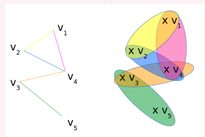
Hb-graphs and their applications

Xavier Ouvrard @ UniGe & CERN supervised by:
Pr. Stéphane Marchand-Maillet @ UniGe (CH) and Dr. Jean-Marie Le Goff @ CERN

Why do we need hb-graphs?

Graphs:
- Set of **vertices** and set of **edges**.
- An edge links two vertices.

In graphs:
- Links are 2-adic : **pairwise relationship**.
- Adjacency of two vertices correspond to the fact that 2 vertices are incident to an edge



Hypergraph view of a graph

- Sets regroup non-repeated elements

Hypergraphs:
- Extend graphs.
- Allow relations between multiple vertices.
- Are a family of **hyperedges** of unempty subsets of a **vertex** set.

In hypergraphs:
- Links are n-adic
- **Which level of adjacency we want ?**

=> Introduction of:
- k-**adjacency**: k vertices are said k-adjacent if they belong to the same hyperedge
- **e-adjacency**: vertices of a same hyperedge are said e-adjacent

Maximal k-adjacency a hypergraph can tolerate: $\overline{k}$-**adjacency**

In ξ-uniform hypergraphs:
- $\overline{k}$-adjacency = ξ-adjacency
- Vertices that are $\overline{k}$-adjacent are also e-adjacent and reciprocally

In general hypergraphs:
- Vertices that are $\overline{k}$-adjacent are e-adjacent but it is no more reciprocal!

- 2-adjacency is modeled by matrix relationship and dimension $|V|$
- $\overline{k}$-adjacency requires a tensor of rank $\overline{k}$ and dimension $|V|$

For ξ-uniform hypergraphs:
- **Cooper and Duttle 2012** propose a degree-normalised ξ-adjacency tensor:

⋆ $\mathcal{A} = \left\{a_{i_1\dots i_\xi}\right\}_{1\leqslant i_1,\dots,i_\xi\leqslant n}$ such that:

$$a_{i_1\dots i_\xi} = \frac{1}{(\xi-1)!}\begin{cases}1 & \text{if } \{v_{i_1},\dots,v_{i_\xi}\}\in E\\ 0 & \text{otherwise.}\end{cases}$$

⋆ Allows to retrieve degree of vertices:

$$\deg(v_i) = \sum_{i_2,\dots,i_\xi=1}^{n} a_{ii_2\dots i_\xi}.$$

⋆ Allows to have hypergraph spectral theory

For general hypergraphs:
- The Art of Filling ! or the Art of Cubing !

What about this?

=> **We need to store additional information**

The aim:
- Making a uniform structure from something that is non-uniform.
- To capture non uniformity: we need e-adjacency

Existing solutions for general hypergraphs:
- push everything in a tensor of order $|r_{\mathcal{H}}|$ and size $|V|$.

=> **Banerjee 2016**: split the storage in the elements of the tensor that is not occupied

The ([Author's note]: **e-1) adjacency hypermatrix** of $\mathcal{H}$ written $\mathcal{A}_{\mathcal{H}} = \left(a_{i_1\dots i_{max}}\right)_{1\leqslant i_1\dots i_{max}\leqslant n}$ is such that for a hyperedge: $\boldsymbol{e} = \{v_{i_1},\dots,v_{i_s}\}$ of cardinality $s\leqslant k_{max}$.

$$a_{p_1\dots p_{k_{max}}} = \frac{s}{\sigma}, \text{ where } \sigma = \sum_{\substack{k_1,\dots,k_s\geqslant 1\\ \sum_{i=1}^{s}k_i=k_{max}}}\frac{k_{max}!}{k_1!\dots k_s!}$$

with $p_1,\dots,p_{k_{max}}$ chosen in all possible way from $\{i_1,\dots,i_s\}$ with at least once from each element of $\{i_1,\dots,i_s\}$.

=> **Sun 2018**: similarly the same, but not symmetric.

The ([Author's note]: **prefix e-1adjacency hypermatrix** of a general hypergraph $\mathcal{H} = (V,E)$, with V as vertex set ([Author's note]: identified to $[\![n]\!]$) and E as hyperedge set, having range $r(\mathcal{H}) = k_{max}$, is an order k_{max} and dimension $|V|$ hypermatrix with entries:

$$a_{i_1\dots i_{k_{max}}} = \begin{cases}\dfrac{1}{(s-1)} & i_1 = \dots = i_{k-s+1}, \{i_{k-s+1},\dots,i_k\}\in E\\ 0 & \text{otherwise.}\end{cases}$$

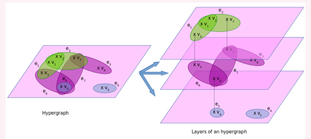
- tensor built using a hypergraph uniformisation process:

=> **Ouvrard 2017**: e-adjacency tensor for general hypergraphs.

