

# Hb-graphs and their applications to the ranking of information in an information space

AIME 2019  
Budapest  
26.11.2019

Xavier Ouvrard<sup>1,2</sup>

supervised by Pr Stéphane Marchand-Maillet<sup>1</sup> and Dr Jean-Marie Le Goff<sup>2</sup>

[www.infos-informatique.net](http://www.infos-informatique.net)



# The example of search on a Scientific Publication Database

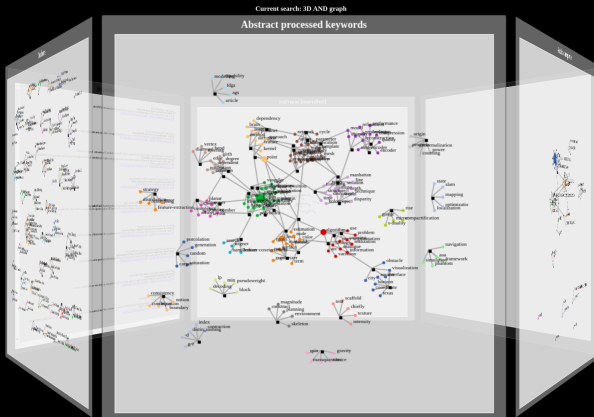
Current search: hypergraph

## References

- Chen Cheng, Liang Qi, Xiang Tian  
The first few acyclic and tripartite hypergraphs with larger spectral radii  
A connected  $k$ -set-regular hypergraph with  $k$  odd vertices and  $k+2$  edges is called  $k$ -cyclic if its rank  $(r) = 1$ .  
<https://doi.org/10.1016/j.amc.2018.03.014> <https://www.sciencedirect.com/science/article/pii/S0191261518300141>
- Yuan Ding, Shao-Fan, Adam Gray, Friedhelm Strüger, Karim M. Maghs, Akram Parfy, Bing Yang, Guangshu Yu  
Inclusion in Hypergraphs  
In this paper a new perspective for hypergraphs called hypergraph inclusion is defined.  
<https://doi.org/10.1007/s11464-018-0691-4> <http://www.springerlink.com/10.1007/s11464-018-0691-4>
- Juho Hojman  
Zero forcing and maximum outflow for hypergraphs  
The concept of zero forcing is extended from graphs to uniform hypergraphs to analogize with the way zero forcing was defined on an upper bound for the maximum outflow of a family of combinatorial structures whose combinatorial problems are identified by a given graph. A family of systematic hypergraphs is associated with a uniform hypergraph and zero forcing is defined on this family.  
<https://doi.org/10.1007/s11464-018-0691-4> <http://www.springerlink.com/10.1007/s11464-018-0691-4>
- Thomas Schuster  
Combinatorial Hypergraph Coloring  
A uniform hypergraph is called  $k$ -colorable if it is a class of hypergraphs that is hereditary and non-trivial, i.e., closed under induced subhypergraphs and it contains a non-empty hypergraph for all  $k$  hypergraphs.  
<https://doi.org/10.1007/s11464-018-0691-4> <http://www.springerlink.com/10.1007/s11464-018-0691-4>
- James J. Bruck  
Ordered Hypergraphs I: Introduction and Balance  
An ordered hypergraph is an ordered multiset of subsets of a finite set.  
<https://doi.org/10.1007/s11464-018-0691-4> <http://www.springerlink.com/10.1007/s11464-018-0691-4>
- Nguyen Dinh  
A Boolean Functions Theoretic Approach to Quantum Hypergraph States and Entanglement  
We establish an one-to-one correspondence between the Boolean functions and hypergraph states, that are entangled multipartite pure quantum states corresponding to hypergraphs.  
<https://doi.org/10.1007/s11464-018-0691-4> <http://www.springerlink.com/10.1007/s11464-018-0691-4>
- Janine Marie-Franz, Muelar Maria, Karl-Ludwig  
Uniform Hypergraphs and dominating sets of graphs  
A uniform hypergraph is a family of pairwise disjoint subsets of a finite set.  
<https://doi.org/10.1007/s11464-018-0691-4> <http://www.springerlink.com/10.1007/s11464-018-0691-4>
- Wei Ewens  
A class of hypergraphs that generalizes chordal graphs  
In this paper we introduce a class of hypergraphs that we call chordal.  
<https://doi.org/10.1007/s11464-018-0691-4> <http://www.springerlink.com/10.1007/s11464-018-0691-4>
- Galina Elsh, Balazs Szegedy  
Limits of Hypergraphs, Random and Regularity Lemmas. A Non-standard Approach  
We study the hypergraph and regularity theory of the submodular of large sets.  
<https://doi.org/10.1007/s11464-018-0691-4> <http://www.springerlink.com/10.1007/s11464-018-0691-4>
- Robert Schmalzer, Eberhard Tschach  
Fundamentals of balanced hypergraphs  
We give a new proof of Kővári's theorem and generalize the Gallai-Edmonds decomposition to balanced hypergraphs in two different ways.  
<https://doi.org/10.1007/s11464-018-0691-4> <http://www.springerlink.com/10.1007/s11464-018-0691-4>
- Jin-Feng, Yuxuan Chen  
Decomposition of  $k$ -uniform hypergraphs  $K_{n, n}^{(k)}$  into hypergraphs  $K_{n, n}^{(k)}$   
In this paper it is established that a decomposition of a  $k$ -uniform hypergraph  $K_{n, n}^{(k)}$  into a special kind of hypergraph  $K_{n, n}^{(k)}$  exists if and only if  $n$  is a multiple of  $k$ .  
<https://doi.org/10.1007/s11464-018-0691-4> <http://www.springerlink.com/10.1007/s11464-018-0691-4>
- David Santos, Andrew Thomason  
Single contains for simple hypergraphs  
We give an easy method for constructing counterexamples for simple hypergraphs.  
<https://doi.org/10.1007/s11464-018-0691-4> <http://www.springerlink.com/10.1007/s11464-018-0691-4>

