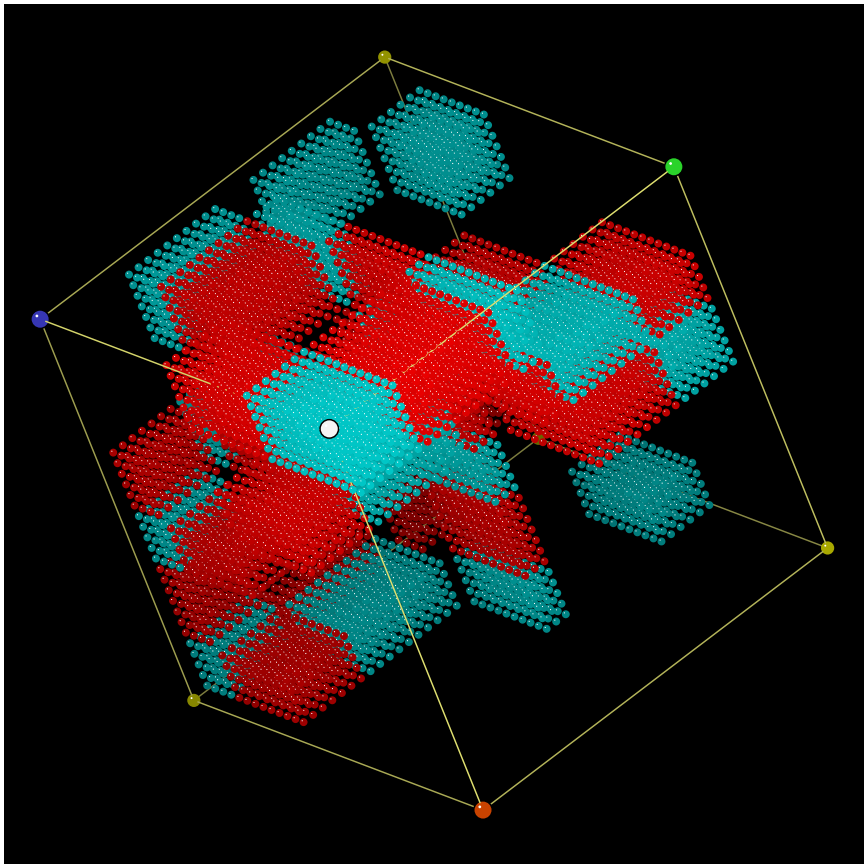
Analysis of weighted multiway networks

Vladimir Batagelj
IMFM Ljubljana

1327. sredin seminar
Ljubljana, 9. november 2022

Outline



Vladimir Batagelj: vladimir.batagelj@fmf.uni-lj.si
Current version of slides (November 8, 2022 at 03:52): slides PDF
https://github.com/bavla/ibm3m/

IFCS 2022

This seminar is a result of joint work with colleagues from Napoli and Salerno. At the IFCS 2022 conference in July in Porto, they were presenting their research on Italian students' mobility. They developed their own approach and asked me for comments and alternatives.

Italian students' data

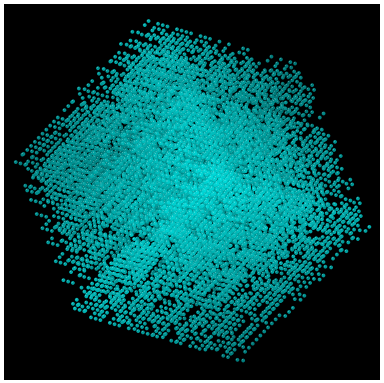
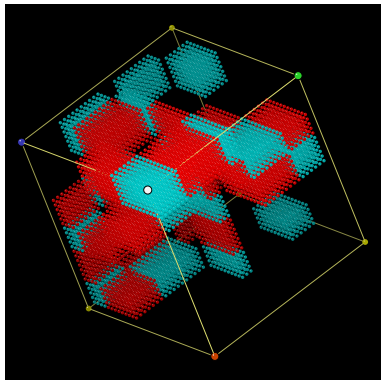
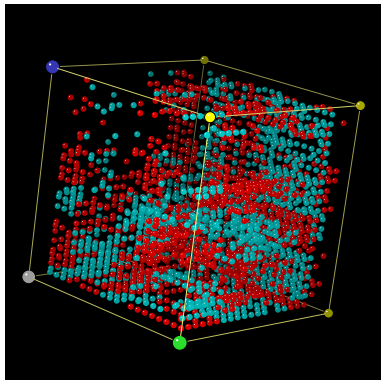
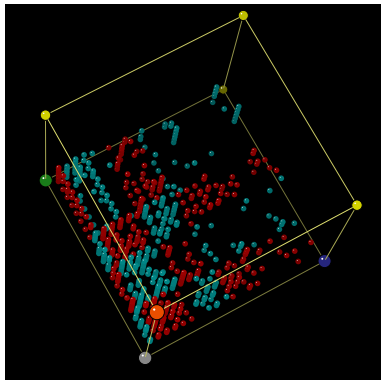
Their data consists of quadruples (province, university, study program, year) with the corresponding value (weight) counting how many students from a given province selected in a given year to study a given program at a given university. Such data are called *weighted multiway networks*.