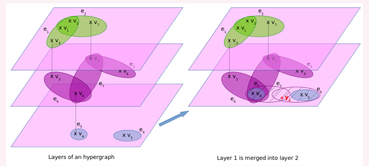
Let $\mathcal{H} = (V,E)$ with $V = \{v_1, v_2, \dots, v_n\}$ and family $E = \{e_1, e_2, \dots, e_p\}$.
Let $k_{max} = \max\{|e_i| : e_i\in E\}$ be the maximum cardinality of the family of hyperedges.
The **layered e-adjacency hypermatrix** of $\mathcal{H}$ written $\mathcal{A}_{\mathcal{H}} = \left(a_{i_1\dots i_{max}}\right)_{1\leqslant i_1\dots i_{max}\leqslant n}$ is such that for each hyperedge: $\boldsymbol{e} = \{v_{i_1},\dots,v_{i_s}\}$ of cardinality $s < k_{max}$ it is completed in a hyperedge $\overline{\boldsymbol{e}} = \{v_{i_1},\dots,v_{i_s},y_{s+1},\dots,y_{k_{max}-1}\}$.

$$a_{\sigma(i_1)\dots\sigma(i_s)\sigma(s+1)\dots\sigma(n+k_{max}-1)} = \frac{1}{(k_{max}-1)!}$$

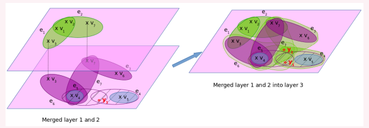
where $\sigma\in\mathcal{S}_{k_{max}}$.



Decomposition of the hypergraph in layers of uniform hypergraph



Additional element in the lower level and merging (step 1)



Additional element in the lower level and merging (step 2)

The filling is done with different vertices at each layer.

The way of filling is not unique:
- Add always the same vertex

=> straightforward approach
- A different vertex per uniform hypergraph layer

=> silo approach
- A vertex per layer (previous approach)

=> iterative approach

In the first two approaches **we need multisets** to keep interpretability

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

⋆ **Support** of the multiset A_m^*: $A_m^* = \{x\in A : m(x)\neq 0\}$

⋆ **m-cardinality** of a multiset A_m: $\#_m A_m = \sum_{x\in A} m(x)$.

⋆ $\mathcal{A}$ is **included** in $\mathcal{B}$ ($\mathcal{A}\subseteq\mathcal{B}$) if $\forall x\in U$: $m_{\mathcal{A}}(x)\leqslant m_{\mathcal{B}}(x)$.
In this case: $\mathcal{A}$ is a **submset** of $\mathcal{B}$.

⋆ **Union** of $\mathcal{A}$ and $\mathcal{B}$: mset $\mathcal{C} = \mathcal{A}\cup\mathcal{B}$, universe U,

$$\forall x\in U : m_{\mathcal{C}}(x) = \max\left(m_{\mathcal{A}}(x), m_{\mathcal{B}}(x)\right).$$

⋆ **Intersection** of $\mathcal{A}$ and $\mathcal{B}$: mset $\mathcal{D} = \mathcal{A}\cap\mathcal{B}$, universe U,

$$\forall x\in U : m_{\mathcal{D}}(x) = \min\left(m_{\mathcal{A}}(x), m_{\mathcal{B}}(x)\right).$$

⋆ **Sum** of $\mathcal{A}$ and $\mathcal{B}$: mset $\mathcal{E} = \mathcal{A}\uplus\mathcal{B}$, universe U,

$$\forall x\in U : m_{\mathcal{E}}(x) = m_{\mathcal{A}}(x) + m_{\mathcal{B}}(x).$$

⋆ **Power multiset** of $\mathcal{A}$: set $\widetilde{\mathcal{P}}(\mathcal{A})$ of all submsets of $\mathcal{A}$.

Hb-graphs

HyperBag-graphs (Hb-graphs):
- **Hb-graph** $\mathcal{H} = (V,E)$: family of multisets $E = (\boldsymbol{e}_i)_{i\in I}$, with $I = [\![p]\!]$ - 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 $m_{\boldsymbol{e}} : V\to\mathbb{W}$ where $\mathbb{W}\subset\mathbb{R}^+$.

⋆ **Hb-graph with no repeated hb-edge**:

$$\forall i_1\in I, \forall i_2\in I : \boldsymbol{e}_{i_1} = \boldsymbol{e}_{i_2}\Rightarrow i_1 = i_2$$

⋆ **Order** of a hb-graph $\mathcal{H}$: $O(\mathcal{H}) = \sum_{j=1}^{n}\max_{e\in E}\left(m_{\boldsymbol{e}}(v_j)\right)$.

⋆ **Size** of a hb-graph: $|E|$.

⋆ **Natural hb-graph**: when all multiplicity functions have their range included in $\mathbb{N}$

⋆ **Support hypergraph** $\underline{\mathcal{H}}$: hypergraph of the support of the multisets

⋆ **Star of a vertex**: $\forall x\in V$: $H(v) = \left\{\boldsymbol{e}_i^{m_{\boldsymbol{e}_i}(v)} : \boldsymbol{e}_i\in E\wedge v\in\boldsymbol{e}_i^*\right\}$.

