Hypergraphs: an introduction and review

Xavier Ouvrard*
CERN,
Esplanade des Particules, 1,
CH-1211 Meyrin (Switzerland)

September 29, 2020

Abstract

Hypergraphs were introduced in 1973 by Berge. This review aims at giving some hints on the main results that we can find in the literature, both on the mathematical side and on their practical usage. Particularly, different definitions of hypergraphs are compared, some unpublished work on the visualisation of large hypergraphs done by the author. This review does not pretend to be exhaustive.

In 1736, Euler was the first to use a graph approach to solve the problem of the Seven Bridges of Königsberg. In 1878, Sylvester coined the word graph itself. Graphs are extensively used in various domains. A lot of developments on graphs have been realized during the first half of the twentieth century, but graphs are still an active field of research and applications, in mathematics, physics, computer science and various other fields, ranging from biology to economic and social network analysis.

In Berge [1967, 1973], the author introduced hypergraphs as a means to generalize the graph approach. Graphs only support pairwise relationships. Hypergraphs preserve the multi-adic relationships and, therefore, become a natural modeling of collaboration networks and various other situations. They already allow a huge step in modeling, but some limitations remain that will be discussed further in this article that has pushed us to introduce an extension of hypergraphs. In this Section, we aim at giving a synthesis of the hypergraphs. Further details on hypergraphs can be found either in Berge [1967, 1973], in Voloshin [2009] or in Bretto [2013] and in many of the references cited throughout this article.

1 Generalities

As given in Berge [1967, 1973], a **hypergraph** $\mathcal{H} = (V, E)$ on a finite **set of vertices** (or nodes) $V \triangleq \{v_i : i \in \llbracket n \rrbracket\}$[2] is defined as a family of hyperedges $E \triangleq (e_j)_{j \in \llbracket p \rrbracket}$ where each **hyperedge** is a non-empty subset of V and such that $\bigcup\limits_{j \in \llbracket p \rrbracket} e_j = V$.

It means that in a hypergraph, a hyperedge links one or more vertices. In Voloshin [2009], the definition of hypergraphs includes also hyperedges that are empty sets as hyperedges are defined as a family of subsets of a finite vertex set and it is not necessary that the union covers the vertex set. Both the vertex set and the family of hyperedges can be empty; it they are at the same time, the hypergraph is then designated as the **empty hypergraph**. This definition of hypergraph opens their use in various collaboration networks. It is the one we choose in this Thesis.

In Bretto [2013], an intermediate definition is taken, relaxing only the covering of the vertex set by the union of the hyperedges enabling only isolated vertices in hypergraphs.

Other interesting definitions of hypergraphs exist. Two of them are given in Stell [2012].

A hypergraph $\mathcal{H}$ is firstly defined as consisting of a set V of vertices and a set E of hyperedges with an **incidence function** $\iota : E \to \mathcal{P}(V)$, where $\mathcal{P}(V)$ is the poset of V. Also $\mathcal{H} = (V, E, \iota)$.

*xavier.ouvrard@cern.ch

[2]We write for $n \in \mathbb{N}^* : \llbracket n \rrbracket = \{i : i \in \mathbb{N}^* \wedge i \leqslant n\}$

The author emphasizes the fact that this definition allows several hyperedges to be incident to the same set of vertices and, also, to have hyperedges that are linked to the empty set of vertices. This definition is tightly linked to the one of Bretto [2013], with the advantage of having the incident function to link hyperedges and their corresponding set of vertices, instead of an equality.

A hypergraph is secondly defined in Stell [2012] as consisting of a set $\mathcal{H}$ containing both vertices and hyperedges and a **binary relation** φ such that:

1. $\forall x \in \mathcal{H}, \forall y \in \mathcal{H}, \quad (x, y) \in \varphi \implies (y, y) \in \varphi;$
2. $\forall x \in \mathcal{H}, \forall y \in \mathcal{H}, \forall z \in \mathcal{H}, \quad (x, y) \in \varphi \wedge (y, z) \in \varphi \implies y = z.$

The link can be made with the other definitions in considering the set V such that:

$$V \triangleq \{v \in \mathcal{H} : (v, v) \in \varphi\}$$

which corresponds to the set of vertices of the hypergraph $\mathcal{H}$ and the set E such that:

$$E \triangleq \{e \in \mathcal{H} : \ (e, e) \notin \varphi\}$$

which corresponds to the set of hyperedges of the hypergraph.

This last definition treats on the same footing vertices and hyperedges. Nonetheless, in this definition, the set aspect of hyperedges is implicit. Effectively, to know which vertex set a hyperedge e contains, the information needs to be rebuilt—using $\iota(e)$ as in the former definition—:

$$\iota(e) = \{v \in \mathcal{H} : \ (e, v) \in \varphi\} .$$

This definition leads to a representation of the hypergraph as a directed multi-graph where the vertices point to themselves and hyperedges are linked to vertices in this order.

Depending on the needs, either the first definition of Stell [2012] or equivalently the one given in Bretto [2013] without specifying the incidence relation will be used and referred as poset definition of hypergraphs—**PoDef** for short. Switching between these two definitions reverts to abusively identifying the hyperedge with the subset of vertices it identifies. The second definition of Stell [2012] will also be used—referred to as **MuRelDef**.

When using a PoDef defined hypergraph, $\mathcal{H} = (V, E, \iota)$ designates a hypergraph. If the MuRelDef is used, the hypergraph $\mathcal{H} = (U, \varphi)$ is used, with $U = V \cup E$, with $V = \{v \in \mathcal{H} : (v, v) \in \varphi\}$ the set of vertices and $E = \{e \in \mathcal{H} : \ (e, e) \notin \varphi\}$. In the following definitions, we put in perspective these two approaches.

2 Particular hypergraphs

When needed, elements of V will be written $(v_i)_{i \in [\![n]\!]}$ and those of E will be written as $(e_j)_{j \in [\![p]\!]}$. Abusively, e_j will also designate the subset $\iota(e_j)$ of V.

Two hyperedges $e_{j_1} \in E$ and $e_{j_2} \in E$, with $j_1, j_2 \in [\![p]\!]$ and $j_1 \neq j_2$ such that $e_{j_1} = e_{j_2}$ are said **repeated hyperedges**.

A hypergraph is said with **no repeated hyperedge**, if it has no repeated hyperedges.

Following Chazelle and Friedman [1988], where the hyperedge family is viewed as a multiset of hyperedges, a hypergraph with repeated hyperedges is called a **multi-hypergraph**.

A hypergraph is said **simple**, if for any two hyperedges $e_{j_1} \in E$ and $e_{j_2} \in E$:

$$e_{j_1} \subseteq e_{j_2} \Rightarrow j_1 = j_2.$$

Hence, a simple hypergraph has no repeated hyperedges.

A hyperedge $e \in E$ such that: $|e| = 1$ is called a **loop**.

A hypergraph is said **linear** if it is simple and such that every pair of hyperedges shares at most one vertex.

A **sub-hypergraph** $\mathcal{K}$ of a hypergraph $\mathcal{H}$ is the hypergraph formed by a subset W of the vertices of $\mathcal{H}$ and the hyperedges of $\mathcal{H}$ that are subsets of W.

