On the structure of the network of European airports and airlines

Vladimir Batagelj

IMFM Ljubljana and IAM UP Koper

HiTEc Working Groups Meetings

3-5 April 2023, Poseidonia Beach Hotel, Limassol, Cyprus

Outline

Vladimir Batagelj: vladimir.batagelj@fmf.uni-lj.si
Current version of slides (April 2, 2023 at 02 : 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 $(\textit{from}, \textit{to}, \textit{line}) \in \mathcal{L}$ 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.

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
```

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* $\mathcal{L} \subseteq \mathcal{V}_1 \times \mathcal{V}_2 \times \cdots \times \mathcal{V}_k$, and the *weight* $w : \mathcal{L} \to \mathbb{R}$. 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 $w : \mathcal{L} \to S_w$.

Let $e \in \mathcal{L}$ be a link. With $e(i)$ we denote the i-th component of $e = (e(1), e(2), \ldots, e(i), \ldots, e(k))$.

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 $\mathcal{L}'$.

Multi-relational networks

A special kind of multiway networks are multi-relational (multiplex) networks.

In a *multi-relational* network

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

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.

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

- $\mathcal{V}_M = (\mathcal{V}, \mathcal{V}, 1..k)$
- 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 $w_M(e) = w_i(u, v)$

Simple networks

A multiway network $\mathcal{N}$ is *simple* if and only if all its links (tuples) are mutually different.

A multiway network $\mathcal{N}' = ((\mathcal{V}'_1, \mathcal{V}'_2, \ldots, \mathcal{V}'_k), \mathcal{L}', w')$ is a *subnetwork* of the multiway network $\mathcal{N} = ((\mathcal{V}_1, \mathcal{V}_2, \ldots, \mathcal{V}_k), \mathcal{L}, w)$ if and only if $\mathcal{V}'_i \subseteq \mathcal{V}_i$ for all $i \in 1..k$, $\mathcal{L}' \subseteq \mathcal{L}$, $\mathcal{L}' \subseteq \mathcal{V}'_1 \times \mathcal{V}'_2 \times \ldots \times \mathcal{V}'_k$, and for all $e \in \mathcal{L}'$ holds $w'(e) = w(e)$.

The notion of a subnetwork could be extended to the omission of some ways.

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

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

Projection to a selected way

Because of the reordering option, we can 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.

The operation can be generalized to the product of compatible multiway networks.

Salton index has the following properties

From the projection p we can get the corresponding measure of similarity – *Salton index* $S(u,t)$ [2]

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

that can be used for clustering the set $\mathcal{V}_1$.

The Salton index has the following properties

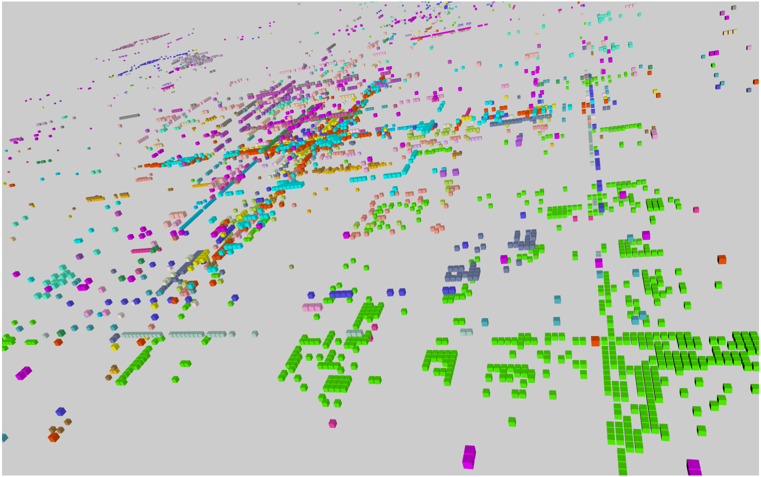
1. $S(u,t) \in [-1,1]$
2. $S(u,t) = S(t,u)$
3. $S(u,u) = 1$
4. $w : L \to \mathbb{R}_0^+ \Rightarrow S(u,t) \in [0,1]$
5. $S(\alpha u, \beta t) = S(u,t), \quad \alpha, \beta > 0$
6. $S(\alpha u, u) = 1, \quad \alpha > 0$

Clustering