⋆ **m-degree of a vertex**: $\deg_m(v) = \#_m H(v) = \sum_{\boldsymbol{e}_i\in E} m_{\boldsymbol{e}}(v)$.

⋆ **m-range** of a hb-graph: $r_m(\mathcal{H}) = \max_{e\in E}\#_m\boldsymbol{e}$.

⋆ **m-co-range** of a hb-graph: $cr_m(\mathcal{H}) = \min_{e\in E}\#_m\boldsymbol{e}$.

⋆ **k-m-uniform** hb-graph: all its hb-edges of same m-cardinality k.

⋆ A hb-graph $\mathcal{H}$ is k-m-uniform if and only if $r_m(\mathcal{H}) = cr_m(\mathcal{H}) = k$.

⋆ A hypergraph can be seen as a natural hb-graph with multiplicity function ranges in $\{0,1\}$

⋆ Paths can be defined in hb-graphs

Incidence matrix of hb-graphs:

⋆ Hb-edges are **incident** if their intersection is not empty

⋆ **Incidence matrix of the hb-graph** $\mathcal{H}$: $H = \left[m_j(v_i)\right]_{\substack{1\leqslant i\leqslant n\\ 1\leqslant j\leqslant p}}$

⋆ Incidence is a pairwise concept: a vertex is incident to a hb-edge.

⋆ The rows allow to see which hb-edges are incident: linked by rows.

Different application of hb-graphs:
- Network of co-occurences 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.

Hb-graph e-adjacency tensor(s)

Elementary hb-graph:

⋆ Hb-graph with only one non repeated hb-edge in its hb-edge family. Typically: $\mathcal{H}_e = (V,\{\boldsymbol{e}\})$.

⋆ For $\mathcal{H}_e = (V,\{\boldsymbol{e}\})$:

◎ $\boldsymbol{e}$ described uniquely by its hypermatrix representation $Q_{\boldsymbol{e}}$.

◎ $\mathcal{H}_e$ also uniquely described by $Q_{\boldsymbol{e}}$.

⋆ Let $\boldsymbol{e} = \left\{v_{j_1}^{m_{j_1}},\dots,v_{j_k}^{m_{j_k}}\right\}$ be a hb-edge of $\mathcal{H}$ of m-rank r.
Normalised $\overline{k}$-adjacency hypermatrix of an elementary hb-graph $\mathcal{H}_e$:

◎ symmetric hypermatrix $Q_{\boldsymbol{e}} = \left(q_{i_1\dots i_r}\right)$ of rank r and dimension n

◎ only nonzero elements are:

$$q_{\sigma(j_1)^{m_{j_1}}\dots\sigma(j_k)^{m_{j_k}}}(\boldsymbol{e}) = \frac{m_{j_1}!\dots m_{j_k}!}{(r-1)!}$$

where $\sigma\in\mathcal{S}_{[r]}$.

⋆ **Claim**: Let $\mathcal{H} = (V,E)$ be a hb-graph with no repeated hb-edge. $V = \{v_i : i\in[\![n]\!]\}$, $\mathcal{H} = \bigoplus_{e\in E}\mathcal{H}_e$, $\mathcal{H}_e = (V,\{\boldsymbol{e}\})$

Elementary operations on hb-graphs

⋆ Are needed for building the hb-graph m-uniformisation process

⋆ **Canonical weighting operation**: $\phi_{can} : \mathcal{H} = (V,E)\mapsto\mathcal{H}_1 = (V,E,\boldsymbol{w}_1)$ where $\forall\boldsymbol{e}\in E$: $\boldsymbol{w}_1(\boldsymbol{e}) = 1$.

⋆ **c-dilatation operation**: $\phi_{c,d} : \mathcal{H}_1 = (V,E,\boldsymbol{w}_1)\mapsto\mathcal{H}_c = (V,E,\boldsymbol{w}_c)$ with $\forall\boldsymbol{e}\in E$: $\boldsymbol{w}_c(\boldsymbol{e}) = c, c\in\mathbb{R}^{*+}$.

⋆ **y-complemented operation**: $\phi_{\overline{y}c} : \mathcal{H}_w = (V,E,\boldsymbol{w})\mapsto\widehat{\mathcal{H}}_{\overline{y}} = \left(\widehat{V},\widehat{E},\widehat{\boldsymbol{w}}\right)$, where $\widehat{\mathcal{H}}_{\overline{y}}$ is the **y-complemented hbgraph** of $\mathcal{H}_w$ with:

◎ $\widehat{V} = V\cup\{y\}$,

◎ $\widehat{E} = (\xi(\boldsymbol{e}))_{\boldsymbol{e}\in E}$ where $\xi : E\to\mathcal{M}\left(\widehat{V}\right)$ is such that: $\forall\boldsymbol{e}\in E$, $\xi(\boldsymbol{e}) = \left\{x^{m_{\xi(\boldsymbol{e})}(x)} : x\in\widehat{V}\right\}\in\mathcal{M}\left(\widehat{V}\right)$, with $m_{\xi(\boldsymbol{e})}(x) = \begin{cases}m_{\boldsymbol{e}}(x) & \text{if } x\in\boldsymbol{e}^*\\ r_{\mathcal{H}}-\#_m\boldsymbol{e} & \text{if } x = y\end{cases}$

◎ $\widehat{\boldsymbol{w}}$ is such that $\forall\boldsymbol{e}\in E$: $\widehat{\boldsymbol{w}}(\xi(\boldsymbol{e})) = \boldsymbol{w}(\boldsymbol{e})$.

