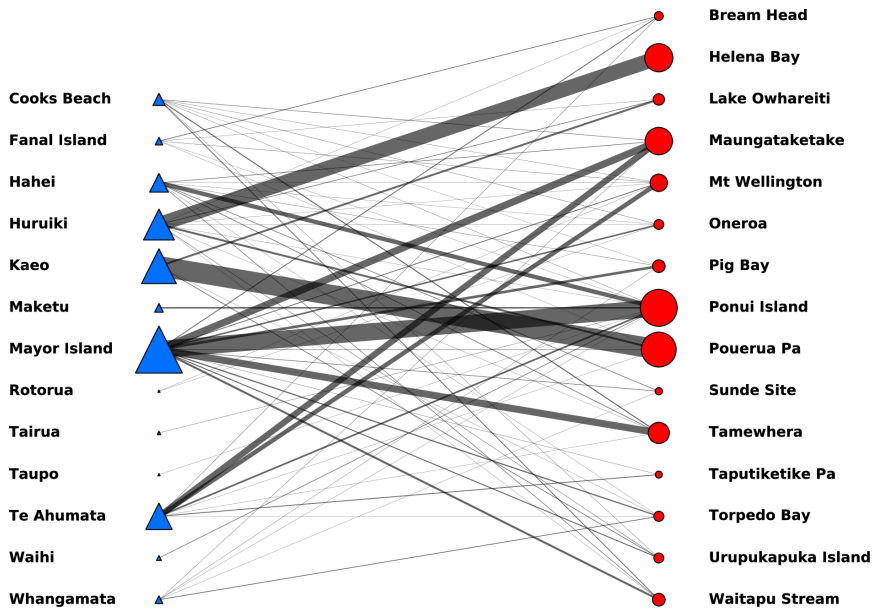
Normalizations and Projections in Analysis of Weighted Two-Mode Networks

Vladimir Batagelj
IMFM Ljubljana, IAM UP Koper, and NRU HSE Moscow

Networks 2021 and XXXXI Sunbelt Social Networks Conference
on Zoom, July 5-10, 2021

Outline



source

Vladimir Batagelj: vladimir.batagelj@fmf.uni-lj.si
Current version of slides (July 5, 2021 at 03 : 50): slides PDF
https://github.com/bavla/NormNet/tree/main/docs

Introduction

There are three main approaches to the analysis of **binary** two-mode networks

- Treat the two-mode network as an ordinary one-mode network (degrees, components, etc.).
- Apply special methods developed for analysis of two-mode networks (two-mode hubs and authorities, two-mode cores, 4-ring weights, etc. [1, 2, 3]).
- Transform (project) the two-mode network to a corresponding one-mode network and use the usual methods to analyze it.

Projections produce weighted networks with a special structure.

. . . Introduction

In the paper

- Batagelj, V, Mrvar, A: Density based approaches to network analysis: Analysis of Reuters terror news network. Workshop on Link Analysis for Detecting Complex Behavior (LinkKDD2003), August 27, 2003. PDF

we proposed some normalizations of projection networks to neutralize large differences in network weights and make nodes comparable.

. . . Introduction

In our papers

- Batagelj, V, Cerinšek, M: On bibliographic networks. Scientometrics 96 (2013) 3, 845-864. PDF
- Batagelj, V.: On fractional approach to analysis of linked networks. Scientometrics 123 (2020) 2: 621-633. PDF

we showed that a projection network is a sum of complete subnetworks (outer-product decomposition) – nodes with a large degree are over-represented in it.

We also proposed some normalized projections that better grasp the network structure.

For this presentation I initially intended to present some generalizations of the above results to **weighted** two-mode networks. Since it is impossible to squeeze them into 15 minutes I decided to present a related handy normalization that is very useful in inspecting weighted (two-mode) networks

Two-mode networks

A simple directed two-mode network $\mathcal{N} = ((U, V), L, w)$ links the set of nodes U to the set of nodes V with the arcs from the set $L \subseteq U \times V$. The mapping $w : L \to \mathbb{R}^+$ assigns to each arc (u, v) its weight $w(u, v)$. The network $\mathcal{N}$ can be represented with the corresponding matrix $\mathbf{W} = [w[u, v]]_{u \in U, v \in V}$

$$w[u, v] = \begin{cases} w(u, v) & (u, v) \in L \\ 0 & \text{otherwise} \end{cases}$$

In traditional two-mode networks we usually assume that $U \cap V = \emptyset$. In the case $U = V$ we get an ordinary one-mode simple directed network.

