Analysis of weighted temporal networks represented by time slices

Vladimir Batagelj

IMFM Ljubljana and IAM UP Koper

Compstat 2023

22-25 August 2023, Birkbeck, London

Outline

Vladimir Batagelj: vladimir.batagelj@fmf.uni-lj.si
Current version of slides (August 23, 2023 at 13 : 27): slides PDF
https://github.com/bavla/wNets/

World trade datasets

We will discuss some approaches to **exploratory analysis** of weighted dense temporal networks of moderate size. For illustration, we use world trade data.

WTO and UN comtrade and some other institutions are collecting world trade data. The problem with the collected data is missing data and inconsistencies.

An early version of world trade flows was provided by NBER ([Feenstra etal.(2005)]; download).

The French research center CEPII is providing a consolidated version of the UN comtrade data as a database BACI/CEPII [Gaulier and Zignago(2010)].

An application of the network analysis to the BACI data was done already by de Benedictis et al.
[De Benedictis etal.(2014), Hoang etal.(2023)].

BACI data

Each BACI trade flow is described by 6 quantities:

- Year,
- Product category (HS 6-digit code),
- Exporter (ISO 3-digit country code),
- Importer (ISO 3-digit country code),
- Value of the trade flow (in thousands current USD),
- Quantity (in metric tons)

HS codes: Harmonized Commodity Description and Coding Systems (HS), Classifications on economic statistics, Harmonized System Codes (HS Code 2017 - Current).

Currently, there are around 5300 different HS codes. They are hierarchically organized. The 97 2-digit codes are also called chapters. They are joined into 21 sections. This enables the user to study the data at different granularity levels. In this presentation, we will limit our attention to the top level – the values of total trade flows between countries in a year.

Weighted temporal multi-relational network

The BACI data essentially determine a *weighted temporal multi-relational network* $\mathcal{N} = (\mathcal{V}, (\mathcal{L}_h)_{h \in HS}, w, \mathcal{T})$ where $\mathcal{V}$ is the set of world countries, relation $\mathcal{L}_h$ is the set of links with the HS code h, $w : \mathcal{L} \to \mathbb{R}^+$ is the flow value (in thousands current USD), and $\mathcal{T} = 1996 : 2021$ is the observed range of years.

In a temporal network, nodes $v \in \mathcal{V}$ and links $e \in \mathcal{L}$ are not necessarily present or active in all time points. Let $T(v)$ be the *activity set* of time points for node v and $T(e)$ the activity set of time points for link $e \in \mathcal{L}$. Besides the presence/absence of nodes and links also their properties can change through time – the weight w becomes a *temporal quantity* [Batagelj and Praprotnik(2016)]

$$w(t) = \begin{cases} w'(t) & t \in T(w) \\ \text{⌘} & t \in \mathcal{T} \setminus T(w) \end{cases}$$

where $T(w)$ is the *activity time set* of w, $w'(t)$ is the value of w in an instant $t \in T(w)$, and ⌘ denotes the value *undefined*.

Weighted temporal multi-relational network

A *time slice* of $\mathcal{N}$ in an instant $t \in \mathcal{T}$ is called the network $\mathcal{N}(t) = (\mathcal{V}(t), \mathcal{L}(t), w)$ where $\mathcal{V}(t) = \{v \in \mathcal{V} : t \in T(v)\}$ and $\mathcal{L}(t) = \{e \in \mathcal{L} : t \in T(e)\}$.

They can be also considered weighted multiway network data on four sets of nodes Exporters, Importers, HS codes, Years, and the flow value as a weight.

We converted the BACI data into Pajek format at total, section, and chapter levels (to Pajek, GitHub).

In this presentation, we limit our attention to time slices of the **totally aggregated network**.