⋆ **Decomposition operation**: $\phi_d : \mathcal{H}\mapsto\{\mathcal{H}_i\}_{i\in I}$ such that: $\mathcal{H} = \bigoplus_{i\in I}\mathcal{H}_i$ where the $\mathcal{H}_i$have all same universe and no pair of $\mathcal{H}_i$ have repeated hb-edge.

⋆ **y^{σ}-vertex-increasing operation**: $\phi_{y^{\sigma}\uparrow} : \mathcal{H} = (V,E,\boldsymbol{w})\mapsto\mathcal{H}^+ = (V^+,E^+,\boldsymbol{w}^+)$ where $\mathcal{H}^+$ is the

◎ $V^+ = V\cup\{y\}$,

◎ $E^+ = (\phi(\boldsymbol{e}))_{\boldsymbol{e}\in E}$ with the map $\phi : E\to\mathcal{M}(V^+)$ such that for all $\boldsymbol{e}\in E$, $\phi(\boldsymbol{e}) = \left\{x^{m_{\phi(\boldsymbol{e})}(x)} : x\in V^+\right\}\in\mathcal{M}(V^+)$ with $m_{\phi(\boldsymbol{e})}(x) = \begin{cases}m_{\boldsymbol{e}}(x) & \text{if } x\in\boldsymbol{e}^*\\ \alpha & \text{if } x = y\end{cases}$

◎ $\boldsymbol{w}^+$ is such that $\forall\boldsymbol{e}\in E$: $\boldsymbol{w}^+(\phi(\boldsymbol{e})) = \boldsymbol{w}(\boldsymbol{e})$.

⋆ **Merging operation**: $\phi_m : \{\mathcal{H}_{in},\mathcal{H}_{ob}\}\mapsto\widehat{\mathcal{H}}_{\overline{m}}$ where: $\widehat{\mathcal{H}}_{\overline{m}} = \left(\widehat{V},\widehat{E},\widehat{\boldsymbol{w}}\right)$ is the **merged hb-graph** of two weighted hb-graphs $\mathcal{H}_{in} = (V_{in},E_{in},\boldsymbol{w}_{in})$ and $\mathcal{H}_{ob} = (V_{ob},E_{ob},\boldsymbol{w}_{ob})$

◎ $\widehat{V} = V_{in}\cup V_{ob}$

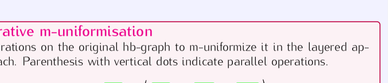
◎ $\widehat{E} = \left(\psi(\boldsymbol{e})\right)_{\boldsymbol{e}\in E_{in}\oplus E_{ob}}$ where $\psi : E_{in}+E_{ob}\to\mathcal{M}\left(\widehat{V}\right)$ such that for all $\boldsymbol{e}\in E_{in}+E_{ob}$, $\psi(\boldsymbol{e}) = \left\{x^{m_{\psi(\boldsymbol{e})}(x)} : x\in\widehat{V}\right\}\in\mathcal{M}\left(\widehat{V}\right)$ with $m_{\psi(\boldsymbol{e})}(x) = \begin{cases}m_x(x) & \text{if } x\in\boldsymbol{e}^*\\ 0 & \text{otherwise}\end{cases}$

◎ $\forall\boldsymbol{e}\in E_{in}$, $\widehat{\boldsymbol{w}}(\boldsymbol{e}) = \boldsymbol{w}_{in}(\boldsymbol{e})$ and $\forall\boldsymbol{e}\in E_{ob}$ $\widehat{\boldsymbol{w}}(\boldsymbol{e}) = \boldsymbol{w}_{ob}(\boldsymbol{e})$.

⋆ These operations preserve e-adjacency either globally or exactly

Straightforward m-uniformisation

Operations on the original hb-graph to m-uniformize it in the straightforward approach.
Parenthesis with vertical dots indicate parallel operations.



⋆ The transformation $\phi_s : \mathcal{H}\mapsto\widehat{\mathcal{H}}_{s,d}$ preserves the e-adjacency.