Degrees, weighted degrees and total weight

We will need some additional functions and quantities:
out-degree $\mathrm{od}(u)$, *in-degree* $\mathrm{id}(v)$,
weighted out/in-degree (row/column sums)

$$\mathrm{wod}(u) = \sum_{v \in V} w[u,v] \qquad \text{and} \qquad \mathrm{wid}(v) = \sum_{u \in U} w[u,v].$$

The *total weight* T of links in the network $\mathcal{N} = (V, L, w)$

$$T = \sum_{(u,v) \in L} w(u,v) = \sum_{u,v} w[u,v] = \sum_{u} \mathrm{wod}(u) = \sum_{v} \mathrm{wid}(v)$$

Deviations

The proportion of activity of the node u is equal to $\text{wod}(u)/T$. The *expected weight* $C[u,v]$ from node u to node v is equal to

$$C[u,v] = \frac{\text{wod}(u)}{T} \cdot \text{wid}(v)$$

The measured weight $w[u,v]$ may deviate from the expected weight $w[u,v] = a[u,v] \cdot C[u,v]$ or

$$a[u,v] = \frac{w[u,v] \cdot T}{\text{wod}(u) \cdot wid(v)}$$

If $a[u,v] > 1$ the measured weight is larger than expected, ...
The factor a is not 'symmetric'. We replace it with b defined as $b = \log(a)$: $(1,\infty) \to (0,\infty)$; $(0,1) \to (-\infty,0)$. The quantity b is defined on $\mathbb{R}$ and is positive for $a > 1$ and negative for $a < 1$. We will use rows/cols of $\mathbf{B} = [\log(a[u,v])]$ as units for clustering.
Note: if the network is one-mode and undirected $\text{wid}(v) = \text{wod}(v)$.

WTO data on international trade

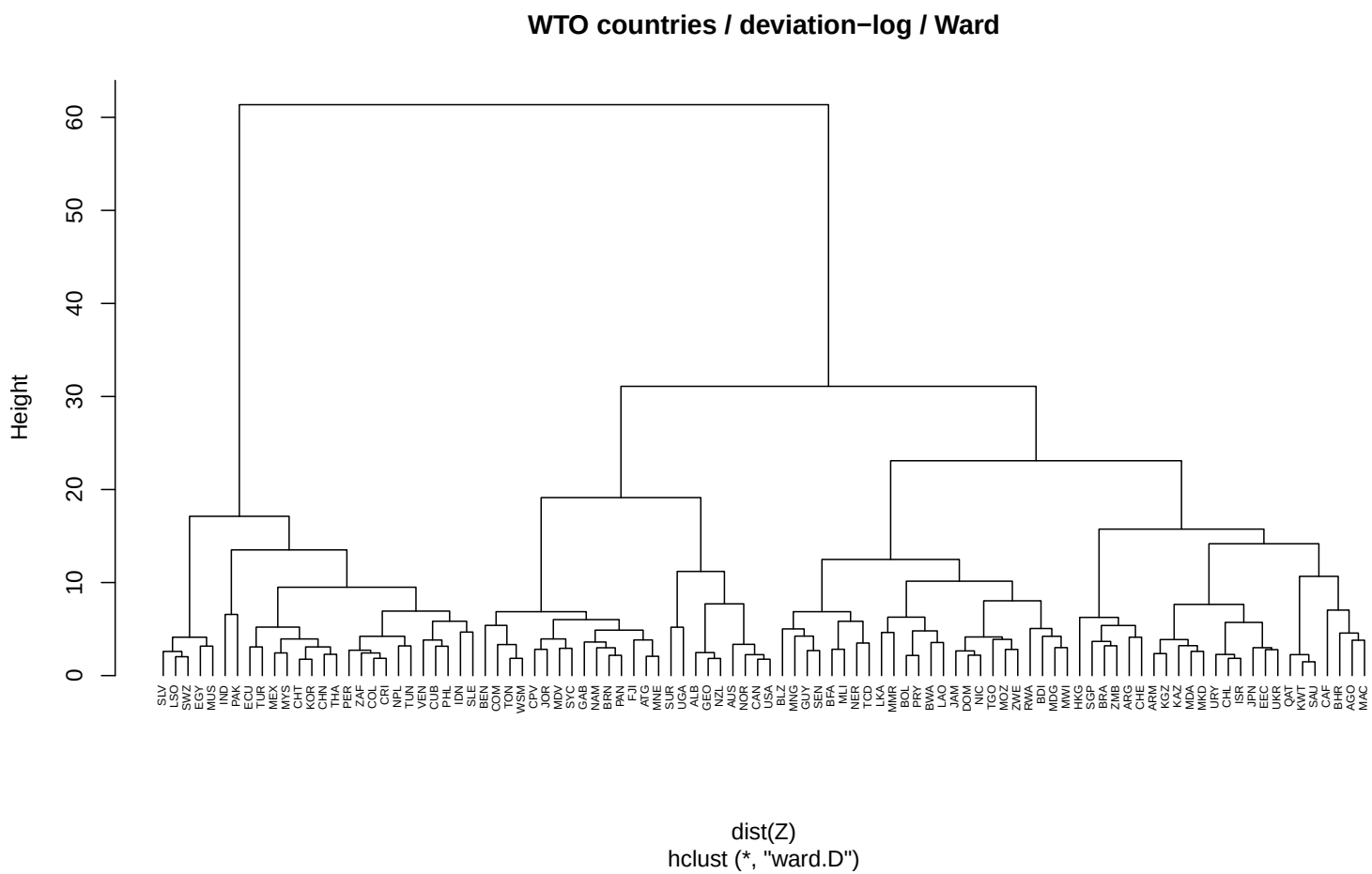
From the WTO database https://data.wto.org/ we downloaded for the year 2015 two data sets that form two weighted two-mode networks (country, sector, value):