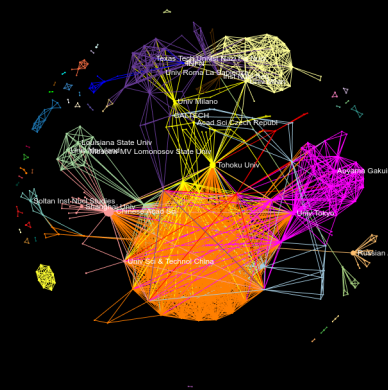
# Searching on a Scientific Publication Database

Abstracts

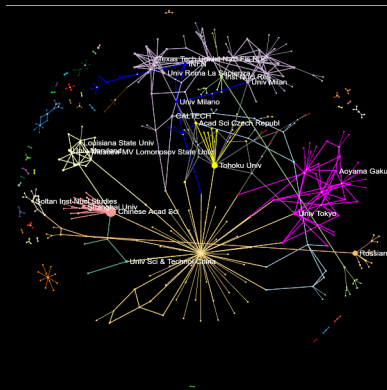


Designed and developed  
by Koenig Overland

# Visualizing facets I



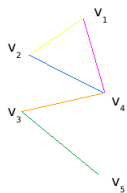
Clique view



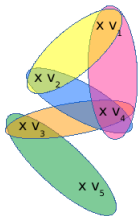
Extra-node view

Figures from Ouvrard et al. [2017]

# From graphs to hypergraphs



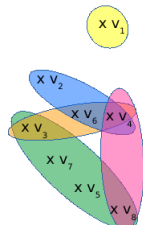
**A graph:**  
An edge links 2 vertices



**A graph with a set view:**  
an edge is now a set of 2 vertices  
=> 2-uniform hypergraph



**A k-uniform hypergraph**  
an edge is now a set of k vertices  
=> k-uniform hypergraph



**A general hypergraph:**  
(hyper)edges are set of 1 and more vertices

# Motivation for a new structure I

- **Background question:**

## How to coarsen a hyper(bag)-graph?

- Task to be solved:
  - spot out the **important structures** of a hypergraph
- Important for
  - **spraying** the information shown
  - give **focus on important information, at vertex and edge level**
- Approach taken: **diffusion approach**

# Motivation for a new structure II

At **which level** the **diffusion process** occurs?