Weighted temporal networks

V. Batagelj

World trade datasets

Matrix layout

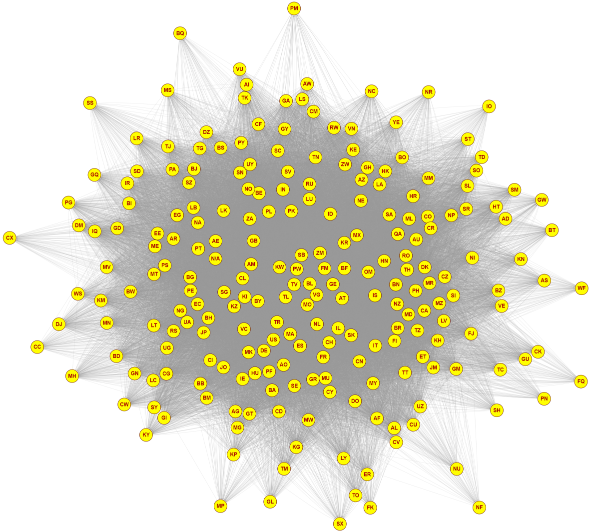
Distributions

Normalizatoins

Skeletons

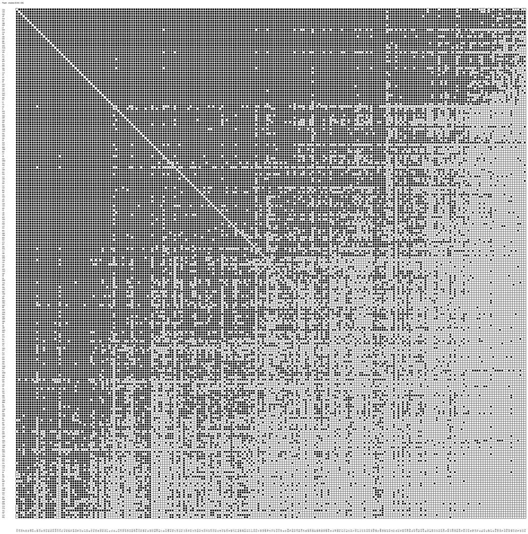
Conclusions

References



Binary matrix layout

nodes ordered by wdeg



Comments

- Larger, $n > 20$, dense graphs can't be presented readably with a graphical layout. For WT2019 the number of nodes is $n = 226$ and the density $\gamma = 0.6231$. For dense graphs of moderate size (up to some hundreds of nodes) a better option is the matrix representation.
- What about weights? They can be represented by link thickness or levels of grey of matrix cells. The problem is a very large range and the distribution of weights – most weights give almost white cells. For WT2019 $w_{\min} = 0.001$ and $w_{\max} = 430317776$. Monotonic transformations such as $w' = a \cdot w$, $a > 0$ or $w' = \sqrt{w}$ or $w' = \log(w)$, etc. In our case, we used $w' = (1000 \cdot w)^{0.08}$.
- A better ordering of rows can be obtained by network clustering [Batagelj etal.(2014)]. Additional reordering of subtrees can be made manually using Pajek by reordering nodes in the hierarchy.

Pajek

network graph and binary matrix layout

```
File/Network/Read [WTtime.net]
Network/Temporal network/Generate in time/Interval [2019,2019] [OK]
File/Network/Change label [WT19]
Draw/Network
save the picture
Network/Create vector/Centrality/Weighted degree/All
Vector/Make permutation
Permutation/Mirror
Network/Create new network/Transform/Line values/Set all to 1
File/Network/Change label [WT19bin]
File/Network/Export as matrix/EPS/Using permutation [WT19bin.eps]
Network/Info/Line values
```

The extracted WT network for the year 2019 has 226 nodes and $w_{\min} = 0.001$ and $w_{\max} = 430317776$. The range is of order 10^{12}. To map it into the interval $[1, 10]$ using the power law transformation $w' = w^k$ we have to select $k \approx 1/12 \approx 0.08$.