- WTO annual bilateral imports by MTN product category (98 countries, 22 sectors);
- WTO annual marchandise export by product group (Million US dollar) (216 countries, 180 sectors);

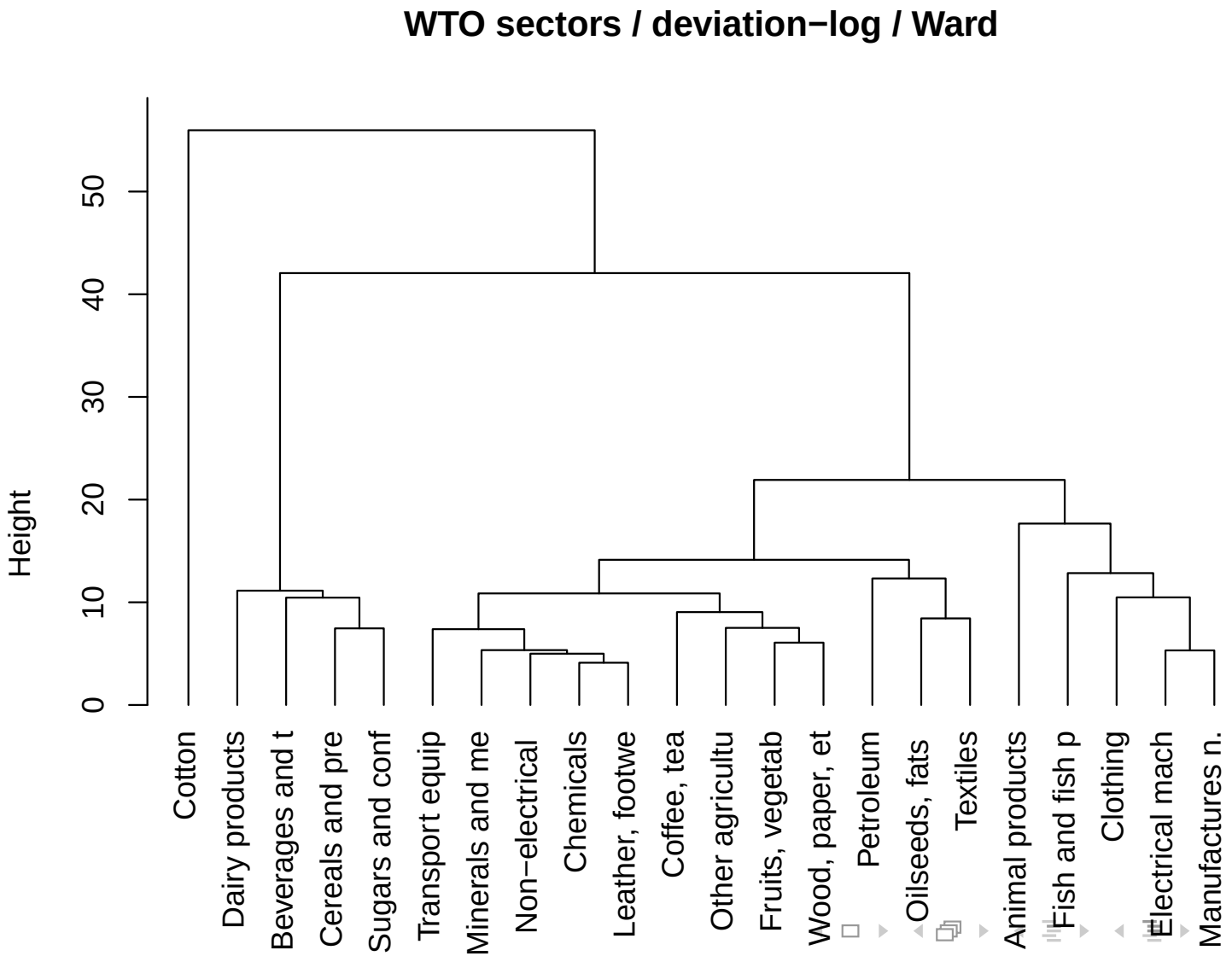
On both networks we

- applied the proposed deviations approach,