Formally, with

- PoDef: $\mathcal{K} = (W, F, \iota|_F)$, such that: $F = \{e_j \in E : \iota|_F(e_j) \subseteq W\}$;
- MuRelDef: $\mathcal{K} = (T, \varphi)$, with $T \subseteq U$, where $\forall t \in T$, $\forall u \in U : (t,u) \in \varphi \implies u \in T$ and $\forall t \in T : (t,t) \in \varphi \implies t \in W$.

A **partial hypergraph** $\mathcal{H}'$ generated by a subset $E' \subseteq E$ of the hyperedges is a hypergraph containing exactly these hyperedges and whose vertex set contains at least all the vertices incident to this hyperedge subset.

Formally, with:

- PoDef: $\mathcal{H}' \triangleq (V', E', \iota|_{E'})$, where $E' = \{e_j : j \in K'\} \subseteq E$ with $K' \subseteq [\![p]\!]$ and V' is such that $\bigcup\limits_{k \in K'} \iota|_{E'}(e_k) \subseteq V'$;
- MuRelDef: $\mathcal{H}' \triangleq (T', \varphi)$, with $T' \subseteq U$, where $\forall t \in U, \forall t' \in T' : (t,t') \in \varphi \implies t \in T'$ and $\forall t \in T' : (t,t) \notin \varphi \implies t \in E'$.

3 Duality

The **star of a vertex** $v \in V$ of a hypergraph $\mathcal{H}$, written $H(v)$, is the family of hyperedges containing v.

Also with:

- PoDef: $H(v) \triangleq \{e_k \in E : v \in \iota(e_k)\}$;
- MuRelDef: $H(v) \triangleq \{t \in U : (t,v) \in \varphi \wedge (t,t) \notin \varphi\}$.

The **dual** of a hypergraph $\mathcal{H}$ is the hypergraph $\mathcal{H}^*$ whose vertices corresponds to the hyperedges of $\mathcal{H}$ and the hyperedges of $\mathcal{H}^*$ are the vertices of $\mathcal{H}$, with the incident relation that links each vertex to its star.

Formally with:

- PoDef: The dual is given by $\mathcal{H}^* = (V^*, E^*, \iota^*)$, with $V^* = E$ and where $E^* = \{v \in V\}$ and $\forall v \in V : \iota^*(v) = H(v)$;
- MuRelDef: The dual is given by: (U, φ^*) where $\varphi^* \triangleq \neg\varphi \cap (1' \cup \smile \varphi)$, with:
- $\neg\varphi$ the complement of the relation $\varphi : \forall x \in U, \forall y \in U : \quad (x,y) \in \neg\varphi \Leftrightarrow (x,y) \notin \varphi$;
- 1' the binary identity: $\forall x \in U, \forall y \in U : \quad (x,y) \in 1' \Leftrightarrow x = y$;
- $\smile \varphi$ the converse relation: $\forall x \in U, \forall y \in U : \quad (x,y) \in \smile \varphi \Leftrightarrow (y,x) \in \varphi$.

Proof for (MuRelDef). (not given in Stell [2012]) Let $x \in U$ such that $(x,x) \in \varphi^*$. Then $(x,x) \in \neg\varphi$ and x is a hyperedge of (U,φ).

If $x \in U$ such that $(x,x) \notin \varphi^*$, then $(x,x) \notin \neg\varphi$ and x is a vertex of (U,φ) as $(x,x) \in 1'$ and, therefore: $(x,x) \in 1' \cup \smile \varphi$.

Let $(x,y) \in \varphi^*$ and suppose that $x \neq y$. Then as $\varphi^* = \neg\varphi \cap (1' \cup \smile \varphi)$, we have: $(x,y) \in 1' \cup \smile \varphi$. As $x \neq y$, $(x,y) \notin 1'$, so we have $(x,y) \in \smile \varphi$, i.e. $(y,x) \in \varphi$. It implies that $(x,x) \in \varphi$ and thus a vertex of (U,φ). As y is a vertex of (U,φ^*), it is a hyperedge of (U,φ).

$\square$

Some interesting properties can be derived from the MuRelDef and might be useful for computation.

Proposition 1. *Let (U,φ) be a hypergraph.*

The set of vertices is given by the binary relation: $\nu \triangleq \varphi \cap 1'$.

The set of hyperedges is given by the binary relation: $\eta \triangleq \varphi \cap (\neg\nu)$.

Proof. Let $x \in U$: $(x,x) \in \varphi \Leftrightarrow (x,x) \in \varphi \cap 1'$, so ν gives exactly the vertices of (U,φ).

Moreover: $(x,x) \in \varphi \Leftrightarrow (x,x) \notin \neg\nu$, so $\eta = \varphi \cap (\neg\nu)$ does not hold the vertices of (U,φ).

Let us take $x, y \in U$ with $x \neq y$. We have $(x,y) \in \varphi \implies (x,y) \notin \varphi \cap 1' \Leftrightarrow (x,y) \in \neg(\varphi \cap 1')$. Therefore: $(x,y) \in \varphi \implies (x,y) \in \varphi \cap (\neg\nu)$.

Reciprocally, if $x \neq y : (x,y) \in \varphi \cap (\neg\nu) \implies (x,y) \in \varphi$, then η gives exactly the hyperedges' content of (U,φ). □

Proposition 2. *Using the dual relation:*

The set of hyperedges can be retrieved by the binary relation: $\nu^* \triangleq \varphi^* \cap 1'$.

The set of incident hyperedges can be retrieved by the binary relation: $\eta^* \triangleq \varphi^* \cap (\neg\nu^*)$.

Proof. Immediate with duality. □

Proposition 3. *Let (U,φ) be a hypergraph and φ^* its dual relation.*

(i) The binary relation $\alpha = (\varphi \circ \varphi^*) \cap (\neg\eta)$ allows to retrieve the adjacency of vertices.

(ii) The binary relation $\beta = (\varphi^* \circ \varphi) \cap (\neg\eta^*)$ allows to retrieve the intersection of hyperedges.

Proof for (i). In (U,φ^*), when we select an element of $x \in U$, either it is a vertex of (U,φ^*), and then $(x,x) \in \varphi^*$, or it is a hyperedge and then $(x,x) \notin \varphi^*$.

Let $x \in (U,\varphi^*)$ be such that $(x,x) \notin \varphi^*$. x is a hyperedge of (U,φ^*) and, therefore, a vertex of (U,φ). If there exists $e \in U$, such that $(x,e) \in \varphi^*$, then e is a vertex of (U,φ^*), also a hyperedge of (U,φ). If this hyperedge is non empty in (U,φ), there exists $x' \in (U,\varphi)$, such that $(e,x') \in \varphi$, and, necessarily by definition, $(x',x') \in \varphi$. (x,x') links two vertices of (U,φ). Either $x = x'$ which implies: $(x,x') \in \neg\eta$ or $x \neq x'$ which implies $(x,x') \notin \varphi$, so $(x,x') \in \neg\eta$. Therefore, $(x,x') \in (\varphi \circ \varphi^*) \cap (\neg\eta)$ links two vertices and points the adjacency of these two vertices.