- At **graph level** => Laplacian matrix ... linked to adjacency matrix
- At **hyper(bag)-graph level** => **incidence matrix**

Pitfalls:  $HH^T$  take us back to pairwise relationships => n-arity not totally taken into account

- An **adjacency tensor** is needed:
  - Well defined for uniform hypergraphs
  - For general hypergraphs:
    - **Adjacency** has to be refined
    - Convenient adjacency tensor to ensure diffusion and Laplacian tensor

# On adjacency in general hypergraphs

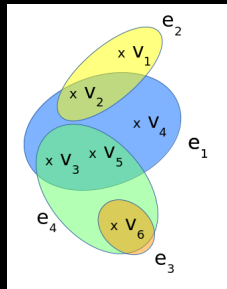
## Refining adjacency in hypergraphs

- **Adjacency is more than pairwise notion**
- Two distinct concepts:
  - **e-adjacency**: vertices of a given hyperedge are said e-adjacent
  - **k-adjacency**:  $k$  given vertices are  $k$ -adjacent if it exists a hyperedge that hold them
  - **$\bar{k}$ -adjacency**: the maximal  $k$ -adjacency a hypergraph holds
- In  $k$ -uniform hypergraph:
  - $\bar{k}$ -adjacency corresponds to  $k$ -adjacency
  - The e-adjacency corresponds to  $\bar{k}$ -adjacency
- In general hypergraphs:

$\bar{k}$ -adjacency  $\neq$  e-adjacency

Hypergraph  $\mathcal{H} = (V, E)$

- $V = \{v_1, v_2, v_3, v_4, v_5, v_6\}$
- $E = \{e_1, e_2, e_3, e_4\}$



- $v_2$  and  $v_3$  are 2-adjacent
- $v_2, v_3$  and  $v_4$  are 3-adjacent
- $v_1$  and  $v_2$  are e-adjacent
- $\bar{k}$ -adjacency corresponds to 4-adjacency



Tensor for general hypergraphs: the art of filling or how to obtain cubic form from a non cubic object



First way: cut everything in small pieces...



First way: ... and gather



tabifinch

# First way: ... and gather

- Uniformisation process based on **hyperedge splitting**

=> push everything in a tensor of order  $|r_{\mathcal{H}}|$  and size  $|V|$ .

=> **Banerjee & al**<sup>2</sup>: split the storage in the elements of the tensor that is not occupied

The ([Author's note]: **e- adjacency hypermatrix** of  $\mathcal{H}$  written

$\mathcal{A}_{\mathcal{H}} = (a_{i_1 \dots i_{k_{\max}}})_{1 \leq i_1, \dots, i_{k_{\max}} \leq n}$  is such that for a hyperedge:  $e = \{v_{l_1}, \dots, v_{l_s}\}$  of cardinality  $s \leq k_{\max}$ .

$$a_{p_1 \dots p_{k_{\max}}} = \frac{s}{\alpha}, \text{ where } \alpha = \sum_{\substack{k_1, \dots, k_s \geq 1 \\ \sum k_i = k_{\max}}} \frac{k_{\max}!}{k_1! \dots k_s!}$$

with  $p_1, \dots, p_{k_{\max}}$  chosen in all possible way from  $\{l_1, \dots, l_s\}$  with at least once from each element of  $\{l_1, \dots, l_s\}$ .

=> **Sun & al**<sup>3</sup>: similarly the same, but not symmetric.