- computed the corresponding row and column Euclidean distance matrices (a kind of projection),
- clustered both rows and columns (using hierarchical Ward method),
- displayed the deviations matrix **B** considering orderings obtained by clustering.

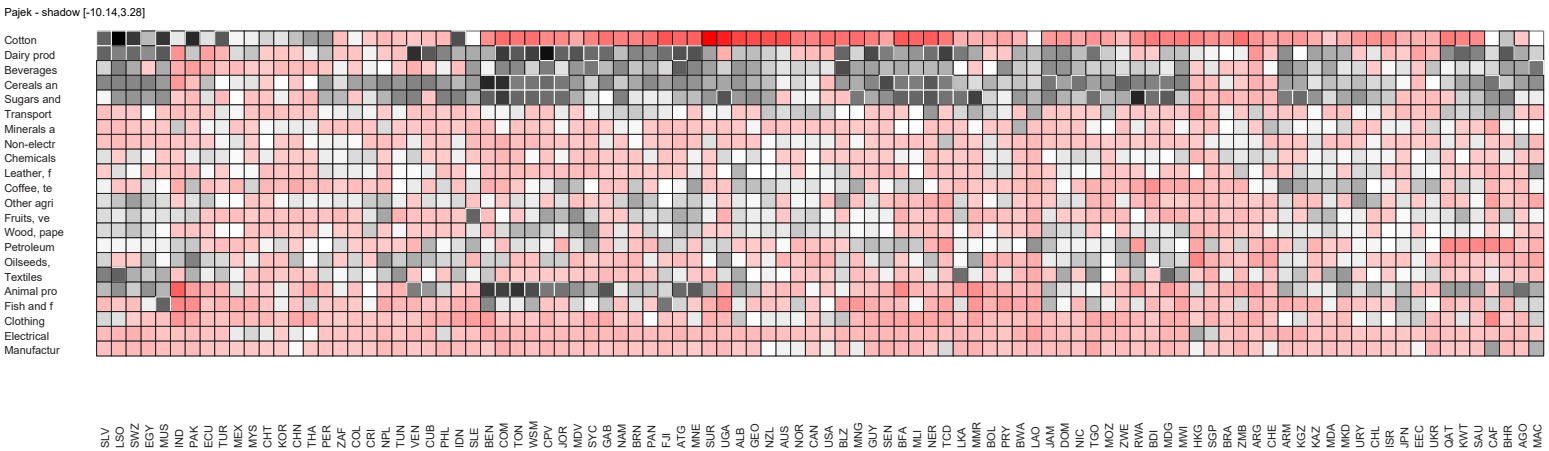
WTO annual bilateral importst 2015 / Countries



