Hypergraph Modeling and Visualisation of Complex Co-occurence Networks

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

- Project leader: J.M. Le Goff
- 4 PhD students

http://collspotting.web.cern.ch

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

Graphs (1)

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

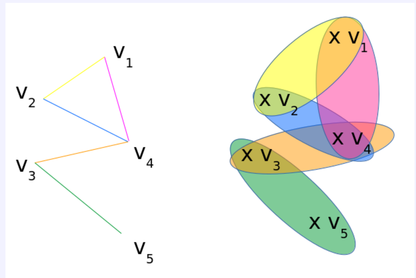
Collaborations and (multi)sets (2)

- **Sets:** group elements with no repetition and no order.
- **Natural multisets**: collection of objects with allowed repetitions.
- **Co-occurences** are *n*-adic **relationships.**
- Co-occurences are **multisets**, often reduced to **sets**.

Hypergraphs (3)

- **Hypergraphs:**
  - $\mathcal{H} = (V, E(, i))$.
  - extend graphs.
  - allow relations between n vertices.
  - are a family of **hyperedges** E of unempty subsets of vertices set V.
  - **incident function** is optional:

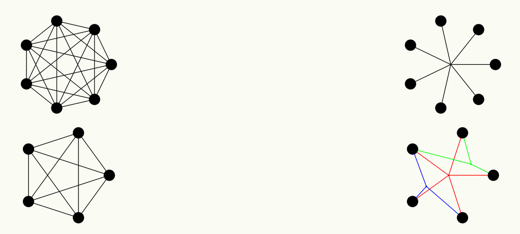
$$i : E \to \mathcal{P}(V).$$



- **Hypergraphs are suited for co-occurence modeling.**

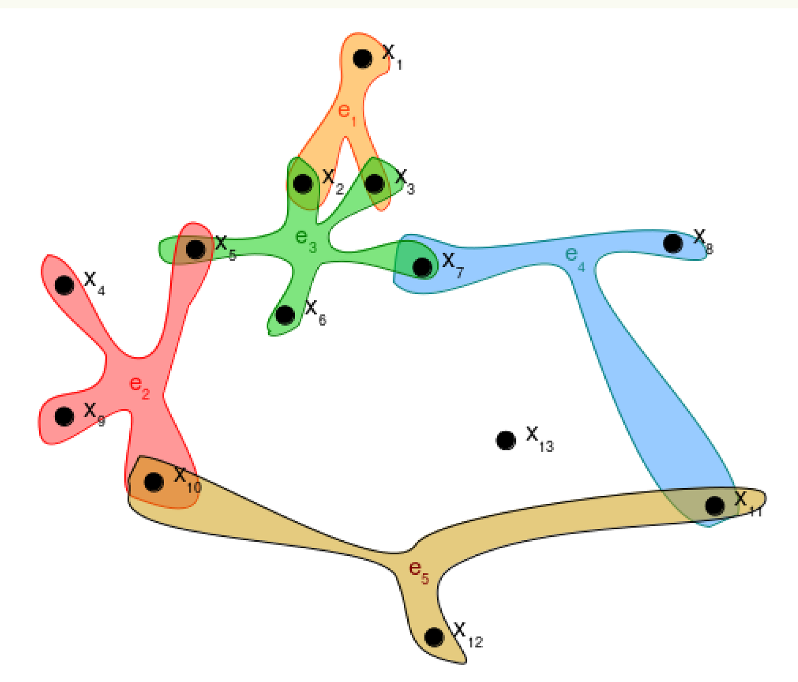
Hypergraph representations (11)

