNT Airways is linking airports CDG and Athens, and Sofia and Henri Coanda

Olympics

The data set *120 years of Olympic history – athletes and results* with basic bio data on athletes and medal results from Athens 1896 to Rio 2016 is available at Kaggle [10]. The original data describe 134732 participants of the Summer and Winter Olympics. We transformed it into the multiway network `Olympics16S` about the medalists of the Summer Olympics till 2016 (available at GitHub/Bavla [1]). It contains data about 34088 medals. Omitting the athlete Name info and simplifying the network we obtain a multiway network with 10429 links on five ways (Games (29), NOC (147), Sport (52), Sex (2), Medal (3)), the weight w counting the medals, and additional weights Age, Height, and Weight.

We decided to search for a core with a large number of medals on the ways Games, NOC, and Sport. Core

Olympics core

After examing the distributions of values of the node property function p_S on all three ways we decided to use the core conditions list P that contains three triples
$(p = pWsum, t = 100, args = (Games, NOC, Sport))$,
$(p = pWsum, t = 100, args = (NOC, Games, Sport))$, and
$(p = pWsum, t = 100, args = (Sport, Games, NOC))$.

The obtained core was too large. We increased the thresholds to (500, 300, 350). The corresponding p_S-core $\mathcal{C} = (C_{\mathrm{Games}}, C_{\mathrm{NOC}}, C_{\mathrm{Sport}})$ is of order (577, 308, 390) – inside the core, each Olympics accounts for at least 577 medals, each country (NOC) for at least 308 medals, and each sport discipline for at least 390 medals.

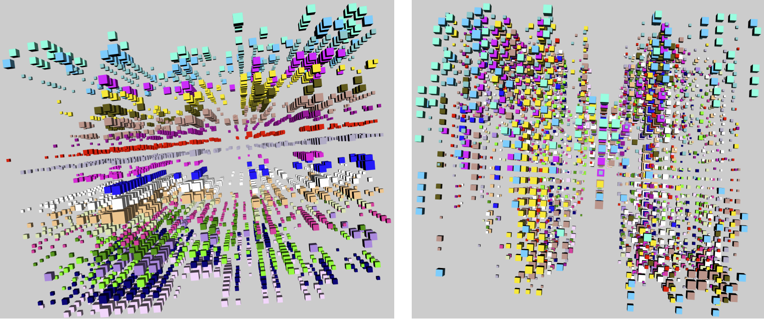
. . . Olympics core

C_{Games} (25): 1988 Seoul (1405), 1984 Los Angeles (1354), 2008 Beijing (1352), 2012 London (1338), 2000 Sydney (1324), 2016 Rio de Janeiro (1320), 2004 Athens (1305), 1980 Moscow (1270), 1996 Atlanta (1251), 1976 Montreal (1220), 1992 Barcelona (1112), 1972 Munich (1080), 1920 Antwerp (1054), 1964 Tokyo (952), 1968 Mexico City (949), 1912 Stockholm (853), 1960 Rome (826), 1952 Helsinki (797), 1956 Melbourne + Stockholm (795), 1936 Berlin (754), 1948 London (731), 1924 Paris (681), 1908 London (656), 1928 Amsterdam (610), 1932 Los Angeles (577).

C_{NOC} (29): USA (4153), URS (2027), GBR (1594), GER (1573), ITA (1325), FRA (1184), AUS (1165), HUN (1052), SWE (1019), NED (854), GDR (843), RUS (776), CHN (662), JPN (661), ROU (632), CAN (613), NOR (546), POL (520), DEN (519), FRG (500), FIN (446), BRA (446), ESP (416), YUG (379), SUI (352), KOR (347), BEL (337), BUL (317), TCH (308).

C_{Sport} (21): Athletics (2911), Swimming (2815), Rowing (2535), Gymnastics (2043), Fencing (1584), Football (1216), Cycling (1129), Hockey (1126), Canoeing (1010), Shooting (995), Sailing (995), Wrestling (941), Handball (924), Basketball (914), Equestrianism (883), Water Polo (876), Volleyball (839), Boxing (596), Judo (425), Weightlifting (419), Diving (390).

Screenshot of a 3D layout of Olympics weighted degree core at levels (577, 308, 390)



It turns out that the core contains $25566/34088 = 75\%$ of all medals, $3607/10429 = 34.6\%$ of all links, and occupies 6.9% of the total space. Note also that the set $\mathcal{C}_{\mathrm{NOC}}$ contains the Soviet Union and Russia; and Germany, West Germany, and East Germany.

The links belonging to the same sport discipline are of the same color. The Olympics are ordered chronologically. The order of NOCs (countries) and sport disciplines was determined using clustering.

Conclusions

- the R package `MWnets` is still in development (additional methods, objects?, dplyr?, robustness)
- additional examples are needed (testing, new problems)
- Python and/or Julia version
- application to multiway networks obtained by conversion of numerical variable(s) to categorical (binning)

References I

- Batagelj, V (2023) GitHub/Bavla, Multiway network analysis in R. `https://github.com/bavla/ibm3m`
- Batagelj, V, Cerinšek, M.: On bibliographic networks. Scientometrics 96 (2013), pages 845–864. Springer
- Batagelj, V.: On fractional approach to the analysis of linked networks. Scientometrics 123 (2020), pages621–633. 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.
- Batagelj, V, Ferligoj, A, Doreian, P: 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, 2007; preprint.
- Borgatti, SP, Everett, MG: Regular blockmodels of multiway, multimode matrices. Social Networks 14(1992)1-2, 91–120.

References II

- Cerinšek, M, Batagelj, V (2015) Generalized two-mode cores. Social Networks 42, 80–87.
- de Domenico, M (2023) Datasets Released for Reproducibility. `https://manliodedomenico.com/data.php`
- Everett, MG, Borgatti, SP: Chapter 9, Partitioning Multimode Networks. In: Doreian, P, Batagelj, V, Ferligoj, A (Eds) Advances in Network Clustering and Blockmodeling , Wiley, 2019.
- rgriffin (2018) 120 years of Olympic history: athletes and results; basic bio data on athletes and medal results from Athens 1896 to Rio 2016 `https://www.kaggle.com/datasets/heesoo37/120-years-of-olympic-history-athletes-and-results`