Let $x \in (U,\varphi^*)$ be such that $(x,x) \in \varphi^*$. x is a vertex of (U,φ^*) and, therefore, a hyperedge of (U,φ), i.e. $(x,x) \notin \varphi$. If this hyperedge is non empty in (U,φ), there exists $x' \in (U,\varphi)$, such that $(x,x') \in \varphi$, and necessarily, by definition, $(x',x') \in \varphi$, and x' is a vertex of (U,φ). Therefore, $(x,x') \in (\varphi \circ \varphi^*) \cap \eta$. □

Proof for (ii). This proof is similar to the former one. □

4 Neighborhood of a vertex

The neighborhood of a vertex in a hypergraph is defined similarly to what is done for graphs.

The **neighborhood of a vertex** $v \in V$ is the set $\Gamma(v)$ of vertices that belongs to the hyperedges this vertex is belonging.

That is with:

- PoDef: $\Gamma(v) = \bigcup\limits_{e \in H(v)} \iota(e)$;
- MuRelDef: For v such that $(v,v) \in \varphi$:

  $$\Gamma(v) = \{s \in U : (s,s) \in \varphi \wedge (\exists w \in U : (w,s) \in \varphi \wedge (w,v) \in \varphi \wedge (w,w) \notin \varphi)\},$$

  which is equivalent to:

  $$\Gamma(v) = \{s \in U : (s,s) \in \varphi \wedge (s,v) \in \varphi \circ \varphi^*\}.$$

Proof for (MuRelDef). As $(v,v) \in \varphi$, $(v,v) \notin \varphi^*$. If there exists $t \in U$ such that $(t,v) \in \varphi^*$, t is a hyperedge of (U,φ). Therefore, any s, if there exists, such that $(s,t) \in \varphi$ is a vertex and is in the same hyperedge than v, so in its neighborhood. The other inclusion is straightforward. □

5 Weighted hypergraph

In Zhou et al. [2007], the definition of a weighted hypergraph is given, based on the definition of Berge [1967, 1973] of a hypergraph.

$\mathcal{H}_{w_e} = (V, E, w_e)$ is a **weighted hypergraph**, if (V,E) is a hypergraph and $w_e : E \to \mathbb{R}$ is a function that associates to each hyperedge $e \in E$ a weight $w_e(e)$.

We can refine this definition to handle weights on individual vertices, by using a second function $w_v : V \rightarrow \mathbb{R}$ that associates to each vertex $v \in V$ a weight $w_v(v)$. But putting weights that are hyperedge dependent cannot be achieved with hypergraphs as it would imply to move to a new algebra, as we will see with the introduction of hb-graphs.

6 Hypergraph features

Hypergraph features are very similar to those of graphs with some arrangements to account for their differences in structure.

The **order** $o_{\mathcal{H}}$ of a hypergraph $\mathcal{H}$ is defined as $o_{\mathcal{H}} \triangleq |V|$.

The **rank** $r_{\mathcal{H}}$ of a hypergraph $\mathcal{H}$ is the maximum of the cardinalities of the hyperedges:

$$r_{\mathcal{H}} \triangleq \max_{e \in E} |e|,$$

while the **anti-rank** $s_{\mathcal{H}}$ corresponds to the minimum:

$$s_{\mathcal{H}} \triangleq \min_{e \in E} |e|.$$

The **degree of a vertex** $v_i \in V$, written **deg** $(v_i) = d_i$, corresponds to the number of hyperedges that this vertex participates in. Hence:

$$\deg(v_i) \triangleq |H(v_i)|.$$

It is also designated as **hyper-degree** in some articles.

7 Paths and related notions

A **path** $v_{i_0} e_{j_1} v_{i_1} \ldots e_{j_s} v_{i_s}$ in a hypergraph $\mathcal{H}$ from a vertex u to a vertex v is a vertex / hyperedge alternation with s hyperedges e_{j_k} such that: $\forall k \in [\![s]\!], j_k \in [\![p]\!]$ and $s+1$ vertices v_{i_k} with $\forall k \in \{0\} \cup [\![s]\!], i_k \in [\![n]\!]$ and such that $v_{i_0} = u$, $v_{i_s} = v$, $u \in e_{j_1}$ and $v \in e_{j_s}$ and that for all $k \in [\![s-1]\!]$, $v_{i_k} \in e_{j_k} \cap e_{j_{k+1}}$.

The **length of a path** from u to v is the number of hyperedges it traverses; given a path $\mathscr{P}$, we write $l(\mathscr{P})$ its length. It holds that if $\mathscr{P} = v_{i_0} e_{j_1} v_{i_1} \ldots e_{j_s} v_{i_s}$, we have: $l(\mathscr{P}) = s$.

The **hypergraph distance** $d(u, v)$ between two vertices u and v of a hypergraph is the length of the shortest path between u and v, if there exists, that can be found in the hypergraph. In the case where there is no path between the two vertices, they are said to be **disconnected**, and we set: $d(u, v) = +\infty$. A hypergraph is said to be **connected** if there exists a path between every pair of vertices of the hypergraph, and **disconnected** otherwise.

A **connected component** of a hypergraph is a maximal subset of the vertex set for which there exists a path between any two vertices of this maximal subset in the hypergraph.

8 Multi-graph, graph, 2-section

A hypergraph with rank at most 2 is called a **multi-graph**. A simple multi-graph without loop is a **graph**.

For the moment, we keep the original concept of adjacency as it is implicitly given in Bretto [2013]; we mention it here as **2-adjacency** since it is a pairwise adjacency.

Two distinct vertices $v_{i_1} \in V$ and $v_{i_2} \in V$ are said **2-adjacent** if there exists $e \in E$ such that $v_{i_1} \in e$ and $v_{i_2} \in e$.

The graph $[\mathcal{H}]_2 \triangleq \left(V_{[2]}, E_{[2]}\right)$ obtained from a hypergraph $\mathcal{H} = (V, E)$ by considering: $V_{[2]} \triangleq V$ and such that if v_{i_1} and v_{i_2} are 2-adjacent in $\mathcal{H}$, $\{v_{i_1}, v_{i_2}\} \in E_{[2]}$ is called the **2-section of the hypergraph** $\mathcal{H}$.

The graph $[\mathcal{H}]_{\mathcal{I}} \triangleq (V_{\mathcal{I}}, E_{\mathcal{I}})$ obtained from a hypergraph $\mathcal{H} = (V, E)$ by considering: $V_{\mathcal{I}} \triangleq V^*$ and, such that, if $e_{j_1} \in E$ and $e_{j_2} \in E$—with $j_1 \neq j_2$—are intersecting hyperedges in $\mathcal{H}$, then $\{e_{j_1}, e_{j_2}\} \in E_{\mathcal{I}}$, is called the **intersection graph** of the hypergraph $\mathcal{H}$.

Let $k \in \mathbb{N}^*$. A hypergraph is said to be ***k*-uniform** if all its hyperedges have the same cardinality k.

A **directed hypergraph** $\mathcal{H} = (V, E)$ on a finite set of n vertices (or vertices) $V = \{v_i : i \in [\![n]\!]\}$ is defined as a family of p **hyperedges** $E = (e_j)_{j \in [\![p]\!]}$ where each hyperedge e_j contains exactly two non-empty subsets of V, one which is called the **source**—written $e_{s\,j}$—and the other one which is the **target**—written $e_{t\,j}$. A hypergraph that is not directed is said to be an **undirected hypergraph**.