- **Edge representation**



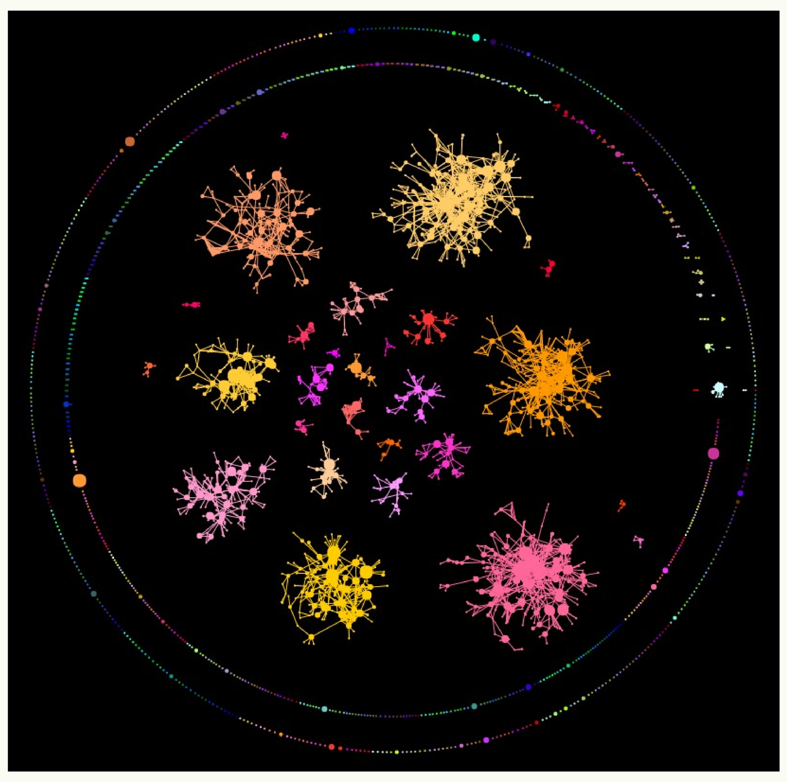
Clique representation — Extra-node representation