⋆ The **straightforward e-adjacency hypermatrix of a hb-graph** $\mathcal{H} = (V,E)$ is the hypermatrix $\widehat{\boldsymbol{A}}_{\mathcal{H},\mathcal{H}}$ defined by:

$$\widehat{\boldsymbol{A}}_{\mathcal{H},\mathcal{H}} = sum_{i\in[\![p]\!]} c_{e_i}\boldsymbol{R}_{\boldsymbol{e}_i},$$

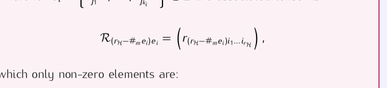
where for $\boldsymbol{e}_i = \left\{v_{j_1}^{m_{j_1}},\dots,v_{j_k}^{m_{j_k}}\right\}\in E$ the associated hypermatrix is: $\boldsymbol{R}_{\boldsymbol{e}_i} = \left(r_{i_1\dots i_{r_{\mathcal{H}}}}\right)$, which only non-zero elements are:

$$r_{j_1^{m_{j_1}}\dots j_k^{m_{j_k}}(m+1)^{m_{m+1}}} = \frac{m_{j_1}!\dots m_{j_k}!m_{l+1}!}{(r-1)!}$$

- with $m_{l+1} = r_{\mathcal{H}} - \#_m\boldsymbol{e}_i$ - and the ones with same value and obtained by permutation of the indices and where $c_{e_i} = \dfrac{r_{\mathcal{H}}}{\#_m\boldsymbol{e}_i}$.

Silo m-uniformisation

Operations on the original hb-graph to m-uniformize it in the silo approach.
Parenthesis with vertical dots indicate parallel operations.



⋆ The transformation $\phi_s : \mathcal{H}\mapsto\widehat{\mathcal{H}}_{\overline{m}}$ preserves the e-adjacency.

⋆ The **silo e-adjacency hypermatrix of a hb-graph** $\mathcal{H} = (V,E)$ is the hypermatrix $\widehat{\boldsymbol{A}}_{\mathcal{H},\mathcal{H}} = \left(a_{i_1\dots i_{r_{\mathcal{H}}}}\right)_{1\leqslant i_1\dots i_{r_{\mathcal{H}}}\leqslant[\![n]\!]}$ defined by $\widehat{\boldsymbol{A}}_{\mathcal{H},\mathcal{H}} = \sum_{i\in[\![p]\!]} c_{e_i}\boldsymbol{R}_{\boldsymbol{e}_i}$

and where for $\boldsymbol{e}_i = \left\{v_{j_1}^{m_{j_1}},\dots,v_{j_k}^{m_{j_k}}\right\}\in E$ the associated hypermatrix is: $\boldsymbol{R}_{\boldsymbol{e}_i} = \left(r_{i_1\dots i_{r_{\mathcal{H}}}}\right)$, which only non-zero elements are:

$$r_{j_1^{m_{j_1}}\dots j_k^{m_{j_k}}(m+\#_m\boldsymbol{e}_i)^{m_{l+m}\#_m\boldsymbol{e}_i}} = \frac{m_{j_1}!\dots m_{j_k}!m_{l+\#_m\boldsymbol{e}_i}!}{r_{\mathcal{H}}!}$$

and all elements of $\boldsymbol{R}_{\boldsymbol{e}_i}$obtained by permutation of the indices and with:

$$m_{l+\#_m\boldsymbol{e}_i} = r_{\mathcal{H}} - \sum_{i\in[\![k]\!]} m_{j_i},$$

and where:

$$c_{e_i} = \frac{r_{\mathcal{H}}}{\#_m\boldsymbol{e}_i}.$$

$\overline{k}$-adjacency hypermatrix of a r-m-uniform hb-graph

Let $\mathcal{H} = (V,E)$ be a **r-m-uniform** hb-graph. $V = \{v_i : i\in[\![n]\!]\}$.

⋆ each hb-edge in a r-m-uniform hb-graph has same m-cardinality r

⋆ **$\overline{k}$-adjacency hypermatrix of** $\mathcal{H}$ is the hypermatrix $A_{\mathcal{H}} = \left(a_{i_1\dots i_r}\right)_{1\leqslant i_1,\dots,i_r\leqslant n}$ defined by:

$$A_{\mathcal{H}} = sum_{i\in[\![p]\!]} Q_{e_i}$$

where Q_{e_i} is the $\overline{k}$-adjacency hypermatrix of the elementary hb-graph associated to the hb-edge $\boldsymbol{e}_i = \left\{v_{j_1}^{m_{j_1}},\dots,v_{j_k}^{m_{j_k}}\right\}\in E$.

⋆ The only non-zero elements of $Q_{\boldsymbol{e}_i}$ are the elements of indices obtained by permutation of the multiset $\left\{j_1^{m_{j_1}},\dots,j_k^{m_{j_k}}\right\}$ and are all equals to $\dfrac{m_{j_1}!\dots m_{j_k}!}{(r-1)!}$

⋆ Remark: when the r-m-uniform hb-graph corresponds to a r-uniform hypergraph => retrieve the result of the degree-normalized tensor of Cooper and Duttle.