9 Sum of hypergraphs

Let $\mathcal{H}_1 = (V_1, E_1)$ and $\mathcal{H}_2 = (V_2, E_2)$ be two hypergraphs. The **sum** of these two hypergraphs is the hypergraph written $\mathcal{H}_1 + \mathcal{H}_2$ defined as:

$$\mathcal{H}_1 + \mathcal{H}_2 \triangleq (V_1 \cup V_2, E_1 \cup E_2)\,.$$

This sum is said **direct** if $E_1 \cap E_2 = \emptyset$. In this case, the sum is written $\mathcal{H}_1 \oplus \mathcal{H}_2$.

10 Matrix definitions of hypergraphs

10.0.1 Incidence matrix

The **incidence matrix**:

$$H \triangleq [h_{ij}]_{\substack{i \in [\![n]\!] \\ j \in [\![p]\!]}}$$

of a hypergraph is the matrix having rows indexed by the corresponding indices of vertices of $\mathcal{H}$ and columns by the hyperedge indices, and where the coefficient $h_{ij} \triangleq 1$ when $v_i \in e_j$, and $h_{ij} \triangleq 0$ when $v_i \notin e_j$.

10.0.2 Adjacency matrix

We focus in this paragraph on the pairwise adjacency as defined in Berge [1967, 1973] and Bretto [2013]. In the latter, the **adjacency matrix of a hypergraph** $\mathcal{H}$ is defined as the square matrix $A \triangleq [a_{ij}]$ having rows and columns indexed by indices corresponding to the indices of vertices of $\mathcal{H}$ and where for all $i, j \in [\![p]\!]\,, i \neq j$:

$$a_{ij} \triangleq |\{e \in E \,:\, v_i \in e \wedge v_j \in e\}|$$

and for all $i \in [\![p]\!]$:

$$a_{ii} \triangleq 0.$$

It holds, following Estrada and Rodriguez-Velazquez [2005]:

$$A = HH^{\intercal} - D_V$$

where $D_V \triangleq \text{diag}\left((d_i)_{i \in [\![n]\!]}\right)$ is the diagonal matrix containing vertex degrees.

The **adjacency matrix of a weighted hypergraph** $\mathcal{H}_w$ is defined in Zhou et al. [2007] as the matrix A_w of size $n \times n$ defined as:

$$A_w \triangleq HWH^{\intercal} - D_w$$

where:

$$W \triangleq \text{diag}\left((w_j)_{j \in [\![p]\!]}\right)$$

is a diagonal matrix of size $p \times p$ containing the weights $(w_j)_{j \in [\![p]\!]}$ of the respective hyperedges $(e_j)_{j \in [\![p]\!]}$ and D_w is the diagonal matrix of size $n \times n$:

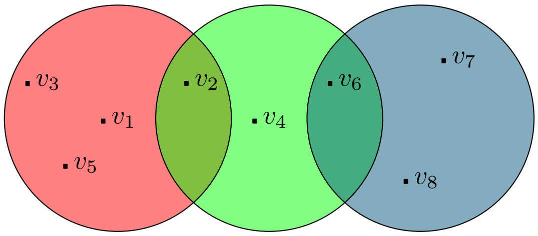
$$D_w \triangleq \text{diag}\left((d_w(v_i))_{i \in [\![n]\!]}\right)$$

and where for all $i \in [\![n]\!]$:

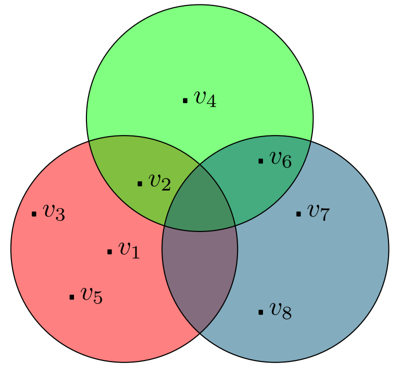
$$d_w(v_i) \triangleq \sum_{j \in \{k : k \in [\![p]\!] \wedge v_i \in e_k\}} w_j$$

is the **weighted degree** of the vertex $v_i \in V$.

We will refine the concept of adjacency in general hypergraphs in the next Chapter.



(a) Euler diagram.



(b) Venn diagram.

Figure 1: Difference between (a) Euler diagram and (b) Venn diagram.

10.0.3 Additional features

In Estrada and Rodriguez-Velazquez [2005], the authors introduce particular features to characterize hypergraphs. They evaluate the **centrality of a vertex** in a simple unweighted hypergraph, by orthogonalizing the adjacency matrix in:

$$A = U\Lambda U^\intercal,$$

where $U \triangleq (u_{ij})_{(i,j)\in[\![n]\!]^2}$ and $\Lambda \triangleq \text{diag}\left((\lambda_i)_{i\in[\![n]\!]}\right)$ is the diagonal matrix formed of the eigenvalues λ_i of A with $i \in [\![n]\!]$.

The **sub-hypergraph centrality** $C_{\text{SH}}(i)$ is defined as the sum of the closed walks of different lengths in the network, starting and ending at a given vertex. For a simple hypergraph, it can be calculated using:

$$C_{\text{SH}}(i) \triangleq \sum_{j=1}^{n} (u_{ij})^2 \, \mathrm{e}^{\lambda_j}.$$

A **clustering coefficient** for hypergraph is also defined as:

$$C(\mathcal{H}) \triangleq \frac{6 \times \text{number of hyper-triangles}}{\text{number of 2-paths}}$$

where a hyper-triangle is defined as a sequence of three different vertices that are separated by three different hyperedges $v_i e_p v_j e_q v_k e_r v_i$ and a 2-path is a sequence $v_i e_p v_j e_q v_k$.

11 Hypergraph visualisation

In Mäkinen [1990], hypergraph visualizations are classified in two categories called the "subset standard" and the "edge standard". These two types of representations reflect the two facets of hypergraphs. The **subset standard** reflects that hyperedges are subsets of the vertex set: the vertices of a hyperedge are drawn as points and hyperedges as closed envelopes. The **edge standard** reflects that vertices of a hyperedge maintain together a multi-adic relationship. We then cover the Zykov representation which can be seen as a bridge between the two representations.

11.1 The subset standard

Euler diagram represents sets by simple closed shapes in a two dimensional plane. Shapes can overlap to represent relationships within the sets; these relationships can be either fully inclusive, partially inclusive or disjoint.

In Venn [1881], the author introduces a particular case of Euler diagram that shows the 2^n logical relations between n sets with the aim to formalize logical propositions by diagrams, now called **Venn diagrams**.

In Mäkinen [1990], the author presents the Venn diagram representation of hypergraphs: a Venn representation of a hypergraph represents hyperedges as closed curves that contains points symbolizing the vertices. In Johnson and Pollak [1987], the authors present two representations with the subset standard. The first representation is a hyperedge-based Venn diagram based on the representation of the hypergraph as a planar graph, an embedding of this planar graph and a one-to-one map between the hyperedges and the embedding faces, such that taking any subset of the vertices, the union of the faces corresponding to this subset includes a region of the plane with a connected interior. Hypergraphs that can be represented in such a way are said hyperedge-planar. The second representation is a vertex-based Venn diagram where the mapping is from the set of vertices to the embedding faces such that for every hyperedge the interior of the face union that contains all the hyperedge vertices is connected, with a corresponding notion of vertex-planar when such a representation is feasible.