- **Set representation**

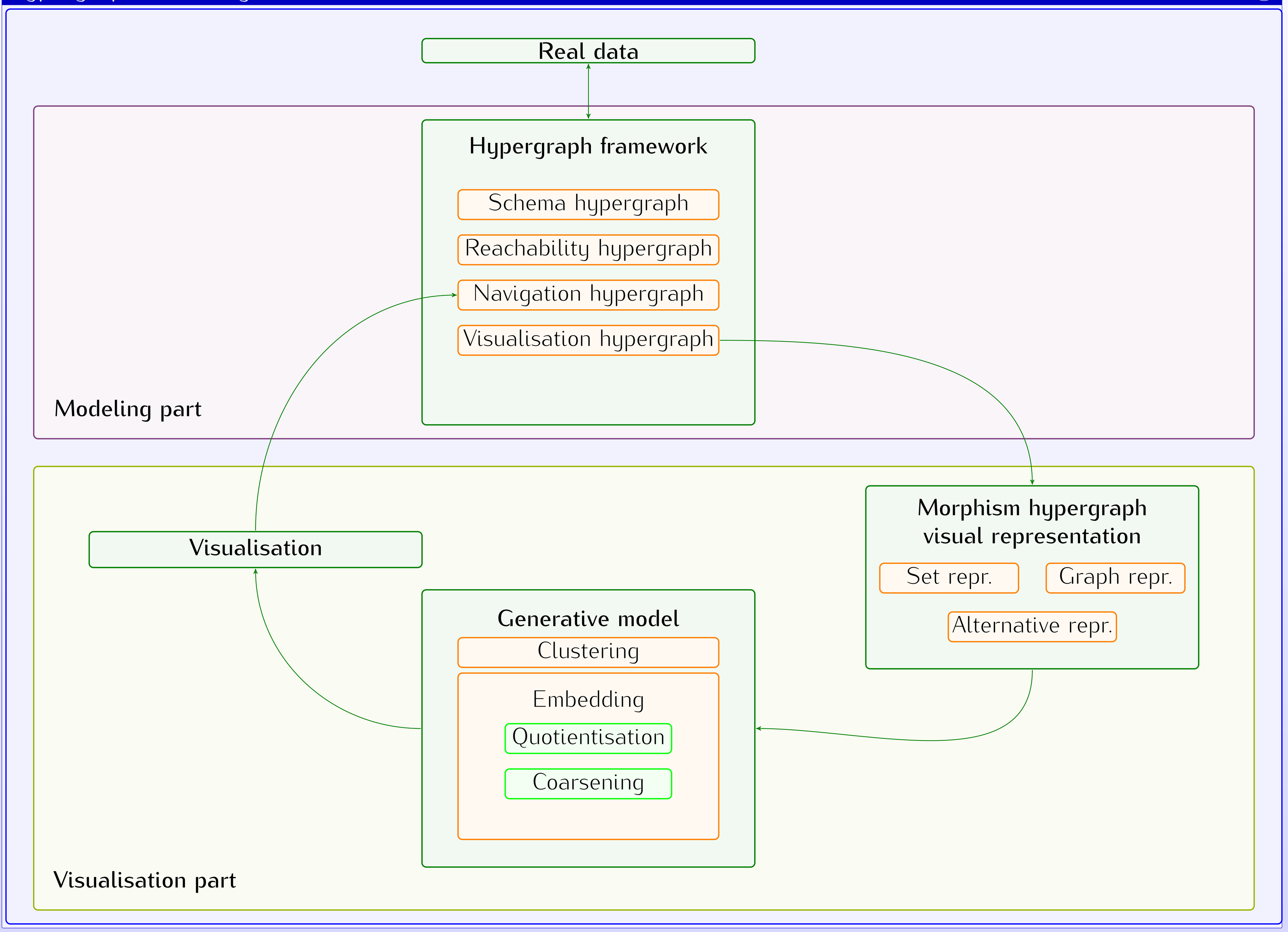


"PaintSplash" representation of hypergraph

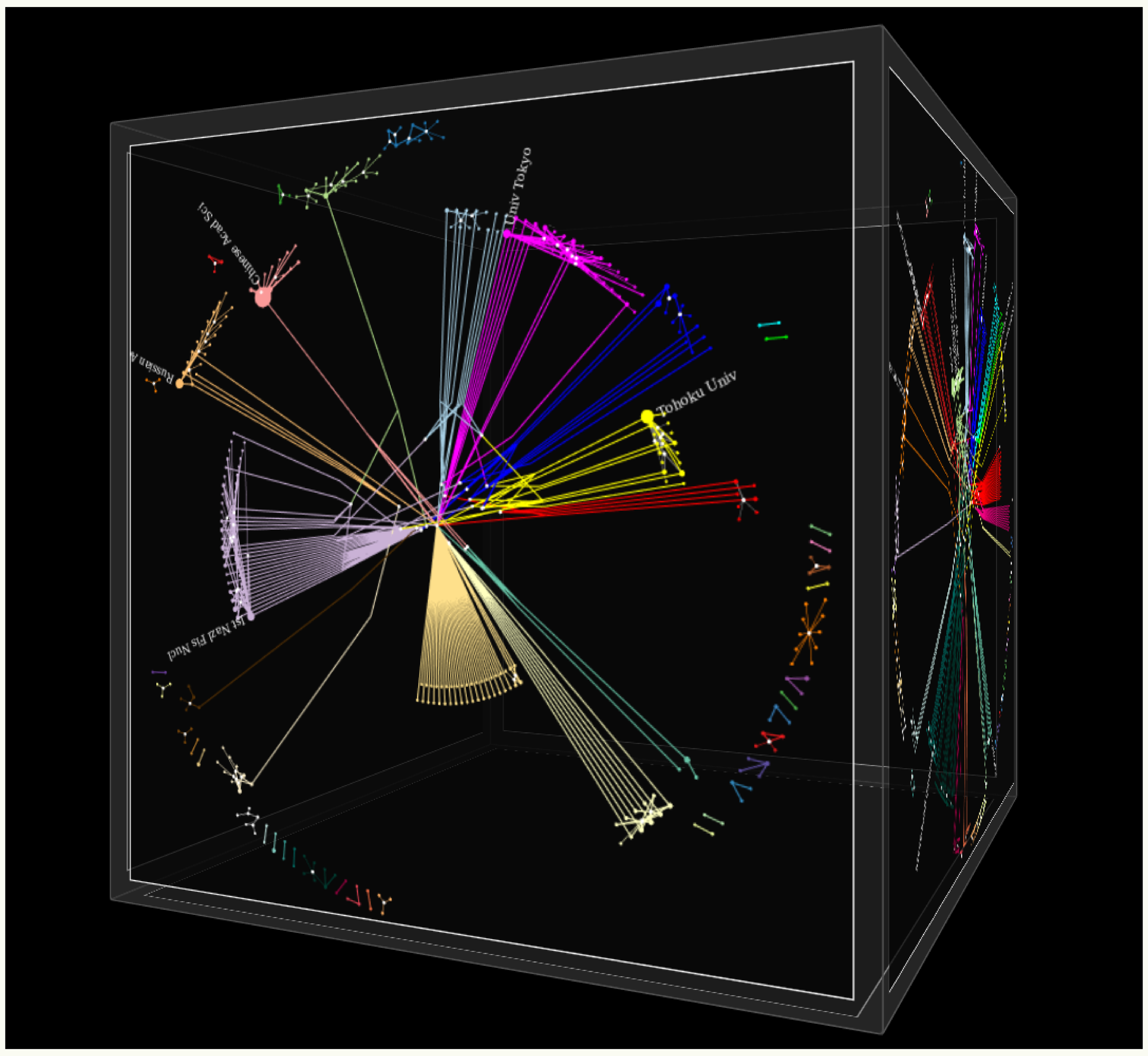
Large co-occurence hypergraph (12)