⋆ Claim: The m-degree of a vertex v_j in a r-m-uniform hb-graph $\mathcal{H}$ of $\overline{k}$-adjacency hypermatrix $\mathcal{A}_{\mathcal{H}}$ is:

$$\deg_m(v_j) = \sum_{1\leqslant i_2,\dots,i_r\leqslant n} a_{jj_2\dots j_r}.$$

e-adjacency hypermatrix of a general hb-graph

⋆ Each hb-graph can be summarized by a polynomial of degree $r_{\mathcal{H}}$:

$$P(z_0) = \sum_{i=1}^{p} c_{e_i}P_{e_i}(z_0) = \sum_{i=1}^{p} c_{e_i}\frac{r_!}{m_{j_1}!\dots m_{j_k}!}q_{j_1^{m_{j_1}}\dots j_k^{m_{j_k}}}z_{j_1}^{m_{j_1}}\dots z_{j_k}^{m_{j_k}}$$

⋆ c_{e_i} is a technical coefficient => choosen to retrieve the m-degree of the vertices from the e-adjacency tensor.

Processes involved for building the e-adjacency tensor:

⋆ **Hb-graph m-uniformisation process** (Hm-UP): transform:

◎ a hb-graph $\mathcal{H}$ of m-range $r_{\mathcal{H}}$

◎ into a $r_{\mathcal{H}}$-m-uniform hb-graph written $\overline{\mathcal{H}}$

⋆ **Polynomial homogenization process** (PHP): homogeneize the hb-graph polynomial

Choice of the c_{e_i} done such that the e-adjacency hypermatrix $A = \left(a_{i_1\dots i_r}\right)_{1\leqslant i_1,\dots,i_r\leqslant[\![n]\!]}$ allows to retrieve:

⋆ the m-degree of the vertices: $\sum_{i_2,\dots,i_r\in[\![n]\!]} a_{ii_2\dots i_r} = \mathbf{deg}_m(v_i)$.

⋆ the number of hb-edges $|E|$.

The hb-graph is decomposed in layers of r-m-uniform hb-graphs

⋆ $\mathcal{H} = \bigoplus_{r\in[\![r_{\mathcal{H}}]\!]}\mathcal{H}_r$.

⋆ Each $\mathcal{H}_r$ can be associated to a $\overline{k}$-adjacency tensor $\mathcal{A}_r$ viewed as a hypermatrix $A_{\mathcal{H}_r} = \left(a_{ji_2\dots i_r}\right)$ of order r, hypercubic and, symmetric of dimension $|V| = n$.

⋆ The number of hb-edges is kept constant in the decomposition of a hb-graph in layers.

⋆ The processes involved in the uniformisation processes keep the number of hb-edges constant.

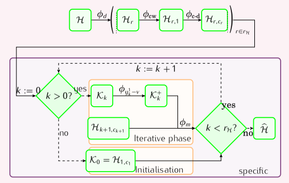
⋆ Generalised hand-shake lemma: $\sum_{i_1,\dots,i_{r_{\mathcal{H}}}\in[\![n]\!]} a_{ii_2\dots i_{r_{\mathcal{H}}}} = r_{\mathcal{H}}|E|$.

⋆ We want: $|E| = \sum_{r=1}^{n}|E_r|$, it imposes: $\sum_{i_1,\dots,i_{r_{\mathcal{H}}}\in[\![n]\!]} a_{i_1\dots i_{r_{\mathcal{H}}}} = \sum_{r=1}^{n}\dfrac{r_{\mathcal{H}}}{r}\sum_{i_1,\dots,i_r\in[\![n]\!]} a_{ji_2\dots i_r}$.

=> $\forall r\in[\![r_{\mathcal{H}}]\!]$: $c_r = \dfrac{r_{\mathcal{H}}}{r}$

Iterative m-uniformisation

Operations on the original hb-graph to m-uniformize it in the layered approach. Parenthesis with vertical dots indicate parallel operations.



⋆ The transformation $\phi_s : \mathcal{H}\mapsto\widehat{\mathcal{H}}_{\overline{m}}$ preserves the e-adjacency.

⋆ The **layered e-adjacency tensor of a hb-graph** $\mathcal{H} = (V,E)$ is the tensor $\mathcal{A}_{lay}(\mathcal{H}) = \left(a_{i_1\dots i_{r_{\mathcal{H}}}}\right)_{1\leqslant i_1,\dots,i_{r_{\mathcal{H}}}\leqslant n}$ defined by:

$$\mathcal{A}_{lay}(\mathcal{H}) = sum_{i\in[\![p]\!]} c_{e_i}\mathcal{R}_{(r_{\mathcal{H}}-\#_m\boldsymbol{e}_i)\boldsymbol{e}_i}$$

where for $\boldsymbol{e}_i = \left\{v_{j_1}^{m_{j_1}},\dots,v_{j_k}^{m_{j_k}}\right\}\in E$ the associated tensor is:

$$\mathcal{R}_{(r_{\mathcal{H}}-\#_m\boldsymbol{e}_i)\boldsymbol{e}_i} = \left(r_{(r_{\mathcal{H}}-\#_m\boldsymbol{e}_i)i_1\dots i_{r_{\mathcal{H}}}}\right),$$

which only non-zero elements are:

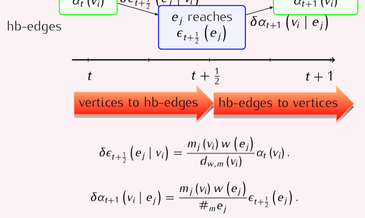
$$r_{(r_{\mathcal{H}}-\#_m\boldsymbol{e}_i)j_1^{m_{j_1}}\dots j_k^{m_{j_k}}(m+\#_m\boldsymbol{e}_i)^{l\dots(m+r_{\mathcal{H}}-1)^l}} = \frac{m_{j_1}!\dots m_{j_k}!}{r_{\mathcal{H}}!}$$

and all elements of $\mathcal{R}_{\boldsymbol{e}_i}$with indices obtained by permutation and **where:**
$c_{e_i} = \dfrac{r_{\mathcal{H}}}{\#_m\boldsymbol{e}_i}$.

⋆ $\mathcal{A}_{lay}(\mathcal{H})$ can also be written: $\mathcal{A}_{lay}(\mathcal{H}) = \sum_{r\in[\![r_{\mathcal{H}}]\!]} c_r\sum_{e_i\in\{e\,|\,\#_m\boldsymbol{e}=r\}}\mathcal{R}_{\boldsymbol{e}_i}$,where $c_r = \dfrac{r_{\mathcal{H}}}{r}$.

The final choice of tensor:
1. based on intrinsinc properties: symmetry, weak irreducibility
2. ensures the retrieval of degrees and additional information on the m-uniformity of the hb-graphs
3. requires to be benchmarked vs the diffusion schema

Diffusion in hb-graphs



$$\delta\epsilon_{t+\frac{1}{2}}(\boldsymbol{e}_j\,|\,v_i) = \frac{m_j(v_i)\,w(\boldsymbol{e}_j)}{d_{w,m}(v_i)}\alpha_t(v_i).$$

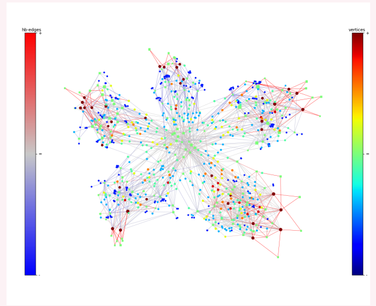
$$\delta\alpha_{t+1}(v_i\,|\,\boldsymbol{e}_j) = \frac{m_j(v_i)\,w(\boldsymbol{e}_j)}{\#_m\boldsymbol{e}_j}\epsilon_{t+\frac{1}{2}}(\boldsymbol{e}_j).$$

- Conservative process
- Valuation of vertices and hb-edges
- Initialisation on an equal foot
- Stochastic process that converges to:

$$\pi_V = \left(\frac{d_{w,v_i}}{\sum_{k=1}^{n}d_{w,v_k}}\right)_{1\leqslant i\leqslant n}$$

and for the hb-edges:

$$\pi_E = \left(\frac{w_{\boldsymbol{e}}(\boldsymbol{e}_j)\times\#_m(\boldsymbol{e}_j)}{\sum_{k=1}^{n}d_{w,v_k}}\right)_{1\leqslant j\leqslant p}.$$



The HyperBagGraph DataEdron: An Enriched Browsing Experience of Scientific Publications

Xavier Ouvrard @ UniGe & CERN supervised by:
Pr. Stéphane Marchand-Maillet @ UniGe (CH) and Dr. Jean-Marie Le Goff @ CERN

This application is a small brother of the **Collaboration Spotting project @ CERN**
- Collspotting Project leader: Dr Jean-Marie Le Goff
- 1 fellowship, 3 PhD students

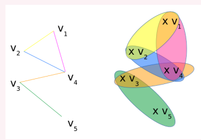
http://collspotting.web.cern.ch

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

Graph:
- Set of **vertices** and set of **edges**.
- An edge links two vertices : **pairwise relationship**.

Sets:
- Regroup elements with no repetition and no order.

Hypergraphs:
- Extend graphs.
- Allow relations between multiple vertices.
- Are a family of **hyperedges** of unempty subsets of a **vertex** set.

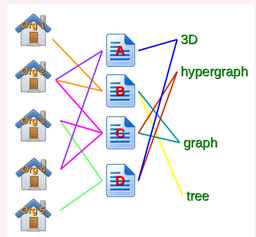


Natural multisets:
- Collection of objects with allowed repetitions.
- Constituted of a universe and of a multiplicity function on this universe
- Natural hb-graphs: the values of the multiplicity function are integer

HyperBag-graphs (Hb-graphs):
- Extend hypergraphs
- Allow duplication of elements
- Are a family of multisets - called **hb-edges** - of same universe called the **vertex** set
- Natural hb-graphs: use natural multisets

In a Scientific Publication Database:
- Metadata store information on the structure of the DataBase
- Metadata have types, that can be used either as dimension or as reference
- Data instances attached to one type of metadata can be regrouped by using a reference => we talk about co-occurences
- **Co-occurences** are **n-adic relationships.**
- Co-occurences are **multisets**, often reduced to **sets**.