In Bertault and Eades [2001], a drawing system is presented, named PATATE, where hypergraphs are represented using the subset standard: a graph is built combining three possibilities. The first representation is achieved by building the incidence graph of the hypergraph, adding a dummy vertex for each hyperedge that is linked by an edge, placed at the barycenter. The second representation is built using a minimal Euclidean spanning-tree that covers the vertices of each hyperedge: the edges of the spanning-tree are added to the graph. The third representation is built using a minimal Steiner tree: a vertex is added for every Steiner point and the edges of the tree are added. A force directed algorithm is applied to the built graph and the convex hull of each hyperedge is then computed.

The Venn diagram representation is also addressed in Estrada and Rodriguez-Velazquez [2005]. The Venn diagram is relevant for small hypergraphs but is harder to use for large hypergraphs.

A large survey on set representations is available in Alsallakh et al. [2016].

11.2 The edge standard

The edge standard includes two main representations: the **clique expansion** and the **extra-node representation** also called the incidence representation in Kaufmann et al. [2009].

The clique expansion is based on the 2-section of the hypergraph where a hyperedge of cardinality n is represented by $\dfrac{n(n-1)}{2}$ undirected edges.

The extra vertex representation corresponds to the incidence graph of the hypergraph, which is a bipartite graph. It is called the König representation of the hypergraph in Zykov [1974]. In this case, a hyperedge of size n is represented with only n edges connecting hyperedge vertices to a central vertex called **extra-node**.

Figure 2 illustrates these two representations.

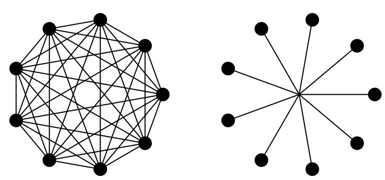


Figure 2: Clique expansion vs extra-node representation of a hyperedge.

The clique expansion generates graph with more edges than the extra-node representation for values of $n \geqslant 4$. Moreover, if a hyperedge e_{j_1} is a strict subset of an other hyperedge e_{j_2}, the clique expansion of e_{j_1} does not add any other edge, while the extra-node representation brings $|e_{j_1}|$ new edges and an additional vertex. An example of an unfavorable case is given in Figure 3.

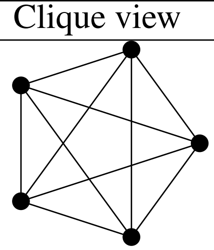
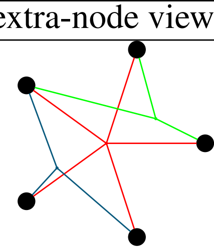
Clique view	extra-node view
	
In this case: 10 edges, 5 vertices	In this case: 11 edges, 5 vertices and 3 extra-nodes

Figure 3: Unfavorable case for the extra-node view.

In the clique representation, vertices of hyperedges are seen as interacting with 2-adic relationships and the information on the meso-structure is lost Taramasco et al. [2010]. The extra-node representation allows to keep the multi-adic relationships. Moreover, we have shown in Ouvrard et al. [2017] that the number of edges can be substantially reduced by the extra-node representation for a power-law hypergraph.

Junghans [2008] focuses on hyperedge drawing so that they do not intersect cluster groups, giving a solution based on force attraction/repulsion.

Other representations in the node standard exist: in Paquette and Tokuyasu [2011], the authors propose a representation using a pie-chart node approach, which is valuable for hypergraphs where the hyperedges do not intersect one another with too high cardinality. In Kerren and Jusufi [2013], a radial approach is presented that fits for small hypergraphs. In Kaufmann et al. [2009], a Steiner tree representation of the hyperedges is considered as the edge representation of a hypergraph.

11.3 The Zykov representation

In Zykov [1974], the author tackles the representation of hypergraphs and proposes a notion of hypergraph planarity. A planar hypergraph is a hypergraph where the vertices are represented as points in a plane and hyperedges as closed curves not intersecting one another but in the neighborhood of a common incident vertex of hyperedges.

This notion of Zykov-planarity corresponds to the planarity of the incidence graph of the hypergraph.

The author proposes a representation, called now the Zykov representation and shown in Figure 4 (c), that is based on a continuous deformation of the edges, starting with the Euler diagram in Figure 4 (a) via the PaintSplash representation in Figure 4 (b) and that stops just before obtaining the incidence graph as shown in Figure 4 (d). Hyperedges are represented as faces of a subdivision realized by the vertices belonging to the corresponding hyperedge. Two hyperedges with a common vertex are incident at that vertex. The representation in the original article is in black and white; hence, this representation requires two colors: one for the background and an other one for the faces—we keep here the colors for helping to the understanding of the continuous deformation of the hyperedges from the Euler representation to the extra-node representation. This representation is intensively used with simplicial complexes which are particular case of hypergraphs.

The Zykov representation cannot be totally put neither in the edge standard nor in the subset standard.

11.4 PaintSplash representation

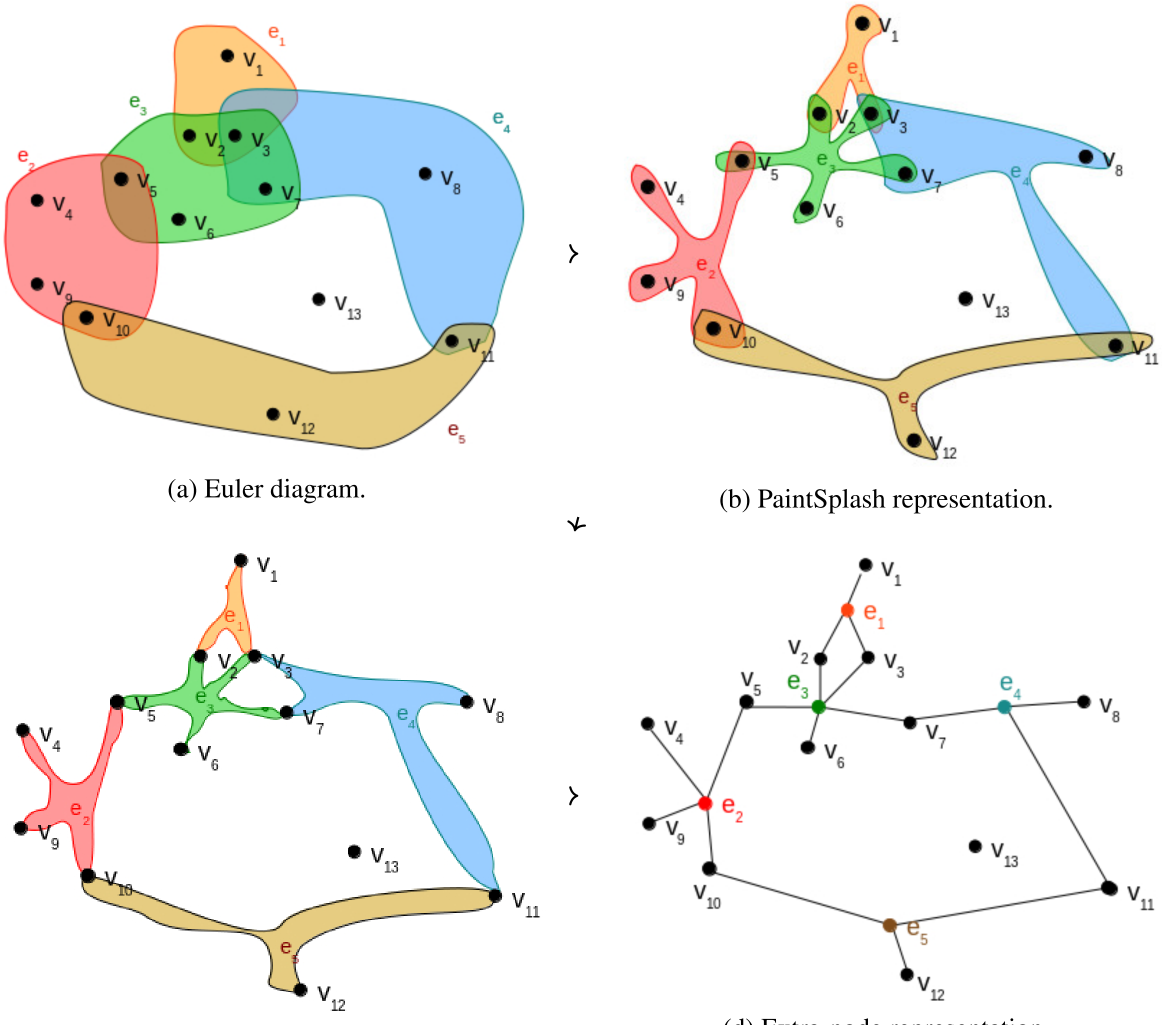
It is worth mentioning that Figure 4 (b) can lead to a good compromise between the subset standard and the edge standard, that we have called in its modern colored version the PaintSplash representation of a hypergraph, presented in Figure 5.