Hypergraph Modeling and Visualisation Framework (4)



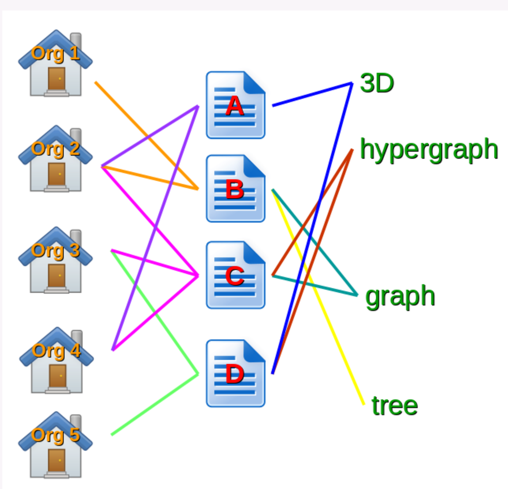
DataEdron



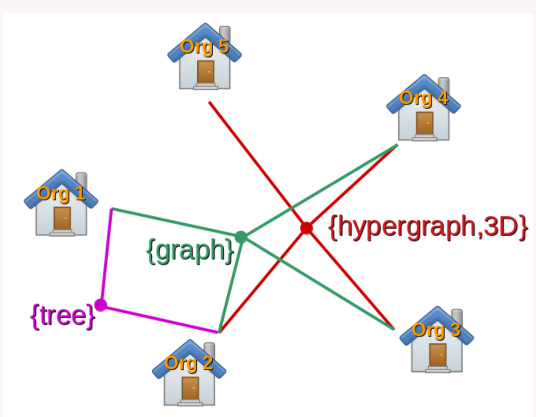
Improving visualisation in large hypergraphs (13)

- **Aim:**
  - keep meaningful information and structures;
  - focus on main information;
  - spreading the information uniformly on the layout.
- **How can it been achieved?**
  - coarsening the visualisation hypergraph;
- **What has been achieved?**
  - retrieval of important nodes using a diffusion process;
  - detection of important hyperedges of the network;
  - spectral comparison of the coarsened hypergraph with the original one.