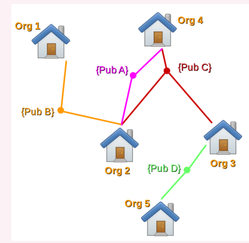
On an example:



We can retrieve in a simplified way a **hypergraph**: we use as **reference** the publication and the **co-occurences** of organisations seen as subsets of all organisations:

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

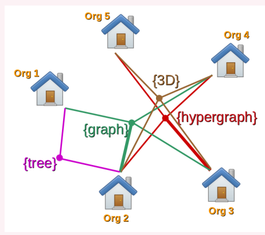
that we can visualize with an extra-node representation:



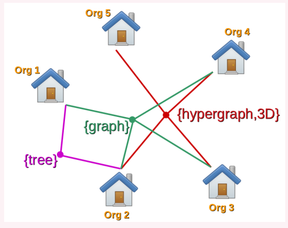
Choosing as reference keywords, we retrieve family of multisets of organisations, called a **hb-graph**

tree	$\left\{\left\{\text{Org }1^1,\text{Org }2^1\right\}\right\}$
graph	$\left\{\left\{\text{Org}1^1,\text{Org}2^2,\text{Org }3^1\right\}\right\}$
hypergraph	$\left\{\left\{\text{Org }2^1,\text{Org }3^2,\text{Org }4^1,\text{Org}5^1\right\}\right\}$
3D	$\left\{\left\{\text{Org }2^1,\text{Org }3^1,\text{Org }4^1,\text{Org}5^1\right\}\right\}$

Represented by a bundled extra-node multipartite graph representation:



Which can be reduced as a hypergraph:

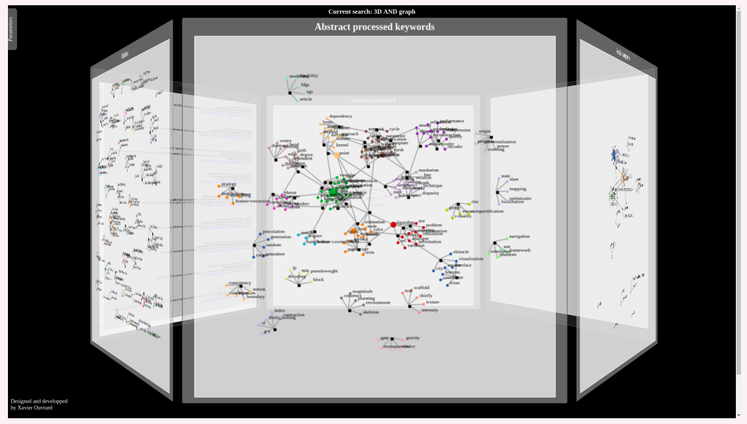


Searching on a Scientific Publication Database

With traditional verbatim browser:
- Give linear information
- To refine information: perform a new search
- Making complex query can be hazardous for most people
- Accessing the different facets of the information space require to perform different searches

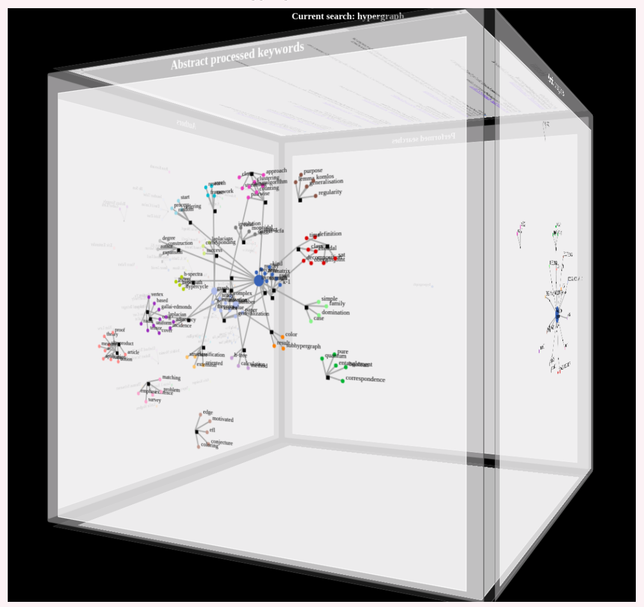
But in fact:
- A space of information is multi-facetted
- Much more information is available or can be extracted
- Use of natural language processing allow to extract keywords
- Hb-graphs highlight how the data instances are linked and allow additional information to be displayed

Switching between the different facets of the information space



Carousel view of the information space

Toggling to the DataEdron



The Hb-graph DataEdron:
switching between the different facets of the information space

Additional information can be displayed:
- Dblp / Linkedin profile
- Publication abstract / full article