```
> 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()
```

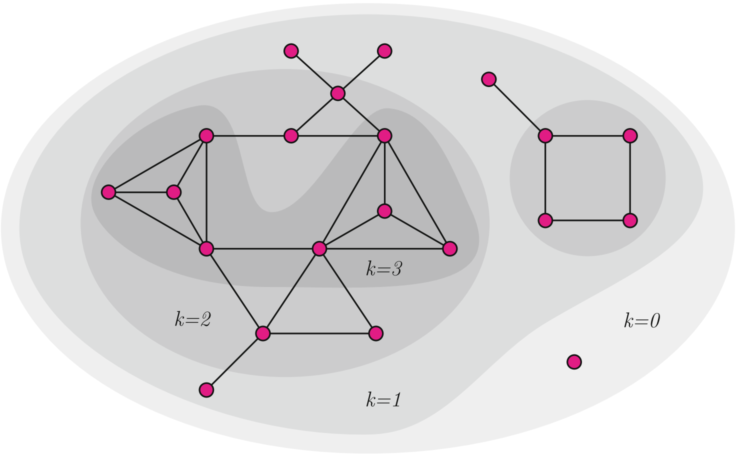
Airports dendrogram, 3D layout.

Screenshot of a part of the 3D layout of `AirEu2013`



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

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 $\mathcal{V}_j$: $p_\delta(u, \mathcal{C}, \mathcal{V}_j) = \mathrm{card}(N(u, \mathcal{C}, \mathcal{V}_j))$

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: $\mathcal{C}_1 \subset \mathcal{C}_2 \Rightarrow \forall v \in \mathcal{V} : p(v, \mathcal{C}_1) \leq p(v, \mathcal{C}_2)$.

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

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

Cores in multiway networks

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. Here we present a simple algorithm that is eliminating 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 $\mathit{args}[1]$.

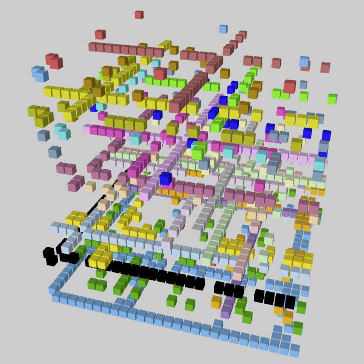
Cores

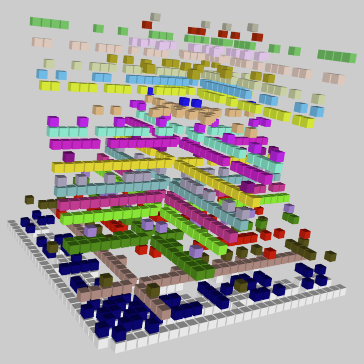
```
> library(jsonlite); library(Polychrome)
> source("https://raw.githubusercontent.com/bavla/ibm3m/master/multiway/MWnets.R")
> MN <- fromJSON("https://raw.githubusercontent.com/bavla/ibm3m/master/data/AirEu2013Ext.json")
> cw <- c("airA","airB")
> ci <- unname(sapply(cw,\(x) which(names(MN$ways)==x)))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



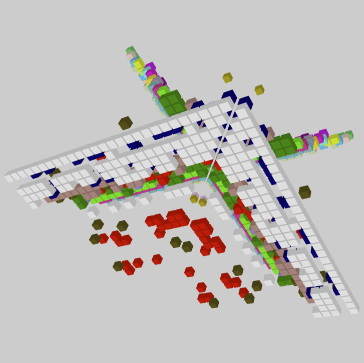


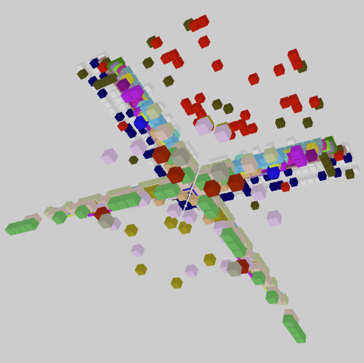
(1) Frankfurt Ap, (2) Berlin Tegel Ap, (3) Munich Ap, (4) Dusseldorf Ap, (5) Hamburg Ap, (6) Zurich Ap, (7) Geneva Ap, (8) Milan-Malpensa Ap, (9) Copenhagen Ap, (10) Stockholm Arlanda Ap, (11) Heathrow Ap, (12) Warsaw Chopin Ap, (13) Vienna IAp, (14) Amsterdam Ap Schiphol, (15) Charles de Gaulle Ap (Roissy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

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)

Acknowledgments

We would like to thank Andreas Plesch for extending the library `X3DOM` with support for `Anchor`'s attribute `description`.

The computational work reported in this presentation was performed using R library MWnets. The code and data are available at Github/Bavla [1].

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 (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.