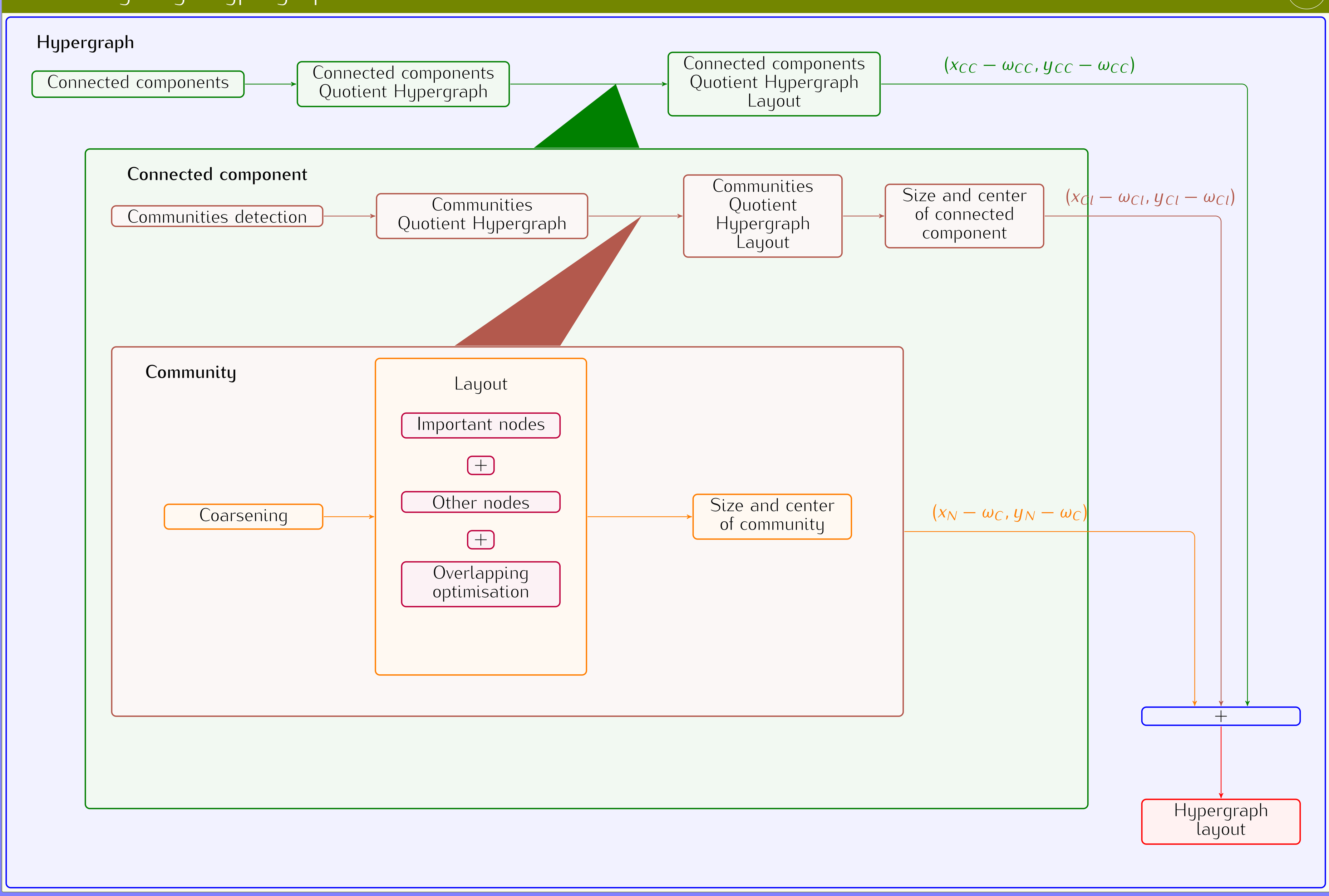
Building co-occurences on a publication dataset (10)



- **Aim:** Visualize **co-occurences of organisations** in reference to keywords.
- **Choose a type**:
  - α to visualize => organisations;
  - ρ to use as reference for co-occurences => keywords.
- **Result:** hypergraph with:
  - vertices: organisations;
  - hyperedges: organisation co-occurences.



