Exploring World Trade Networks

Vladimir Batagelj

IMFM Ljubljana, UL FMF Ljubljana, and UP IAM Koper

ARS'25

Tenth International Workshop on Social Network Analysis

30-31 October 2025, Naples (Italy)

Outline

Vladimir Batagelj: vladimir.batagelj@fmf.uni-lj.si
Current version of slides (October 30, 2025 at 09 : 02): slides PDF
https://github.com/bavla/Nets/

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 total trade flow values 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 $\mathcal{T} = 1996 : 2023$ 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 $\mathcal{N}(t) = (\mathcal{V}(t), \mathcal{L}(t), w(t))$ where $\mathcal{V}(t) = \{v \in \mathcal{V} : t \in T(v)\}$ and $\mathcal{L}(t) = \{e \in \mathcal{L} : t \in T(e)\}$.

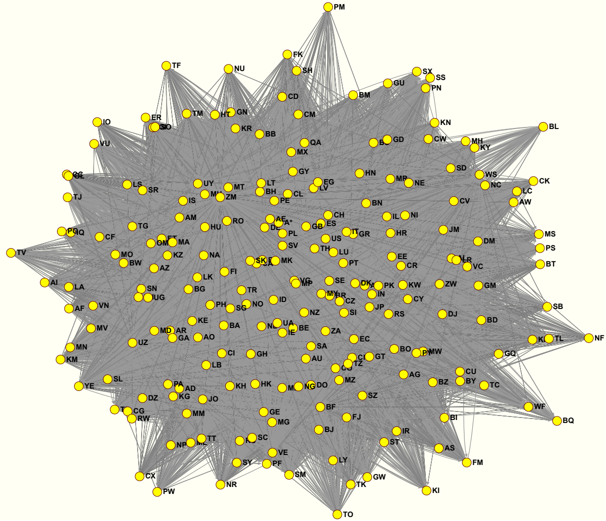
They can also be considered weighted multiway network data on four sets of nodes: Exporters, Importers, HS codes, and Years, with the flow value as a weight.

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

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

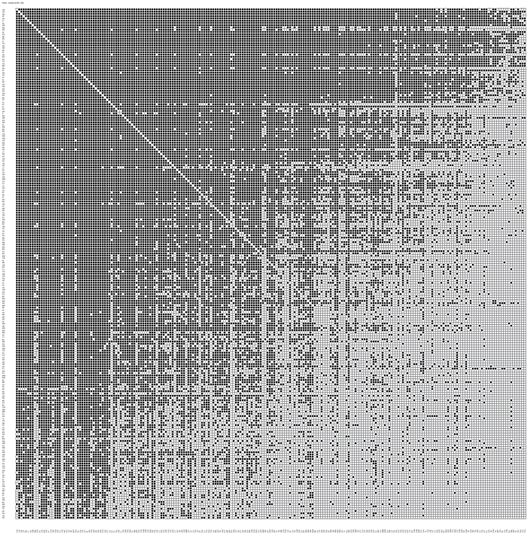
World trade 2023 graph

spring embedder

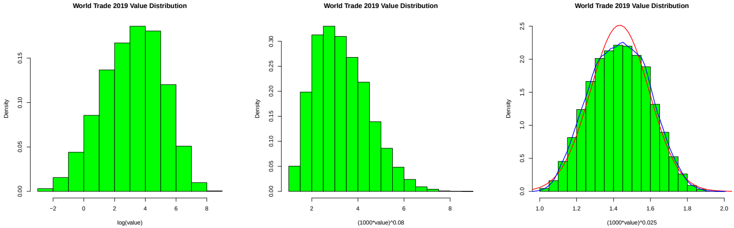


WT 2023 binary matrix layout

nodes ordered by wdeg



World trade 2019 aggregated flows distribution



Comments

- Larger, $n > 20$, dense graphs can't be presented readably with a graphical layout. For WT2023 the number of nodes is $n = 226$ and the density $\gamma = 0.6068$. 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 WT2023 $w_{\text{min}} = 0.001$ and $w_{\text{max}} = 456484584$. 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.

Corrected Euclidean distance and Salton index

Corrected Euclidean distance

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

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

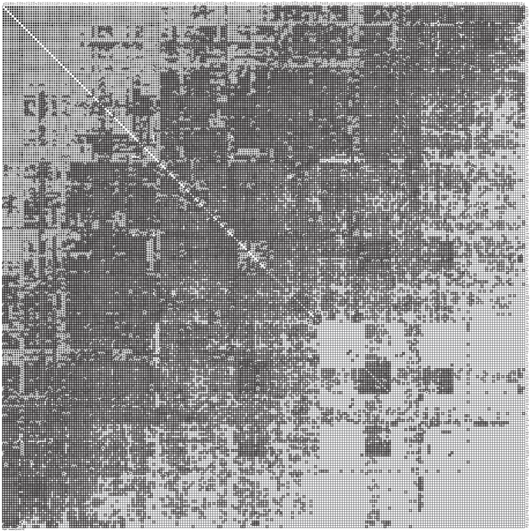
$$S[u,v] = \frac{w[u,.] \bullet w[v,.] + (w[u,u]-w[u,v]) \cdot (w[v,v]-w[v,u])}{\sqrt{w[u,.]^2 \cdot w[v,.]^2}}$$