Pajek

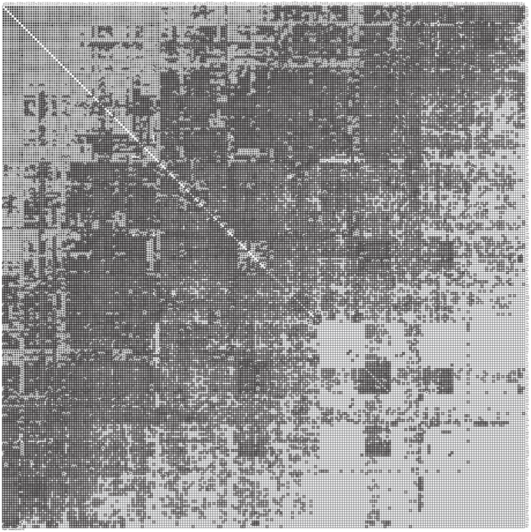
transforming weights

We create two transformed networks, WT19t – transformed WT19, and WT19s – the network WT19 is first symmetrized by adding the weights of opposite arcs and afterward transformed.

```
select the network WT19
Network/Create new network/Transform/Line values/Multiply by [1000]
Network/Create new network/Transform/Line values/Power [0.08]
File/Network/Change label [WT19t]
select the network "Multiply by [1000]"
Network/Create new network/Transform/Arcs -> Edges/All [1]
Network/Create new network/Transform/Line values/Power [0.08]
File/Network/Change label [WT19s]
```

Weighted matrix

order determined by clustering + manual editing



Comments

- center-periphery
- blocks

LC S Lucia	CR Costa Rica	AF Afghanistan	AI Anguilla
VC S Vincent Grenad	DO Dominican R	UZ Uzbekistan	TC Turks Caicos
AG Antigua Barbuda	EC Ecuador	KG Kyrgyzstan	MS Montserrat
BB Barbados	GT Guatemala	MN Mongolia	AS American Samoa
BM Bermuda	UY Uruguay	TJ Tajikistan	GU Guam
GY Guyana	CO Colombia	TM Turkmenistan	TK Tokelau
GD Grenada	PE Peru		KP DPR Korea
BZ Belize	PA Panama		KY Cayman Is
SR Suriname	TT Trinidad Tobago		KN S Kitts Nevis
JM Jamaica	SV El Salvador		VG Br Virgin Is
BS Bahamas	HN Honduras		DM Dominica
	NI Nicaragua		CW Curaçao
	BO Bolivia		AW Aruba

Pajek

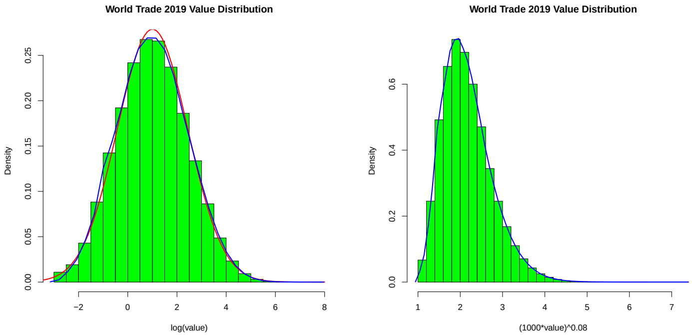
network clustering

```
Select the network WT19s
Cluster/Create complete cluster [226] [OK]
Operations/Network+cluster/Dissimilarity/Network based/d5
   [1] [OK] [dendro.eps]
Select the transformed WT19
File/Network/Export as matrix/Options/Labels up-right
File/Network/Export as matrix/EPS/Using permutation [WT19mat.eps]
```

Double-click on the clustering in the Hierarchy register. Drill into the structure of the clustering. Select a subtree node and using Edit/Push Up or Edit/Push Down change the order of subtrees. When done

```
Hierarchy/Make permutation
File/Network/Export as matrix/EPS/Using permutation [WT19mat.eps]
```

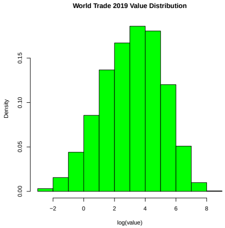
World trade 2019 all flows value distribution