Visualising large hypergraphs (14)



More info? (15)

- Find my work on Arxiv:

  1707.00115, 1712.08189, 1805.11952v2
- Read more on diffusion by exchange:

  IEEE CBMI 2018 Proceedings.
- Read more on this:

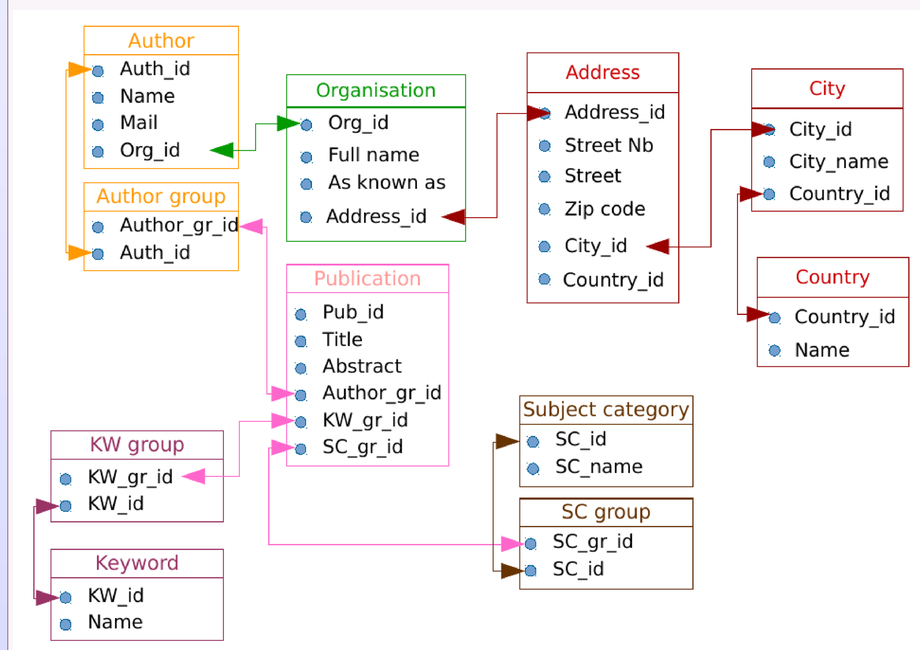
  Proceedings of the 2nd IMA CTCDM.
- Contact: xavier.ouvrard@cern.ch

Schema hypergraph (5)

- In relational databases:
  - metadata instances => vertices.
  - tables => hyperedges.
  - foreign keys allow connection between hyperedges.
- In graph databases:
  - schema represents link between metadata.
  - schema not compulsory.

=> can be represented by a hypergraph.

- **Schema hypergraph** $\mathcal{H}_{\mathrm{Sch}} = (V_{\mathrm{Sch}}, E_{\mathrm{Sch}}, i_{\mathrm{Sch}})$ represents the relations between metadata.



Schema hypergraph: exploded view.
Shown on publication metadata example.

- **Extracted schema hypergraph** $\mathcal{H}_X$: only metadata instances of interest are kept in U, for:
  - visualisation;
  - reference;
  - search;
  - keeping connectivity in the hypergraph.

$\mathcal{H}_X = (V_X, E_X, i_X)$, with $V_X = U$, $E_X = \{e \cap U : e \in E_{\mathrm{Sch}}\}$ and $i_X = i_{\mathrm{Sch}}|_{E_X}$.

Reachability hypergraph (6)

- **Reachability hypergraph** $\mathcal{H}_R = (V_R, E_R, i_R)$: obtained from $\mathcal{H}_X$ by calculating its connected components.
- Vertices of $\mathcal{H}_R$: $V_R = V_X$.
- Hyperedges of $\mathcal{H}_R$: connected components of $\mathcal{H}_X$

$$\forall e_R \in E_R : i_R(e_R) = \bigcup_{E_{CC} \in C_X} \bigcup_{e \in E_{CC}} i_X(e).$$

Navigation hypergraph (7)

- **Navigation hypergraph** $\mathcal{H}_N = (V_N, E_N)$: obtained from $\mathcal{H}_R$ by choosing one hyperedge $e_R \in E_R$.
- Vertices of $\mathcal{H}_N$: $V_N = e_R$.

Possible reference vertices in R_{ref}.