WTO annual bilateral imports 2015 / Sectors



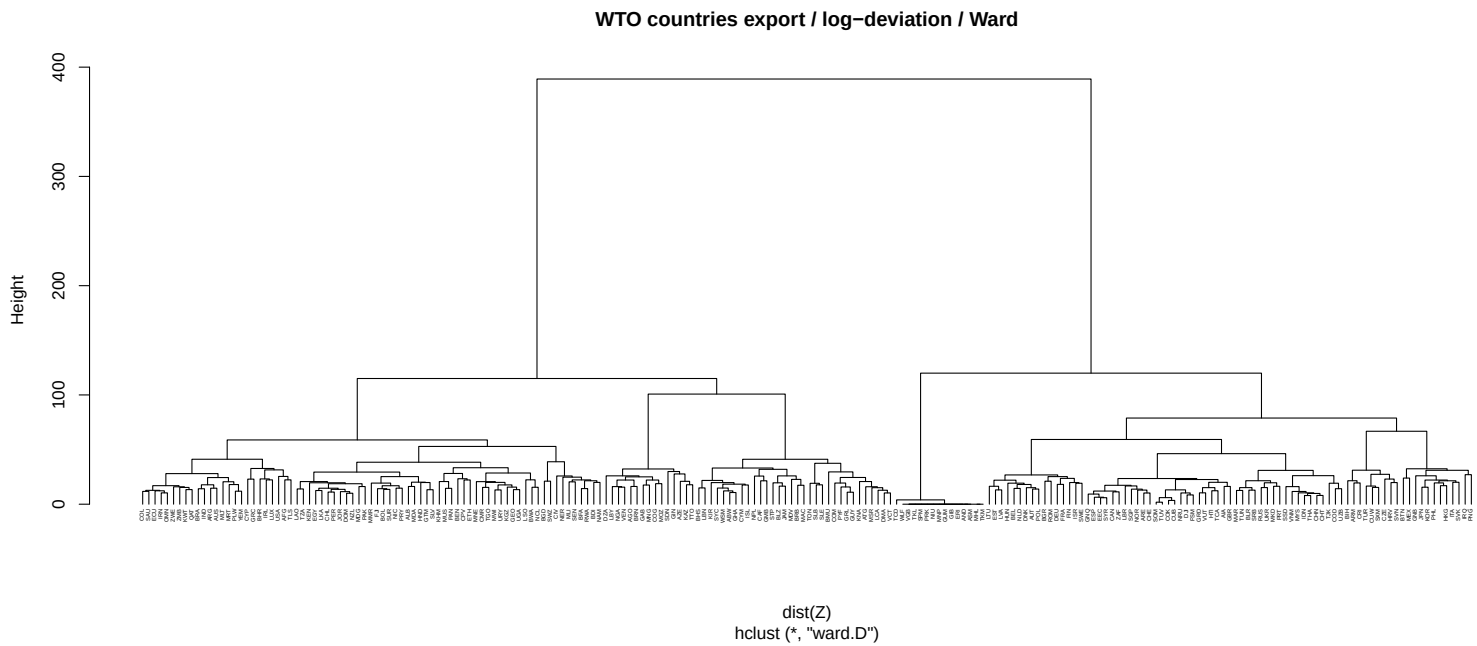
WTO annual bilateral imports 2015 / Matrix