---

<sup>2</sup> **Banerjee et al. [2017]**

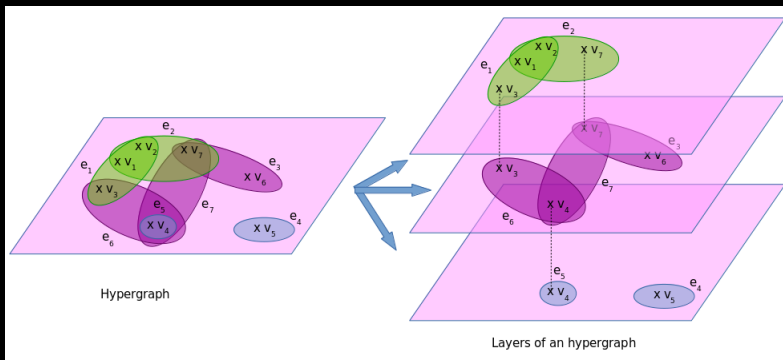
<sup>3</sup> **Sun et al. [2018]**

## Second way: the art of filling



# e-adjacency tensor of a general hypergraph I

- split the hypergraph in layers of uniformity:



# e-adjacency tensor of a general hypergraph II

- hypergraph uniformisation process based on **hyperedge filling**:

## **Various filling options:**

- Always add the **same vertex**

=> **straightforward approach**

- Add a **different vertex** per uniform hypergraph layer up to fulfillment

=> **silo approach**

- Add a vertex for each layer (previous approach)

=> **iterative approach**

In the first two approaches, **multisets** are required to keep interpretability

**First two approaches require multisets**

# A parenthesis on multisets

## Multisets:

**Multiset:** a universe and a multiplicity function  $A_m = (A, m)$

**Natural multiset:** the range of the multiplicity function is a subset of  $\mathbb{N}$ .

In natural multisets: two views:

weighted set:  $A_m = \{x_1^{m_1}, \dots, x_n^{m_n}\}$

collection of objects  $\left\{ \left\{ \underbrace{x_1, \dots, x_1}_{m_1 \text{ times}}, \dots, \underbrace{x_n, \dots, x_n}_{m_n \text{ times}} \right\} \right\}$



# Hb-graphs I

**Hb-graph  $\mathcal{H} = (V, E)$ :** family of multisets  $E = (e_i)_{i \in I}$ , with  $I = \llbracket p \rrbracket$  - called **hb-edges** - where the hb-edges have:

- same universe  $V = \{v_1, \dots, v_n\}$ , called **vertex set**.
- support a subset of  $V$ .
- each hb-edge has its own multiplicity function  $m_e : V \rightarrow \mathbb{W}$  where  $\mathbb{W} \subset \mathbb{R}^+$ .

**Incidence matrix of hb-graphs:**

$$H = [m_j(v_i)]_{\substack{1 \leq i \leq n \\ 1 \leq j \leq p}}$$

# Hb-graphs II

## Different application of hb-graphs:

- General hypergraph e-adjacency tensor
- **Network of co-occurrences** and the hb-graph framework
- **Prime decomposition** and hb-graphs

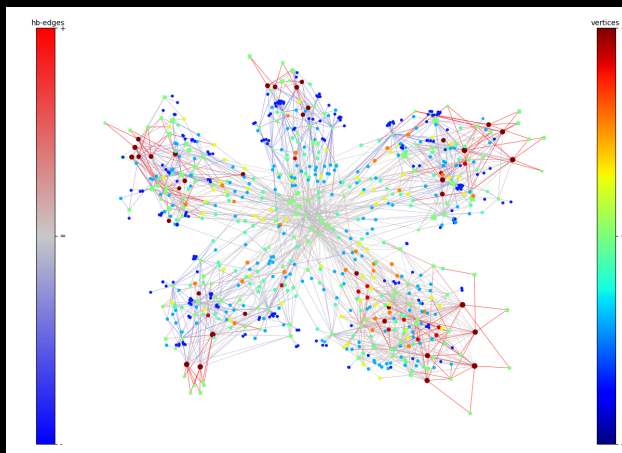
### • **Text and hb-graphs:**

=> bag of words to represent text are efficient modeling for information retrieval

### • **Image and hb-graphs:**

=> bag of visual words are often used in image.