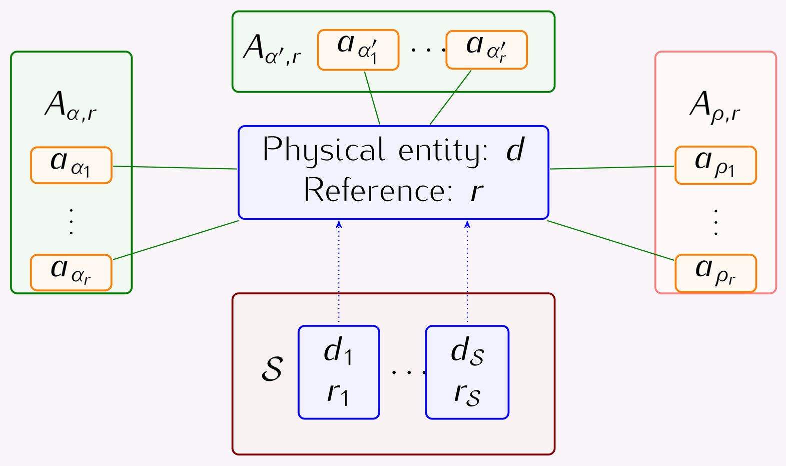
- Hyperedges of $\mathcal{H}_N$:

$$E_N = \{e_R \backslash R : R \subseteq R_{\mathrm{ref}} \wedge R \neq \emptyset\}.$$

- Navigation is possible without changing reference inside a hyperedge of $\mathcal{H}_N$.

Visualisation hypergraphs (8)

- In a dataset $\mathcal{D}$, a physical entity d of reference r is fully described by: $(r, \{A_{\alpha,r} : \alpha \in V_{\mathrm{Sch}}\})$.
- $A_{\alpha,r} = \{a_1, ..., a_{\alpha_r}\}$: set of values of type α that are attached to d.



- For each $v \in \bigcup_{r \in \mathcal{S}} A_{\rho,r} = \Sigma_\rho$, we build a set of physical references corresponding to data d that have v in attributes of type ρ: $R_v = \{r : v \in A_{\rho,r}\}$.
- Set of values of type α relatively to the reference v: $\bigcup_{r \in R_v} A_{\alpha,r} = e_{\alpha,v}$.
- **Raw visualisation hypergraph** for the facet of type α/ρ attached to the search $\mathcal{S}$ is:

$$\mathcal{H}_{\alpha/\rho,\mathcal{S}} = \left(\bigcup_{r \in \mathcal{S}} A_{\alpha,r}, \left(e_{\alpha,v}\right)_{v \in S_\rho}\right).$$

- By quotienting Σ_ρ and weighting => **reduced visualisation weighted hypergraph** for the search $\mathcal{S}$:

$$\mathcal{H}_{\alpha/\rho,w_\alpha,\mathcal{S}} = \left(\bigcup_{r \in \mathcal{S}} A_{\alpha,r}, \overline{E_\alpha}, w_\alpha\right).$$

Navigating through facets (9)

- Reference type: ρ, current type α, target type: α'.
- Selecting vertices of type $A \subseteq A_{\alpha,\mathcal{S}}$.

Allows to:

- retrieve a subset of hyperedges of $\overline{E_\alpha}$:

$$\overline{E_\alpha}\big|_A = \{e : e \in \overline{E_\alpha} \wedge (\exists x \in e : x \in A)\}.$$

- retrieve the class $\overline{v}$ attached to each $e \in \overline{E_\alpha}\big|_A$ => $\overline{V}\big|_A$ set of class $\overline{v}$.
- retrieve the references of type ρ: $\mathcal{V}_{\rho,A} = \{v : \forall \overline{v} \in \overline{V}\big|_A : v \in \overline{v}\}$.
- R_v remains the same between facets

=> group of references: $\mathcal{S}_A = \bigcup_{v \in \mathcal{V}_{\rho,A}} R_v$.

- switching to the facet of type α is then possible:

$$\mathcal{H}_{\alpha'/\rho}\big|_A = \left(\bigcup_{r \in \mathcal{S}_A} A_{\alpha',r}, \left(e_{\alpha',v}\right)_{v \in \mathcal{V}_{\rho,A}}\right).$$