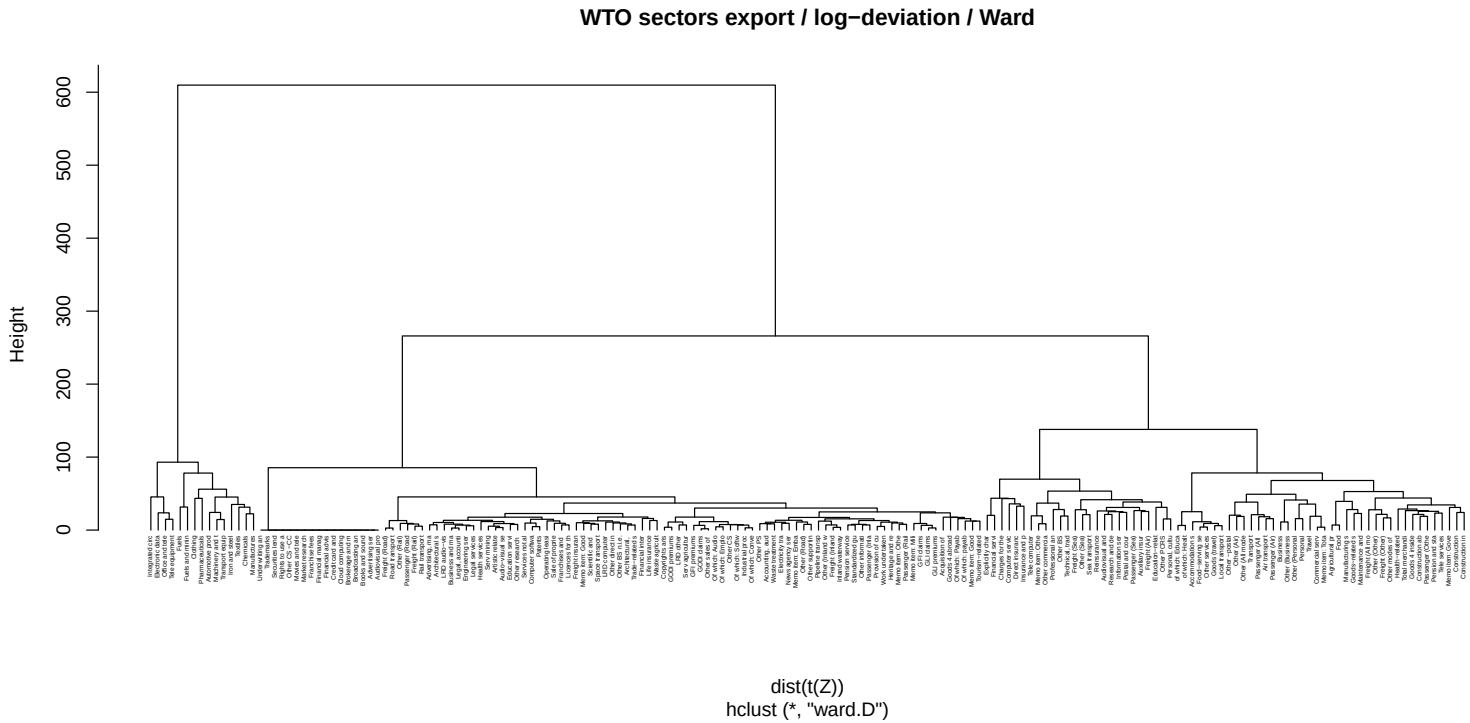

```
Cotton:

     CAF       LAO       MAC       SLE       SUR       UGA     ...
  0.0000    0.0000    0.0000    0.0000   62.4901 770.1000     ...

       IND        KOR        PAK        THA         IDN         TUR         CHN
 394102448  485391882  546190127  549158614  1088290181  1241859599  2653891735
```

white cells – missing data (or no deviation)