# ML interest of hb-graphs



## Exchange-based diffusion in hb-graphs:

- **Stochastic process**
- Allows **generalised random walk**
- Defines **a ranking** of vertices and hb-edges (akin to PageRank)
- Enables **coarsening** of hb-graphs and thus data landscape

# Exchange-based diffusion algorithm

## Given:

A hb-graph  $\mathcal{H} = (V, E, w_e)$  with  $|V| = n$  and  $|E| = p$

Number of iterations:  $T$

## Initialisation:

For all  $v_i \in V$  :  $\alpha_i := \frac{1}{n}$

For all  $e_j \in E$  :  $\epsilon_j := 0$

## DiffuseFromVerticesToHbEdges():

For  $j := 1$  to  $p$ :

$\epsilon_j := 0$

For  $v_i \in e_j^*$ :

$$\epsilon_j := \epsilon_j + \frac{m_j(v_i) w_e(e_j)}{d_{w,m}(v_i)} \alpha_i$$

## DiffuseFromHbEdgesToVertices():

For  $i := 1$  to  $n$ :

$\alpha_i := 0$

For  $e_j$  such that  $v_i \in e_j^*$ :

$$\alpha_i := \alpha_i + \frac{m_j(v_i)}{\#_m e_j} \epsilon_j$$

## Main():

Calculate for all  $i$  :  $d_{w,m}(v_i)$  and for all  $j$  :  $\#_m e_j$

For  $t = 1$  to  $T$ :

DiffuseFromVerticesToHbEdges()

DiffuseFromHbEdgesToVertices()

# Exchange-based diffusion algorithm

## Given:

A hb-graph  $\mathcal{H} = (V, E, w_e)$  with  $|V| = n$  and  $|E| = p$

Number of iterations:  $T$

## Initialisation:

For all  $v_i \in V$  :  $\alpha_i := \frac{1}{n}$

For all  $e_j \in E$  :  $\epsilon_j := 0$

## DiffuseFromVerticesToHbEdges():

For  $j := 1$  to  $p$ :

$\epsilon_j := 0$

For  $v_i \in e_j^*$ :

$$\epsilon_j := \epsilon_j + \frac{m_j(v_i) w_e(e_j)}{d_{w,m}(v_i)} \alpha_i$$

## DiffuseFromHbEdgesToVertices():

For  $i := 1$  to  $n$ :

$\alpha_i := 0$

For  $e_j$  such that  $v_i \in e_j^*$ :

$$\alpha_i := \alpha_i + \frac{m_j(v_i)}{\#_m e_j} \epsilon_j$$

## Main():

Calculate for all  $i$  :  $d_{w,m}(v_i)$  and for all  $j$  :  $\#_m e_j$

For  $t = 1$  to  $T$ :

DiffuseFromVerticesToHbEdges()

DiffuseFromHbEdgesToVertices()

## Time complexity:

$$O(T (d_{\mathcal{H}} n + r_{\mathcal{H}} p))$$

where:

$$d_{\mathcal{H}} = \max_{v_i \in V} (d_i)$$

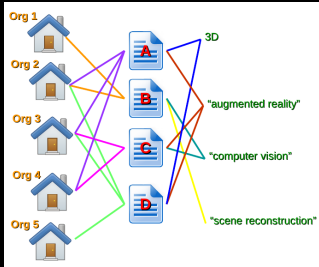
$$r_{\mathcal{H}} = \max_{e_j \in E} |e_j^*|$$

# Hb-graph m-uniformisation process and diffusion

Using a m-uniformisation process modifies the exchange-based diffusion:

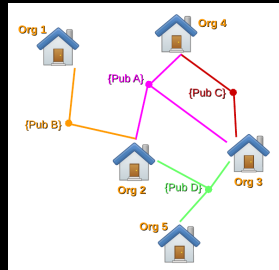
- **Explainable in the case of the hb-edge filling**
- Unclear how it is modified in the case of the hb-edge splitting