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

$$S[\alpha u, \beta v] = S[u,v], \text{ for } \alpha, \beta > 0$$

Weighted matrix WT2019

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

Skeletons

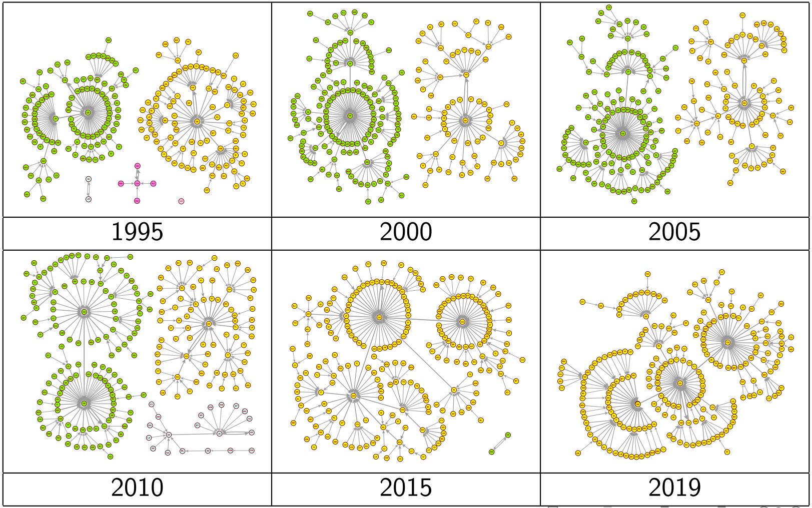
To get insight into the structure of a large network we can reduce it to its skeleton by removing less important links and/or nodes [Batagelj (2011)].

- 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.

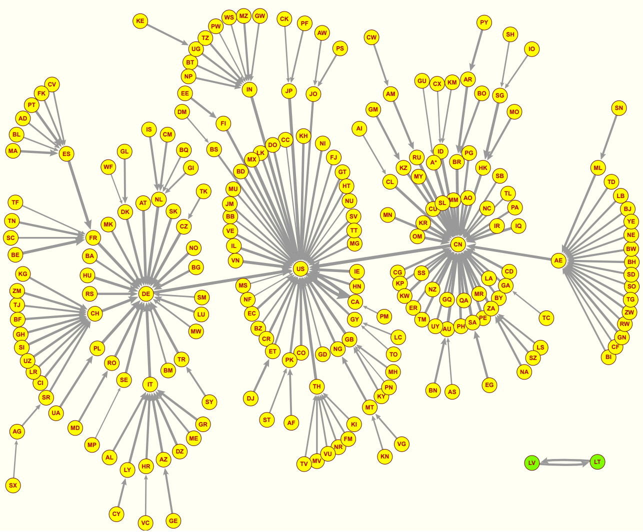
P_s cores

Rank	1995 Id	1995 Value	2019 Id	2019 Value	2023 Id	2023 Value
1	US	257693269	MX	658761422	MX	799627502
2	CA	257693269	US	658761422	US	799627502
3	GB	234199078	CN	658761422	CN	799627502
4	JP	234199078	CA	658761422	CA	799627502
5	IT	234199078	JP	520093954	KR	628044889
6	DE	234199078	KR	504015412	JP	628044889
7	BE	234199078	DE	504015412	DE	628044889
8	FR	234199078	GB	500566393	GB	594923889
9	NL	234199078	FR	500566393	IT	594923889
10	N/A	162498504	NL	500566393	FR	594923889
11	KR	162498504	IT	468723319	NL	594923889
12	HK	162498504	BE	452244774	BE	590202643
13	CN	162498504	HK	441557136	A*	512227117
14	ES	143785957	N/A	435748876	HK	512227117
15	CH	136849606	ES	403209053	CH	511548201
16	SG	136849606	CH	398697522	ES	511548201
17	MX	135153628	VN	398697522	VN	511548201
18	MY	126873428	SG	395561634	IN	511548201
19	TH	104745321	IN	395561634	AU	511548201
20	AT	94004220	AU	395561634	MY	509054674
21	SE	93440360	MY	383134139	SG	509054674
22	AU	87125103	RU	383134139	TH	466862364
23	ID	79999822	PL	359390045	PL	450228447
24	BR	70865384	TH	343993122	AE	446632123
25	DK	68975678	AE	310082847	BR	437245724
26	IE	68975678	ID	310082847	ID	437245724
27	NO	68022683	BR	310082847	RU	435325632
28	SA	67949085	SA	310082847	TR	411617795
29	RU	64736216	CZ	274596220	SA	411617795
30	FI	61497981	AT	274174974	IE	349367822