11.5 Large hypergraph visualisation

Large hypergraphs require strategies to be properly visualized. We have shown in Ouvrard et al. [2017] that switching from the 2-section representation to the incidence representation reduces considerably the overall cognitive load, as the number of edges to be represented is effectively highly diminished in real networks; moreover, true collaborations are seen more distinguishably. An example is given in Figure 6, that shows how the representation in the two modes impact on the cognitive load.

Other strategies have been implemented in order to improve the global layout and information retrieval of large hypergraphs. It includes the segmentation of the hypergraph into connected components, ordering them by the number of references they represent, and then with respect to the number of vertices within co-occurrences they contain. The layout of each connected component is computed separately, by considering the communities based on the Louvain algorithm presented in Blondel et al. [2008] and using a force directed algorithm, named ForceAtlas2, presented in Jacomy et al. [2014], to ensure the layout of the communities and of the vertices inside the communities. To achieve a proper layout, vertices of importance for the community are placed first and the remaining vertices are then placed around those fixed important vertices, with an optimization of the layout. The principle of the calculus is given in Figure 9 and the results is shown in Figure 11.5 and Figure 8.

Optimizing the layout of a large hypergraph is a challenging task. It requires to capture important vertices of the hypergraph, i.e. the ones we do not want to loose by overwhelming or overlapping in the representation. This requirement has been the starting point of our quest for a diffusion tensor.



(a) Euler diagram.

(b) PaintSplash representation.

(c) Zykov representation.

(d) Extra-node representation.

Figure 4: Moving from the Euler diagram (a) via the PaintSplash representation (b) and the Zykov representation (c) to the incident representation (d) based on Zykov [1974]—in the order of the arrows. The original figure is in black and white.

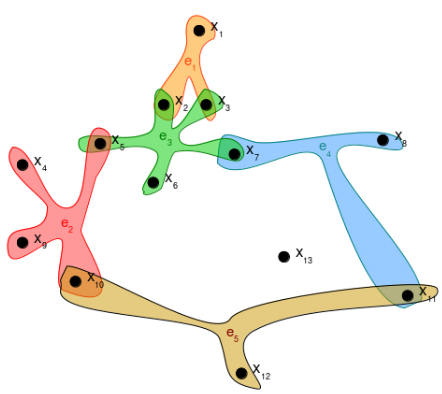
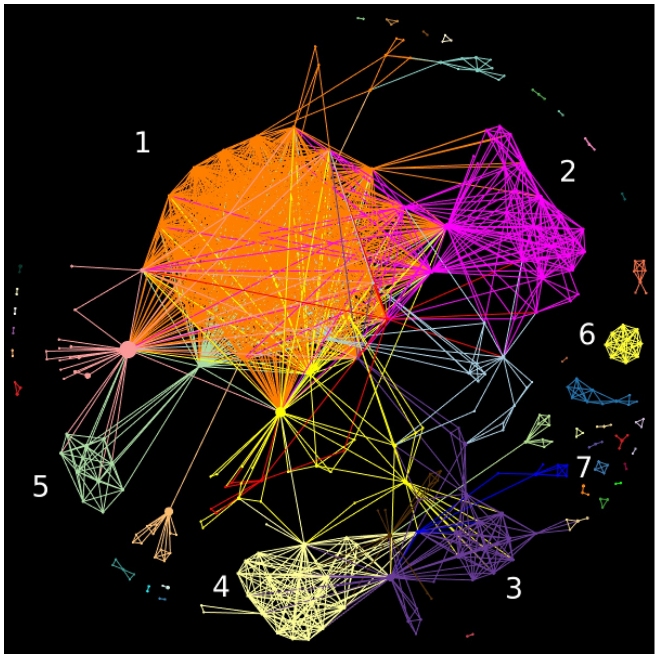
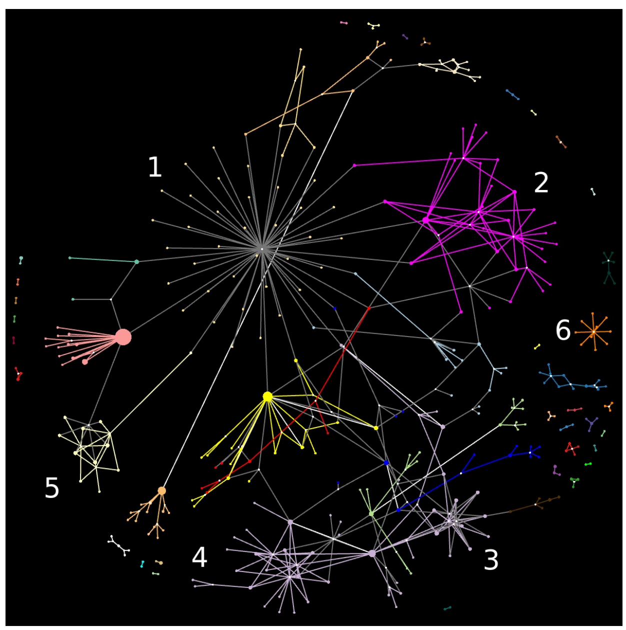


Figure 5: PaintSplash representation of a hypergraph.



Sub-figure 6 (a): Clique representation: The coordinates of the nodes are calculated by ForceAtlas2 on the extra-node view and then transferred to this view.



Sub-figure 6 (b): Extra-node representation: The coordinates of the nodes are calculated by ForceAtlas2 for this representation.

Figure 6: Hypergraph of organizations: Sub-figures (a) and (b) refer to the search:
title:((bgo AND cryst*) OR (bgo AND calor*)) abstract:((bgo AND cryst*) OR (bgo AND calor*)) from Ouvrard et al. [2017].

12 Morphisms, category and hypergraphs

12.1 A parenthesis on categories

The notions presented in this paragraph are taken from different sources including the page on Category[3] and the reference book Adámek et al. [2004].