# In a Scientific Publication Database

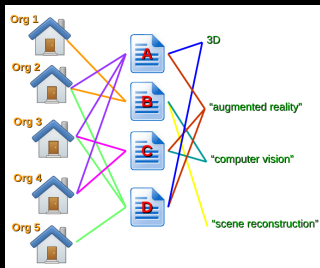


Reference: **Publication**, Facet: **Organization**

Pub A	Org 2, Org3, Org 4
Pub B	Org 1, Org 2
Pub C	Org 3, Org 4
Pub D	Org2, Org 3, Org 5

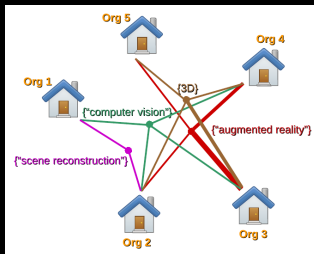


# In a Scientific Publication Database



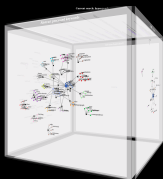
Reference: **Keywords**, Facet: **Organization**

scene reconstruction	{ { Org 1 <sup>1</sup> , Org 2 <sup>1</sup> } }
computer vision	{ { Org 1 <sup>1</sup> , Org 2 <sup>1</sup> , Org 3 <sup>1</sup> , Org 4 <sup>1</sup> } }
augmented reality	{ { Org 2 <sup>2</sup> , Org 3 <sup>3</sup> , Org 4 <sup>2</sup> , Org 5 <sup>1</sup> } }
3D	{ { Org 2 <sup>2</sup> , Org 3 <sup>2</sup> , Org 4 <sup>1</sup> , Org 5 <sup>1</sup> } }





# Facets of the information space

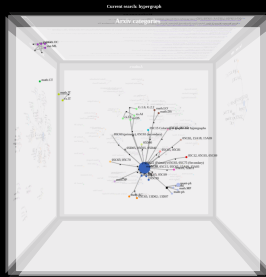
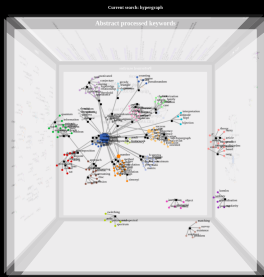
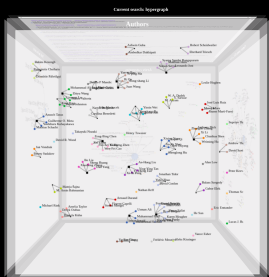


Facet choice

Authors

Processed keywords

Arxiv Categories



# Aggregating the information

## Aggregating the information obtained on the different facets:

- **We have proposed a modified MC4, called WT-MC4**

=> allows ranking of references using weights on facets

- We also proposed a biased ranking of the information on facets, to put emphasize on some kind of features

# A proposal

On each facet  $F_1, \dots, F_m \Rightarrow$  a hb-graph  $\mathcal{H}_k = (V_k, E_k)$  with  $1 \leq k \leq m$

1. **Exchange-base diffusion on each facet hb-graph**  $\Rightarrow$  hb-edges  $\epsilon_k(e_j)$
2. **Each hb-edge is linked to different physical references**: 1 to many.  $\Rightarrow$  ranking  $R_k$  of references  $r_i$  with possible ties = rank of the corresponding hb-edge.
3. Facets have **associated weights** (tunable)  $w_1, \dots, w_m$  such that  $\sum_{i \in [m]} w_i = 1, w_i \geq 0$ .
4. We start by computing a weighted majority matrix for each couple of references:

$$M(r_{i_1}, r_{i_2}) = \sum_{k \in [K]} w_k \mathbf{1}_{R_k(r_{i_1}) < R_k(r_{i_2})} - \sum_{k \in [K]} w_k \mathbf{1}_{R_k(r_{i_1}) > R_k(r_{i_2})}$$

5. We use a **modified MC4 of Dwork et al. [2001] with teleportation and weights**:

Current state reference:  $r_{\text{current}}$ .

- Choose a random number  $\gamma$
- Choose an other reference  $r_{\text{next}}$  uniformly among all the references ranked.