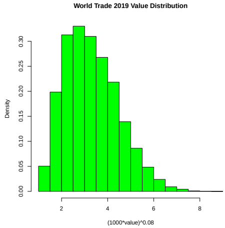


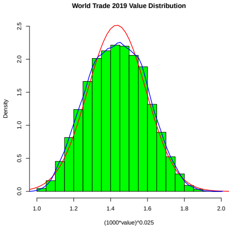
World trade 2019 all flows value distribution


```
> source("https://raw.githubusercontent.com/bavla/Rnet/master/R/Pajek.R"
> C <- read.csv("country_codes_V202301.csv")
> B <- read.csv(unzip("../BACI_HS92_V202301.zip",
+   files="BACI_HS92_Y2019_V202301.csv"))
> str(B)
'data.frame':   10823621 obs. of  6 variables:
...
 $ v: num  1.596 1.114 0.484 6.153 2.399 ...
...
> library(ggplot2)
> hist(log10(B$v),col="green",border="black",prob=TRUE,
+   xlab="log(value)",main="World Trade 2019 Value Distribution")
> m <- mean(log10(B$v)); s <- sd(log10(B$v))
> m
[1] 1.005583
> s
[1] 1.429857
> curve(dnorm(x,m,s),from=-4,to=8,lwd=2,col="red",xaxt="n",yaxt="n",
+   add=TRUE)
> lines(density(log10(B$v),n=32),lwd=2,col="blue")
> hist((1000*B$v)**0.08,col="green",border="black",prob=TRUE,
+   xlab="(1000*value)^0.08",main="World Trade 2019 Value Distribution")
> lines(density((1000*B$v)**0.08,n=64),lwd=2,col="blue")
```

World trade 2019 aggregated flows distribution







World trade 2019 aggregated flows distribution

```
> library(ggplot2)
> P <- net2matrix("WT2019.net");  V <- P[P>0]
> hist(log10(V),col="green",border="black",prob=TRUE,
+   xlab="log(value)",main="World Trade 2019 Value Distribution")
> hist((1000*V)**0.08,col="green",border="black",prob=TRUE,
+   xlab="(1000*value)^0.08",main="World Trade 2019 Value Distribution")
> a <- 0.025; T <- (1000*V)**a
> hist(T,col="green",border="black",prob=TRUE,ylim=c(0,2.5),
+   xlab="(1000*value)^0.025",main="World Trade 2019 Value Distribution"
> m <- mean(T); s <- sd(T)
> m
[1] 1.433429
> s
[1] 0.1585575
> curve(dnorm(x,m,s),from=0.6,to=2.6,lwd=2,col="red",xaxt="n",yaxt="n",
+   add=TRUE)
> lines(density(T,n=64),lwd=2,col="blue")
```

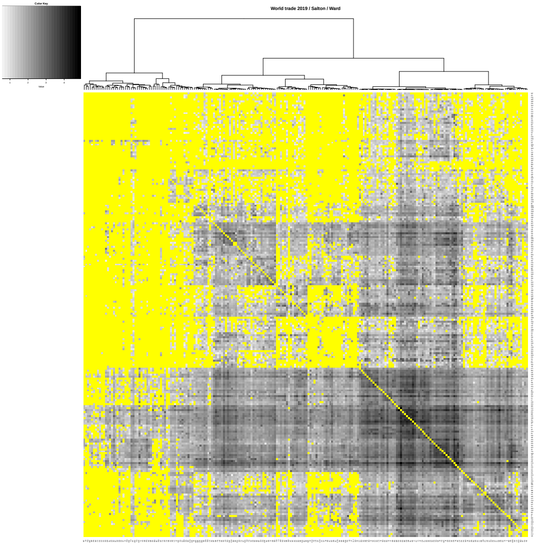
Corrected Salton index

Corrected Salton index of the link $(u, v) \in \mathcal{L}$

$$S[u, v] = \frac{w[u, .] \bullet w[v, .] + (w[u, u] - w[u, v]) \cdot (w[v, v] - w[v, u])}{\sqrt{w[u, .]^2 \cdot w[v, .]^2}}$$

where $\mathbf{x} \bullet \mathbf{y} = \sum_i x_i \cdot y_i$ and $\mathbf{x}^2 = \mathbf{x} \bullet \mathbf{x}$.

World trade 2019 Salton



Salton

```
> source("https://raw.githubusercontent.com/bavla/Rnet/master/R/Pajek.R"
> library(gplots)
> Salton <- function(W,p=1){
+    S <- W; diag(S) <- 1; n = nrow(S)
+    for(u in 1:(n-1)) for(v in (u+1):n) S[v,u] <- S[u,v] <-
+       (as.vector(W[u,]%*%W[v,])+(W[u,u]-W[u,v])*(W[v,v]-W[v,u]))/
+       sqrt(as.vector(W[u,]%*%W[u,])*as.vector(W[v,]%*%W[v,]))
+    return(S)
+ }

> Z <- P <- net2matrix("WT2019.net")
> n <- nrow(P); diag(P) <- 0
> for(u in 1:n) for(v in 1:n) Z[u,v] <- P[u,v]**0.08
> t <- hclust(1-as.dist(Salton(Z)),method="ward.D")
> Z[P == 0] <- NA
> pdf(file="WT19salt.pdf",width=30,height=30)
> heatmap.2(Z,Rowv=as.dendrogram(t),Colv="Rowv",
+   dendrogram="column",scale="none",revC=TRUE,
+   col=colorpanel(30,low="grey95",high="black"),na.color="yellow",
+   trace="none", density.info="none", keysize=0.8, symkey=FALSE,
+   main=paste("World trade ",2019," / Salton / Ward",sep=""))
> dev.off()
```