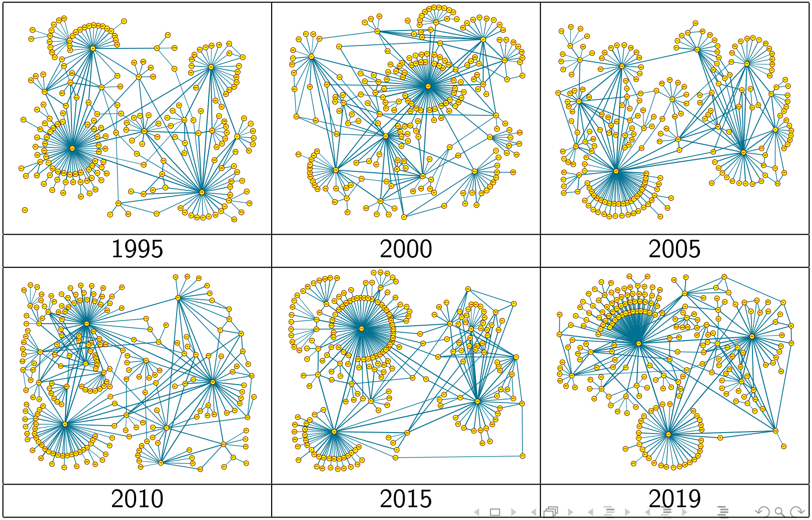
World trade closest neighbor skeletons



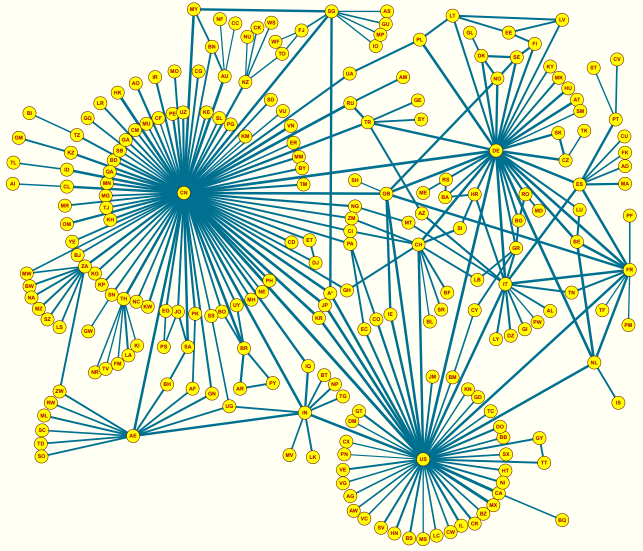
World trade 2023 closest neighbor skeletons



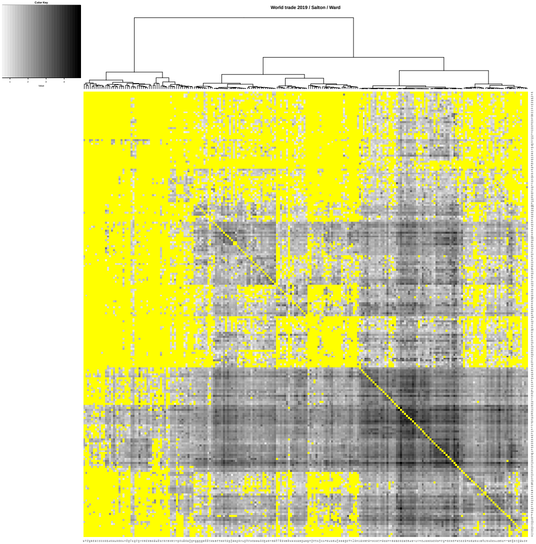
World trade Pathfinder skeletons



WT2023s Pathfinder skeleton



World trade 2019 Salton



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)].

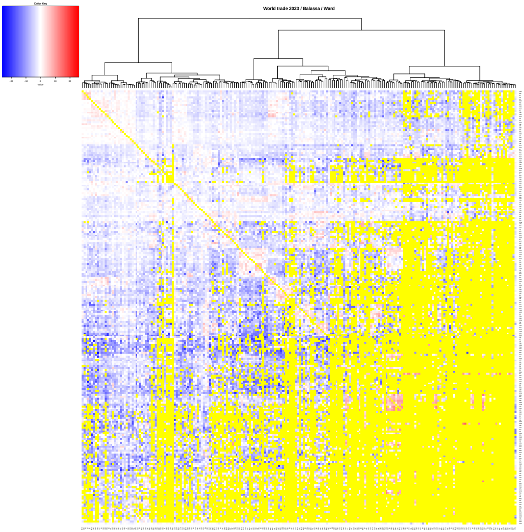
Let $T = \sum_{e \in \mathcal{L}} w(e)$ and for $(u, v) \in \mathcal{L}$ (Balassa index)

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

then the *activity normalization* w'

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

World trade 2023 activity



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 Github/bavla/Nets.

This work is supported in part by the Slovenian Research Agency (research program P1-0294 and research project 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.