If  $\gamma > \gamma_0$ :

go to  $r_{\text{next}}$

else:

If  $M(r_{\text{next}}, r_{\text{current}}) > 0$

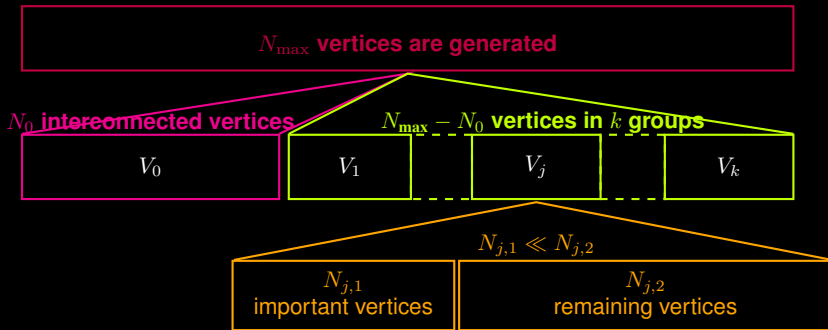
go to  $r_{\text{next}}$

else:

stay in  $r_{\text{current}}$ .

# Methodology of evaluation

- 2 parts in experimentation:
  - **generation of random hb-graphs** => 1 per facet



- **a generated reference hb-graph is built out of the facets**
- **Perform diffusion on each facet separately (multi-diffusion)**
- Aggregation using the modified MC4 and comparison to Borda results

# Results for our modified MC4 I

On the generated information space, we observe for **100 generated information spaces**, the following results for Kendall tau between Borda ranking and rankings obtained by:

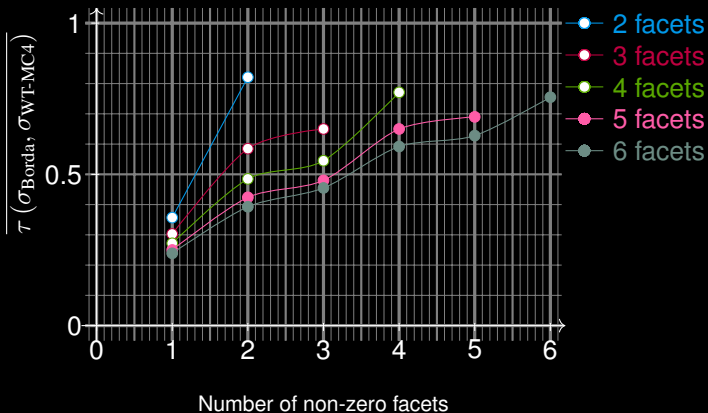
Current	$\tau(\sigma_{\text{Borda}}, \sigma_{\text{current}})$	$\sigma(\tau(\sigma_{\text{Borda}}, \sigma_{\text{current}}))$
Facet 0	0.262	0.09
Facet 1	0.261	0.08
Facet 2	0.237	0.104
References	0.317	0.283
<b>WT-MC4{'0': 0.33, '1': 0.33, '2': 0.33}</b>	<b>0.649</b>	<b>0.116</b>
WT-MC4{'0': 0.5, '1': 0.5, '2': 0.0}	0.581	0.114
WT-MC4{'0': 0.5, '1': 0.0, '2': 0.5}	0.549	0.115
WT-MC4{'0': 0.0, '1': 0.5, '2': 0.5}	0.563	0.123
WT-MC4{'0': 0.0, '1': 0.0, '2': 1.0}	0.261	0.107
WT-MC4{'0': 0.0, '1': 1.0, '2': 0.0}	0.279	0.084
WT-MC4{'0': 1.0, '1': 0.0, '2': 0.0}	0.286	0.096

# Results for our modified MC4 II

Average Kendall tau of the WT-MC4 ranking aggregation compared to the non-zero equal weight facet rankings (average on 1000 information spaces) depending on the number of facets having non-zero weights.