Normalizations – activity or Balassa index

In networks with weights with a large range usually a few strong nodes prevail. To diminish or neutralize the influence of size on results different normalizations were proposed and used [Batagelj and Mrvar(2003), Matveeva etal.(2023)].

Let $T = \sum_{e \in \mathcal{L}} w(e)$ and for $(u, v) \in \mathcal{L}$ (Balassa index)

$$A(u,v) = \frac{w[u,v] \cdot T}{\text{woutdeg}(u) \cdot \text{windeg}(v)}$$

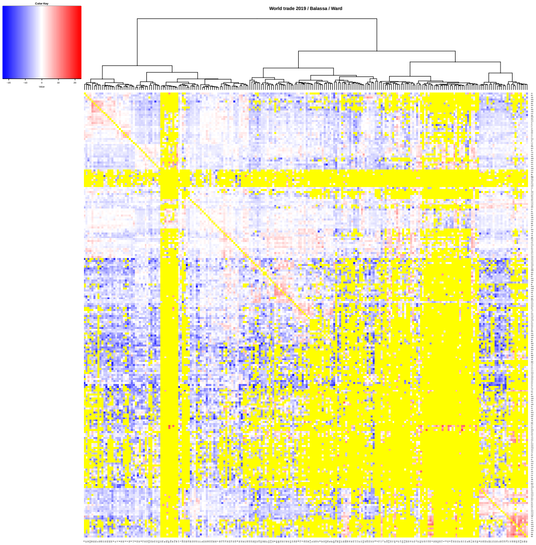
then the *activity normalization* w'

$$w'(u,v) = \log_2 A(u,v)$$

Corrected Euclidean distance

$$D[u,v] = \sqrt{(w[u,v] - w[v,u])^2 + (w[u,u] - w[v,v])^2 + \sum_{t:t \neq u, t \neq v} (w[u,t] - w[v,t])^2}$$

World trade 2019 activity



Activity or Balassa index


```
> source("https://raw.githubusercontent.com/bavla/Rnet/master/R/Pajek.R"
> library(gplots)
> CorEu <- function(W,p=1){
+    D <- W; diag(D) <- 0; n = nrow(D)
+    for(u in 1:(n-1)) for(v in (u+1):n) D[v,u] <- D[u,v] <-
+       sqrt(sum((W[u,]-W[v,])**2) -
+       (W[u,u]-W[v,u])**2 - (W[u,v]-W[v,v])**2 +
+       p*((W[u,u]-W[v,v])**2 + (W[u,v]-W[v,u])**2))
+    return(D)
+ }

> P <- net2matrix("WT2019.net"); diag(P) <- 0
> D <- rowSums(P); T <- sum(D); n <- nrow(P)
> for(u in 1:n) for(v in 1:n) P[u,v] <- P[u,v]*T/D[u]/D[v]
> Z <- log2(P); Z[Z == -Inf] <- 0;
> t <- hclust(as.dist(CorEu(Z)),method="ward.D")
> Z[Z == 0] <- NA
> pdf(file="WT19ba.pdf",width=30,height=30)
> heatmap.2(Z,Rowv=as.dendrogram(t),Colv="Rowv",dendrogram="column",
+   scale="none",revC=TRUE,col=bluered(100),na.color="yellow",
+   trace="none", density.info="none", keysize=0.8, symkey=FALSE,
+   main=paste("World trade ",2019," / Balassa / Ward",sep=""))
> dev.off()
```