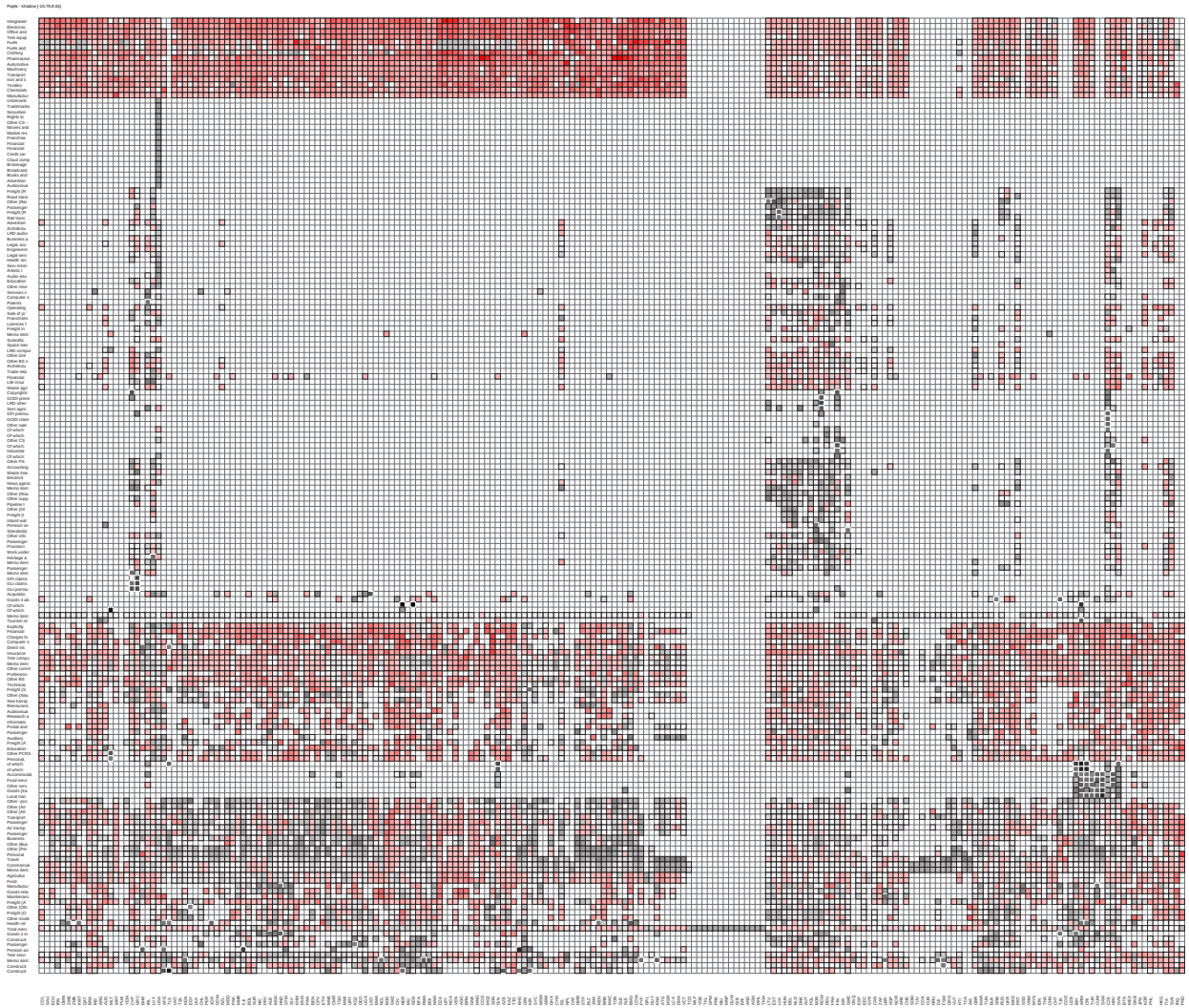
WTO annual marchandise export 2015 / Countries



WTO annual marchandise export 2015 / Sectors



WTO annual marchandise export 2015 / Matrix



Conclusions

- the proposed deviations approach can be used also for one-mode weighted networks,
- the rows/columns of the deviations matrix can be reordered respecting the clustering hierarchy,
- we can include in the display also clusters (blockmodeling),
- for analytical tasks some numerical/statistical evaluation of the obtained structures is needed,
- a collection of nice, well documented weighted two-mode networks is needed.

https://github.com/bavla/NormNet/tree/main/data/WTO

Acknowledgments

This work is supported in part by the Slovenian Research Agency (research program P1-0294 and research projects J1-9187 and J5-2557), and prepared within the framework of the HSE University Basic Research Program.

Additional references

Ahmed, A. ; Batagelj, V.; Fu, X.; Hong, S.-H.; Merrick, D.; Mrvar, A.: Visualisation and analysis of the Internet movie database. Asia-Pacific Symposium on Visualisation 2007 (IEEE Cat. No. 07EX1615), 2007, p 17-24. PDF

Batagelj, V, Doreian, P, Ferligoj, A, Kejžar, N: Understanding Large Temporal Networks and Spatial Networks: Exploration, Pattern Searching, Visualization and Network Evolution. Wiley Series in Computational and Quantitative Social Science. Wiley, 2014.

Cerinšek, M., Batagelj, V.: Generalized two-mode cores. Social Networks 42 (2015), 80–87. PDF