The approach proposed by my Italian colleagues was to analyze the data for each year separately by joining the *ways* university and study program into a single way [university, study program] thus obtaining a weighted two/mode network that can be analyzed using available methods.

I first transformed their data into Pajek format and applied some additional methods on them. Genova

Afterward, I transformed the data back into a multiway network. I saved the data in JSON format.

Blockmodeling

I already worked (with Anuška and Pat, in 2006) on the indirect approach to blockmodeling of binary 3-way networks (Github bavla/ibm3m) [3]. Lazega, Krackhardt [4]

We were dealing with 3-way networks also at the INSNA 2009 Viszards session analyzing the Bibsonomy data.

Drilling into the data of my Italian friends I got some ideas for possible approaches.

Multiway networks

A *weighted multiway network* $N = (V, L, w)$ is based on *nodes* from k pairwise disjoint sets (ways or dimensions)

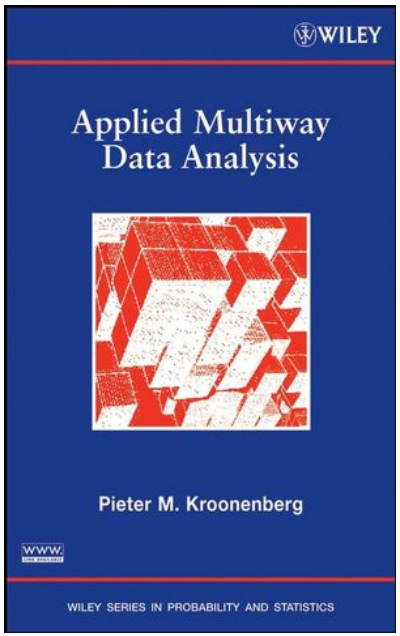
$$V = \{V_1, V_2, \ldots, V_k\}, \quad V_i \cap V_j = \emptyset \text{ for } i \neq j$$

The set of *links* $L \subseteq V_1 \times V_2 \times \cdots \times V_k$. The weight $w : L \to \mathbb{R}$.
It can be represented by a k-dimensional array W

$$W[v_1, v_2, \ldots, v_k] = w(v_1, v_2, \ldots, v_k) \text{ for } (v_1, v_2, \ldots, v_k) \in L$$

otherwise $W[v_1, v_2, \ldots, v_k] = 0$.
In a general multiway network, different additional data can be known for nodes and/or links.

Multiway analysis



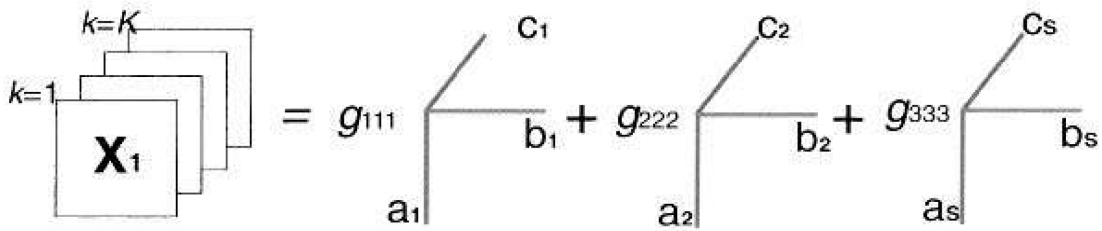


Figure 4.9 The Parafac model.