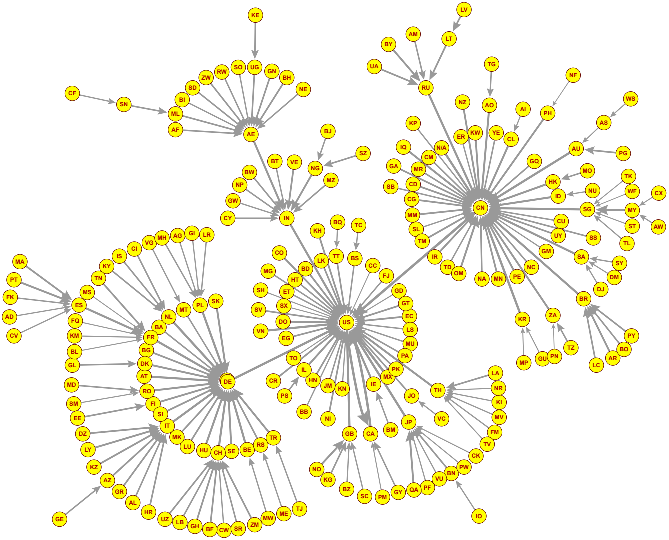
Skeletons

To get insight into the structure of a large network we can reduce it to its skeleton by removing less important links and/or nodes [Batagelj (2011)].

- Most often the spanning tree, link cut, or node cut are used.
- In the closest k-neighbor skeleton for each node only the largest k incident links are preserved. Invariant for monotonic transformations.
- The Pathfinder algorithm was proposed in the 1980s by Schvaneveldt [Schvaneveldt etal.(1988), Schvaneveldt(1990), Batagelj etal.(2014)]. It removes from the network with a dissimilarity weight all links that do not satisfy the triangle inequality – of a shorter path exists that connects the link's end-nodes then the link is removed.
- Cores are a very efficient tool to determine the most cohesive (active) subnetworks [2]. The subset of nodes $\mathbf{C} \subseteq \mathcal{V}$ induces a P_s core at level t if for all $v \in \mathbf{C}$ it holds $\mathrm{wdeg}_{\mathbf{C}}(v) \geq t$, and $\mathbf{C}$ is the maximum such subset.

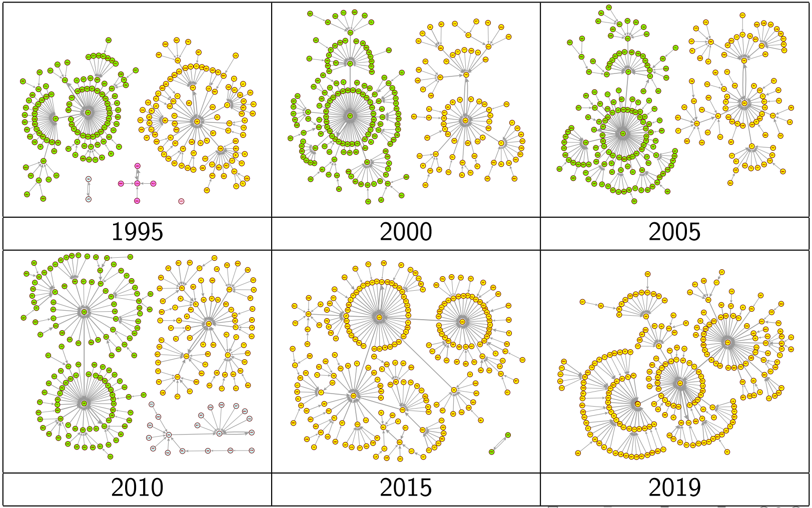
P_s cores

Rank	1995 Id	1995 Value	2019 Id	2019 Value
1	US	257693269.0000	MX	658761422.0000
2	CA	257693269.0000	US	658761422.0000
3	GB	234199078.8000	CN	658761422.0000
4	JP	234199078.8000	CA	658761422.0000
5	IT	234199078.8000	JP	520093954.0000
6	DE	234199078.8000	KR	504015412.0000
7	BE	234199078.8000	DE	504015412.0000
8	FR	234199078.8000	GB	500566393.0000
9	NL	234199078.8000	FR	500566393.0000
10	N/A	162498504.7000	NL	500566393.0000
11	KR	162498504.7000	IT	468723319.0000
12	HK	162498504.7000	BE	452244774.0000
13	CN	162498504.7000	HK	441557136.3000
14	ES	143785957.0000	N/A	435748876.4000
15	CH	136849606.8000	ES	403209053.3000
16	SG	136849606.8000	CH	398697522.1000
17	MX	135153628.7000	VN	398697522.1000
18	MY	126873428.8500	SG	395561634.0000
19	TH	104745321.6000	IN	395561634.0000
20	AT	94004220.3200	AU	395561634.0000
21	SE	93440360.4800	MY	383134139.7000
22	AU	87125103.4300	RU	383134139.7000
23	ID	79999822.3200	PL	359390045.2000
24	BR	70865384.5000	TH	343993122.8000
25	DK	68975678.4900	AE	310082847.8000
26	IE	68975678.4900	ID	310082847.8000
27	NO	68022683.4700	BR	310082847.8000
28	SA	67949085.4900	SA	310082847.8000
29	RU	64736216.2870	CZ	274596220.6000
30	FI	61497981.5330	AT	274174974.0100