A **category** $\mathcal{C}$ is formed by:

- a class of **objects** denoted Ob $(\mathcal{C})$
- a class of elements called **arrows** or **morphisms** mor $(\mathcal{C})$

and such that:

- given a pair (X, Y) , with $X \in$ Ob $(\mathcal{C})$ and $Y \in$ Ob $(\mathcal{C})$, there exists a set: Hom$(X, Y) \in$ mor $(\mathcal{C})$ called the morphisms from X to Y in $\mathcal{C}$. If f is such an homomorphism, we write it $f : X \rightarrow Y$. X is called the source and Y the target.
- for every $X \in$ Ob $(\mathcal{C})$, there exists a morphism written id$_X \in$ Hom (X, X) called the identity on X.
- a composition law $\circ$ such that for two morphisms $f : A \rightarrow B$ and $g : B \rightarrow C$ are associated to a morphism $g \circ f : A \rightarrow C$ called composition of f and g, such that:
- the composition is an associative law: for any $f \in$ Hom (A, B), $g\in$ Hom (B, C) and $h \in$ Hom (C, D) :

$$(h \circ g) \circ f = h \circ (g \circ f)$$

- the identity are the neutral elements of the composition: given any morphism $f \in$ Hom (A, B) :

$$f \circ \mathrm{Id}_A = f = \mathrm{Id}_B \circ f$$

For instance, the category where the objects are sets and where morphisms are applications between sets with the natural composition of applications is the Set category, written Set.

As summarized in Isah and Tella [2015], different kinds of morphisms exist. We consider a morphism $f : A \rightarrow B$ of a category $\mathcal{C}$ where A and B are two objects of $\mathcal{C}$. f is:

- a **monomorphism** if for all pairs of morphisms of $\mathcal{C}$ with $h : C \rightarrow A$ and $k : C \rightarrow A$, $f \circ h = f \circ k$ implies $h = k$ i.e. if f is left-cancellable.
- a **split monomorphism** (or **section**) if it is left-invertible i.e. there exists a morphism $g : B \rightarrow A$ of $\mathcal{C}$ such that: $g \circ f = \mathrm{Id}_A$.
- an **epimorphism** if it is right-cancellable i.e. for all pairs $h : B \rightarrow C$ and $k : B \rightarrow C$, $h \circ f = k \circ f$ implies $h = k$.
- a **split epimorphism** (or **retraction**) if it is right-invertible i.e. there exists $g : B \rightarrow A$ such that $f \circ g = \mathrm{Id}_B$.
- a **bimorphism** if it is both a monomorphism and an epimorphism.
- an **isomorphism** if it is both a split monomorphism and a split epimorphism.

A **sub-category** $\mathcal{S}$ of a category $\mathcal{C}$ is a category where the objects are objects of $\mathcal{C}$ and where the morphisms are morphism of $\mathcal{C}$ between objects of $\mathcal{S}$.

When a sub-category $\mathcal{S}$ of a category $\mathcal{C}$ has its morphisms that are exactly all the morphisms of $\mathcal{C}$ between objects of $\mathcal{S}$, this sub-category is said **full**.

Correspondences between two categories are possible by associating to each object (resp. a morphism) of the first category, an object (resp. a morphism) of the second category: such correspondences are called functors.

Let $\mathcal{C}$ and $\mathcal{D}$ be two categories. A **functor** from $\mathcal{C}$ to $\mathcal{D}$ (or covariant functor), written $F : \mathcal{C} \rightarrow \mathcal{D}$, is defined by giving:

- a function such that for all $X \in$ Obj $(\mathcal{C})$, there exists an object $F (X) \in$ Obj $(\mathcal{D})$;
- a function such that for all $f \in$ mor $(\mathcal{C})$, with $f : X \rightarrow Y$, there exists $F (f) \in$ mor $(\mathcal{D})$ such that: $F(f) : F (X) \rightarrow F (Y)$;

[3] https://en.wikipedia.org/wiki/Category_(mathematics)

and, such that:

- for all $X \in \mathcal{C}$, we have (identity conservation):

$$F\left(\mathrm{id}_X\right) = \mathrm{id}_{F(X)};$$

- for all $X, Y, Z \in \mathcal{C}$ and $f \in \mathrm{Hom}\left(X, Y\right)$ and $g \in \mathrm{Hom}\left(Y, Z\right)$, it holds (composition conservation):

$$F\left(g \circ f\right) = F\left(g\right) \circ F\left(f\right).$$

The construction of new categories can be achieved by using the **dual of a category** $\mathcal{C}$, written $\mathcal{C}^{\mathrm{op}}$, where the objects are the ones of $\mathcal{C}$ and the arrows of $\mathcal{C}^{\mathrm{op}}$ are the ones of $\mathcal{C}$ taken reversely. An other usual way is to consider the product category of two categories $\mathcal{C}$ and $\mathcal{C}'$, denoted $\mathcal{C} \times \mathcal{C}'$, where the objects are pairs of objects of $\mathcal{C}$ and $\mathcal{C}'$ taken in this order and the arrows are constituted of pairs of arrows of $\mathcal{C}$ and $\mathcal{C}'$ taken in this order.

12.2 Hypergraph homomorphism and category

In Dörfler and Waller [1980], the authors use the category theory to consider hypergraph product. In this article, a hypergraph (V, E, f) is defined as a triple with V the vertex set, E the hyperedge set and $f : E \to \mathcal{P}\left(V\right) \setminus \{\emptyset\}$ a function that assigns to each hyperedge its non-empty set of vertices. Duplicated hyperedges are allowed, i.e. two hyperedges can have the same image by f.

The following definitions and properties are taken from Dörfler and Waller [1980].

The **covariant power set functor** $\mathcal{P} : \mathrm{Set} \to \mathrm{Set}$ is defined by:

- for $A \in \mathrm{Obj}\left(\mathrm{Set}\right)$, $\mathcal{P}\left(A\right)$ is the set of all subsets of A;
- for $h \in \mathrm{mor(Set)}$, $h : A \to B$, $\mathcal{P}\left(h\right)$ is the homomorphism: $\mathcal{P}\left(A\right) \to \mathcal{P}\left(B\right)$ such that for all $A' \subset A$, $\mathcal{P}\left(h\right) : A' \to h\left(A'\right)$.

Considering two hypergraphs $\mathcal{H}_1 = \left(V_1, E_1, f_1\right)$ and $\mathcal{H}_2 = \left(V_2, E_2, f_2\right)$, a **homomorphism of hypergraphs** $h : \mathcal{H}_1 \to \mathcal{H}_2$ consists of two functions, abusively written h, $h : V_1 \to V_2$ and $h : E_1 \to E_2$ such that the diagram:

$$\begin{tikzcd}
E_1 \arrow[r, "f_1"] \arrow[d, "h"] & \mathcal{P}\left(V_1\right) \setminus \{\emptyset\} \arrow[d, "\mathcal{P}(h)"] \\
E_2 \arrow[r, "f_2"] & \mathcal{P}\left(V_2\right) \setminus \{\emptyset\}
\end{tikzcd}$$

is commutative.

We write $\boldsymbol{\mathcal{H}yp}$ the category whose objects are hypergraphs and whose morphisms are hypergraph homomorphisms. We write $\boldsymbol{\mathcal{G}raph}$ the category of simple graphs and their homomorphisms.