In data analysis already exists an approach called *multiway analysis* which deals with generalizations of ideas from the multivariate analysis to multiway arrays. They are known as Tucker, Parafac, SimPCA, . . . models [6, 7].

We will continue in another direction – a reduction of multiway networks analysis to weighted networks analysis.

Transformations

In data analysis of multi-way networks, some transformations could prove to be useful:

- reordering of ways
- joining the ways
- flattening of a way
- projection to a selected way
- normalization
- recoding (binarization)

Representation

A natural representation of a weighted multi-way network is by a multi-dimensional array. In real-life networks many (most of) array entries have the value 0. In such cases, this representation is computationally inefficient – takes more space and requires unnecessary computations with 0s.

Because of this, for an experimental implementation of a library for the analysis of weighted multiway networks, I selected the k-tuples representation as a basic data representation. It allows efficient work also with large sparse multiway networks.

Representation

An alternative representation follows the formal definition

```
{
title = "TITLE";

nodes1 = ["v11", ...,"v1n1"];
...
nodesk = ["vk1", ...,"vknk"];
nodes = [ nodes1, ..., nodesk ];
links = [
...
[link: [v1,v2,...,vk], w: W ];
...
]
}
```

lazegaMN.json

Reordering of ways

Trivial.

Wiki

Joining the ways

Selected ways V_i and V_j, $i < j$, are replaced by a new joint way

$$V_{ij} = \{(u : v) : u \in V_i \wedge v \in V_j \wedge \exists(\ldots, u, \ldots, v, \ldots) \in L\}$$

*** add a detailed description of the transformed network.
This transformation reduces the number of ways for 1.
"Commutativity".

Wiki

Flattening of a way

A selected way V_i is removed from the network.

$$V' = \{V_1, V_2, \ldots, V_{i-1}, V_{i+1}, \ldots, V_k\}$$

$$w'(v_1, v_2, \ldots, v_{i-1}, v_{i+1}, \ldots, v_k) = \sum_{v \in V_i} w(v_1, v_2, \ldots, v_{i-1}, v, v_{i+1}, \ldots, v_k)$$

This transformation reduces the number of ways for 1.

"Commutativity".

wiki

Normalization

log w

TBA [2]

Recoding

Let $P = \{P_1, P_2, \ldots, P_p\}$ be a partition of $\mathbb{R}$. The recoding transformation transforms the weight function w into a new weight w' determined for a link $(v_1, v_2, \ldots, v_k) \in L$ as

$$w'(v_1, v_2, \ldots, v_k) = i \Leftrightarrow w(v_1, v_2, \ldots, v_k) \in P_i$$

Code 0 corresponds to the case $(v_1, v_2, \ldots, v_k) \notin L$ which is usually equivalent to $w(v_1, v_2, \ldots, v_k) = 0$ in the array representation. If 0 is also a legal weight value we have to introduce another zero, $\square$, that indicates the absence of a link.

A special case is a binarization for which $P_0 = \{0\}$ and $P_1 = \mathbb{R} \setminus P_0$.

Recoding is often used to get more readable matrix visualizations of a given network.

wiki

Projection to a selected way

Because of the reordering option, we can assume that we selected the way 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 (V_1, A, p), $A \subseteq V_1 \times V_1$ and $p : A \to \mathbb{R}$. Let $u, t \in V_1$ then

$$p(u,t) = \sum_{(v_2,\ldots,v_k) \in V_2 \times \cdots \times V_k} w(u, v_2, \ldots, v_k) \cdot w(t, v_2, \ldots, v_k)$$

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.

wiki

Salton index has the following properties

From the projection p we can get the corresponding measure of similarity – *Salton index* $S(u,t)$ [1]

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

that can be used for clustering the set V_1. wiki

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. $S(\alpha u, u) = 1, \quad \alpha > 0$

Acknowledgments

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 HSE University Basic Research Program.

References I


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, 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.


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.


Kroonenberg, PM: Applied multi-way data analysis. Wiley, 2008.

References II

Schoonees, PC, Groenen, PJF, van de Velden, M: Least-squares bilinear clustering of three-way data. Advances in Data Analysis and Classification (2022) 16:1001–1037. paper.