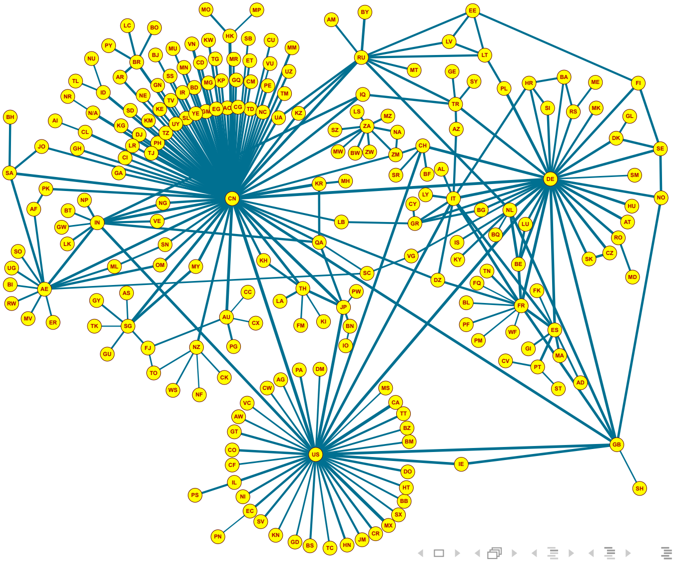


KE
LV
AM
BY
LT
UA
TG
NF
RU
WS
CF
SN
RW
SO
UG
GN
BH
ZW
SD
NE
BI
ML
AF
AE
NZ
AO
AI
PH
KP
ER
KW
YE
CL
AS
BJ
IQ
N/A
CM
AU
PG
GQ
SZ
BT
VE
NG
GA
MR
SB
CD
MO
HK
TK
BW
NP
MZ
CG
NU
ID
WF
CX
GW
CN
SG
MY
AW
CY
IN
MM
ST
SL
CU
KH
BQ
TC
TM
UY
SS
TL
CO
GM
IR
SA
SY
LK
TT
BS
TD
OM
NA
MN
PE
NC
DM
MG
BD
CC
FJ
DJ
HT
SH
ET
GD
BR
SV
SX
GT
US
DO
EC
KR
ZA
VN
EG
LS
PY
MU
TZ
PA
MP
GU
PN
BO
LC
AR
TO
PK
IL
MX
CR
HN
JM
KN
IE
JO
TH
LA
NR
KI
MV
FM
TV
CK
PW
BN
VU
PF
QA
GY
PM
SC
BZ
KG
NO
GB
CA
BM
JP
VC
NI
BB
PS
IO
MA
PT
FK
AD
CV
MS
ES
FQ
KM
BL
GL
MD
SM
EE
DZ
LY
KZ
AZ
GE
GR
AL
HR
UZ
LB
GH
BF
CW
SR
ZM
MW
ME
TJ
TN
KY
IS
CI
VG
MH
AG
GI
LR
SK
PL
MT
NL
BA
FR
BG
DK
AT
RO
FI
SI
IT
MK
LU
HU
CH
SE
BE
RS
TR
DE