# non-zero weight facets $\Rightarrow$	1		2		3		4		5		6	
# facets $\downarrow$	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev
2	0.357	0.096	0.821	0.050								
3	0.303	0.094	0.585	0.127	0.650	0.112						
4	0.272	0.093	0.485	0.116	0.545	0.116	0.771	0.066				
5	0.250	0.091	0.424	0.108	0.480	0.109	0.650	0.094	0.690	0.085		
6	0.239	0.086	0.393	0.096	0.455	0.101	0.592	0.089	0.628	0.090	0.755	0.058

# Results for our modified MC4 III



# Results for our modified MC4 IV

- Putting weight on only one facet: 'authors'

**Comparison of rankings obtained by Dwork\_MC4** with weights:

{'authors' : 1.0, 'keywords' : 0, 'tags' : 0}

and rankings on the facet:

	<b>Kendall's Tau coefficient</b>	<b>Scaled Spearman Footrule coeff</b>	<b>Jaccard Index @10 / @25</b>
<b>authors</b>	<b>0.819</b>	0.125	<b>0.818 / 0.923</b>
keywords	0.076	0.598	0.125 / 0.231
tags	-0.112	0.710	0.171 / 0.315

- **Same kind of results for putting all the weight on keywords or tags**



# Results for our modified MC4 V

- Putting equal weights on each facet:

**Comparison of rankings obtained by Dwork\_MC4** with weights:

$$\left\{ \text{'authors'} : \frac{1}{3}, \text{'keywords'} : \frac{1}{3}, \text{'tags'} : \frac{1}{3} \right\}$$

and other rankings:

	<b>Kendall's Tau coefficient</b>	<b>Scaled Spearman Footrule coeff</b>	<b>Jaccard Index @10 / @25</b>
<b>authors</b>	0.315	0.444	0.333 / 0.515
keywords	0.508	0.327	0.385 / 0.455
tags	0.228	0.612	0.242 / 0.449
Borda	<b>0.818</b>	<b>0.127</b>	<b>0.636 / 0.846</b>
Refs	-0.438	0.995	0.044 / 0.192

# Results for our modified MC4 VI

- **Comparison of rankings obtained by Dwork\_MC4** with weights:

{'authors' : 0.5, 'keywords' : 0, 'tags' : 0.5}

and other rankings:

	<b>Kendall's Tau coefficient</b>	<b>Scaled Spearman Footrule coeff</b>	<b>Jaccard Index @10 / @25</b>
<b>authors</b>	0.354	0.434	0.333 / 0.471
keywords	0.441	0.369	0.385 / 0.455
tags	0.264	0.620	0.242 / 0.449
Borda	<b>0.811</b>	<b>0.134</b>	<b>0.636 / 0.846</b>
Refs	-0.486	0.984	0.043 / 0.212

# Perform visual queries

A **demo** is worth a thousand words...

# Thank you for your attention!



*Leveraging insight into your data network by viewing co-occurrences while navigating across different perspectives.*

## **The HbGraphDataEdron:**

- Is part of the Collaboration Spotting family
- Collspotting Project leader: **Dr Jean-Marie Le Goff**
- Team members: Dimitrios Dardanis, Richard Forster, André Rattinger and Xavier Ouvrard

## **More information:**

- <http://collspotting.web.cern.ch>
- <https://www.infos-informatique.net>
- [xavier.ouvrard@cern.ch](mailto:xavier.ouvrard@cern.ch)

# Bibliography I

Anirban Banerjee, Arnab Char, and Bibhash Mondal. Spectra of general hypergraphs. *Linear Algebra and its Applications*, 518:14–30, 2017.

Cynthia Dwork, Ravi Kumar, Moni Naor, and Dandapani Sivakumar. Rank aggregation methods for the web. In *Proceedings of the 10th international conference on World Wide Web*, pages 613–622. ACM, 2001.

Xavier Ouvrard, Jean-Marie Le Goff, and Stéphane Marchand-Maillet. Networks of collaborations: Hypergraph modeling and visualisation. *arXiv preprint arXiv:1707.00115*, 2017.

Lizhu Sun, Jiang Zhou, and Changjiang Bu. Spectral properties of general hypergraphs. *Linear Algebra and its Applications*, 2018.