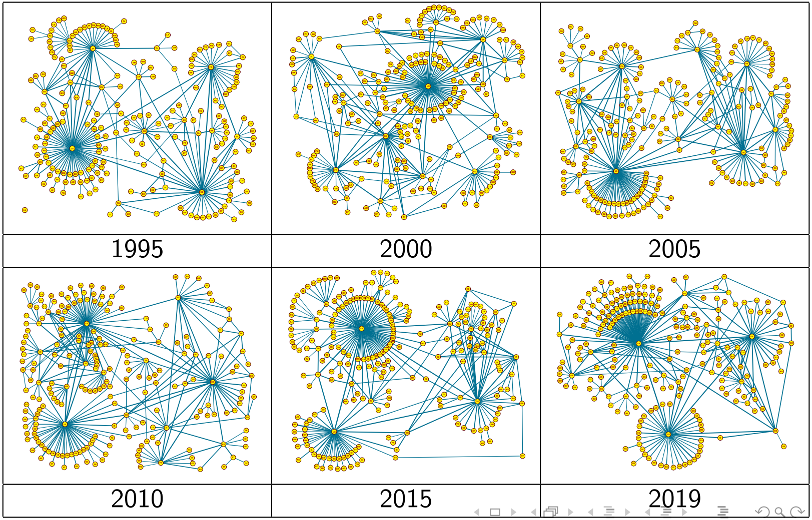
World trade closest neighbor skeletons



WT19s Pathfinder skeleton



World trade Pathfinder skeletons



Pajek

skeletons

```
select the network WT19t
Network/Create new network/Transform/Remove/All arcs except/
   k with highest [1]
Draw network
Kamada-Kawai, manual edit, export picture in SVG
select the network WT19s
Network/Create new network/Transform/Line values/Power [-1]
Network/Create new network/Transform/Reduction/Pathfinder [10] [OK]
select the network WT19s as the second network
Networks/Cross-intersection/Second
Draw network
Kamada-Kawai, manual edit, export picture in SVG
```

Conclusions

- BACI is a gold mine.
- Margin labels in large matrices are not very handy – SVG/JS support with tooltips.
- We intend to look at the data also as a multiway network.

Acknowledgments

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

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

- Vladimir Batagelj, Patrick Doreian, Anuška Ferligoj, Nataša Kejžar (2014) Understanding Large Temporal Networks and Spatial Networks: Exploration, Pattern Searching, Visualization and Network Evolution. Wiley. ISBN: 978-0-470-71452-2
- Batagelj, Vladimir: Large-Scale Network Analysis. in John Scott, Peter J. Carrington eds. The SAGE Handbook of Social Network Analysis SAGE Publications, 2011.
- 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., Praprotnik, S. (2016) An algebraic approach to temporal network analysis based on temporal quantities. *Social Network Analysis and Mining*, 6(1), 1-22
- Batagelj, V.: On fractional approach to the analysis of linked networks. Scientometrics 123 (2020), pages621–633. Springer

References II

- 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.
- De Benedictis, Luca and Nenci, Silvia and Santoni, Gianluca and Tajoli, Lucia and Vicarelli, Claudio, Network Analysis of World Trade Using the BACI-CEPII Dataset (January 2014). Banque de France Working Paper No. 471, Available at SSRN: https://ssrn.com/abstract=2374354 or http://dx.doi.org/10.2139/ssrn.2374354
- Gaulier, G. and Zignago, S. (2010) BACI: International Trade Database at the Product-Level. The 1994-2007 Version. CEPII Working Paper, N 2010-23.
- Hoang, V.P., Piccardi, C. & Tajoli, L. Reshaping the structure of the World Trade Network: a pivotal role for China?. Appl Netw Sci 8, 35 (2023). https://doi.org/10.1007/s41109-023-00560-9
- Robert C. Feenstra, Robert E. Lipsey, Haiyan Deng, Alyson C. Ma and Hengyong Mo: World Trade Flows: 1962-2000. NBER Working Papers 11040, January 2005. https://www.nber.org/papers/w11040

References III

- Nataliya Matveeva, Vladimir Batagelj, Anuška Ferligoj Scientific collaboration of post-Soviet countries: the effects of different network normalizations Scientometrics, Volume 128, issue 8 2023, Pages: 4219 – 4242
- Schvaneveldt, R. W. (Editor) (1990). Pathfinder associative networks: Studies in knowledge organization. Norwood, NJ: Ablex. book
- Schvaneveldt, R. W., Dearholt, D. W., & Durso, F. T. (1988). Graph theoretic foundations of Pathfinder networks. Computers and Mathematics with Applications, 15, 337-345.