The results of interest for this article are the followings:

Lemma 1. *$\boldsymbol{\mathcal{G}raph}$ is a full subcategory of $\boldsymbol{\mathcal{H}yp}$.*

Lemma 2. *Duality of hypergraphs is a covariant functor $d : \boldsymbol{\mathcal{H}yp} \to \boldsymbol{\mathcal{H}yp}$.*

Lemma 3. *The incidence bipartite graph of a hypergraph allows a covariant functor $B : \boldsymbol{\mathcal{H}yp} \to \boldsymbol{\mathcal{G}raph}$.*

Given two hypergraphs $\mathcal{H}_1$ and $\mathcal{H}_2$ and their respective incidence bipartite graph $\mathcal{B}_{\mathcal{H}_1}$ and $\mathcal{B}_{\mathcal{H}_2}$, and a homomorphism $h : \mathcal{H}_1 \to \mathcal{H}_2$, $B\left(h\right) : \mathcal{B}_{\mathcal{H}_1} \to \mathcal{B}_{\mathcal{H}_2}$ with $x \mapsto h(x)$ and $e \mapsto h(e)$.

It is shown in Dörfler and Waller [1980] that B confuses a hypergraph with its dual.

Lemma 4. *The 2-section of a hypergraph induces the definition of a covariant functor $G : \boldsymbol{\mathcal{H}yp} \to \boldsymbol{\mathcal{G}raph}$.*

Lemma 5. *To the intersection graph of a hypergraph corresponds the definition of a covariant functor $I : \boldsymbol{\mathcal{H}yp} \to \boldsymbol{\mathcal{G}raph}$.*

13 Abstract simplicial complexes

We just make a parenthesis on abstract simplicial complexes which are in fact very particular hypergraphs.

In Lee [2010], the author introduces **abstract simplicial complexes** as a collection $\mathcal{K}$ of nonempty finite sets called abstract simplexes that is conditioned to only one rule: if $S \in \mathcal{K}$, then any nonempty subset $S' \subseteq S$, $S' \neq \emptyset$ is also in $\mathcal{K}$. Such a subset is called a face of S and the elements of S are called the **vertices** of S.

Defining $V = \bigcup_{S \in \mathcal{K}} S$ and for every $S \in \mathcal{K}$, $E_S = (e_S)_{e_S \in \mathcal{P}(S)}$ and $E = \underset{S \in \mathcal{K}}{\odot} E_S$ where $\odot$ is the concatenation operation on families, i.e. the family of all the elements of every family that is concatenated. $\mathcal{H} = (V, E)$ is then the unique hypergraph that corresponds to this simplicial complex.

Reciprocally, given a hypergraph $\mathcal{H} = (V, E)$, such that for every $e \in E$, if for all $e' \subseteq e$, $e' \in E$, then $\mathcal{H}$ is in correspondence with a unique abstract simplicial complex.

Simplicial complexes find their strength in the fact that simplicial homology can be used on them, to study the number of holes of their structure. Simplicial homology depends only on the associated topological space and hence gives a way to distinguish simplicial complexes. The interested reader can refer to Hatcher [2005] for an introduction on simplicial homology[4].

As mentioned in Wasserman and Faust [1994], simplicial complexes are useful to study the overlaps between subsets and the connectivity of the network, as well as defining a dimensionality of the network. More complex than hypergraphs by their hierarchical structure, they have been used in different studies such as the structural analysis of a game, using q-holes in Gould and Gatrell [1979].

Abstract simplicial complexes are used quite often in the literature for studying co-occurrence networks. In Nikolaev [2016], the author studies systems with higher-order interaction by considering different models and particularly using a simplicial complex model and a hypergraph model. It is also the case in Salnikov et al. [2018] where co-occurrence simplicial complexes are used to identify the homological holes in mathematic knowledge, in order to discover emerging knowledge. In Benson et al. [2018], the authors study the temporal evolution of different co-occurrence datasets in order to detect and predict higher-order interaction using simplicial closure, studying particularly the life-cycles of triples and quadruples of nodes, showing that the edge density and tie strength are two preponderant factors in the evolution of the group to a closed simplex.

14 Hypergraph coloring

The subject of hypergraph coloring extends the one of graph coloring; but there is not a single way of doing so. Bujtás et al. [2015] is a full survey on the subject: we just give here a very basic definition extracted from it and send the interested reader to this reference for further details.

A **proper hypergraph coloring** is a mapping of a vertex of V to a color number taken in $[\![\lambda]\!]$ where λ is a nonnegative integer such that for every hyperedge there are at least two vertices of different colors. The minimum value of λ for which a proper coloring exists is called the **chromatic number of the hypergraph** $\mathcal{H}$ and written $\chi(\mathcal{H})$. A hypergraph is k-colorable if its chromatic number is no more than k and k-chromatic if $\chi(\mathcal{H}) = k$.

Hypergraph coloring has found applications in finding bounds of the chromatic number of some graphs in Kierstead and Rodl [1996]. They are used in different optimization problems such as divide and conquer approach in Lu [2004]. In Voloshin [1993, 2002], the author explores the coloring of mixed hypergraphs, in which the hyperedge family is partitioned in two sub-families called the hyperedges and anti-hyperedges. The author applies it to problem of energy supply. Coloring is also used in partition problems in Lu [2004].

The subject of hyperedge coloring is also tackled in Gyárfás and Sárközy [2013] to find monochromatic paths and cycles in hypergraphs.

15 Hypergraph partitioning and clustering

Hypergraph partitioning is defined in the Encyclopedia of Parallel Computing[5] as the "task of dividing a hypergraph into two or more roughly equal-sized parts such that a cost function on the hyperedges connecting vertices in different parts is minimized."

Very often this definition is too restrictive and requires strictly more than two parts. The problem of partitioning a hypergraph is at least NP-hard as mentioned in Garey and Johnson [2002]. In Karypis et al. [1999], the authors present a hypergraph partitioning algorithm, called hMetis, based on a multilevel coarsening of the hypergraph, working by

[4] Course of A. Hatcher on Cornell.edu

[5] Hypergraph Partitioning in Encyclopedia of Parallel computing

iterative bisections of the coarsened hypergraphs, starting from the smallest one. In Karypis and Kumar [2000], the authors present an other hypergraph partitioning that is k ways, called hMeTiS-Kway algorithm.

In Papa and Markov [2007], the author gives several methods for hypergraph partitioning and see the action of clustering as "The process of computing a coarser hypergraph from an input hypergraph by merging vertices into larger groups of vertices called clusters." Moreover, the author gives several applications of partitioning and clustering, including VLSI design, numerical linear algebra, automated theorem-proving and formal verification.

The literature is abundant in applications and methods: for more details a survey on clustering ensembles techniques has been done in Ghaemi et al. [2009], which includes hypergraph partitioning.

Clustering and partitioning require often multi-level strategies approach: this is a well studied domain, as it has been extensively used for VLSI design—see for instance, Karypis et al. [1999]—, for parallel scientific computing—Catalyurek and Aykanat [1999], Devine et al. [2006], Ballard et al. [2016]—, for image categorization—Yuchi Huang et al. [2011]—, for social networks—Zhou et al. [2007], Yang et al. [2017].

16 Application of hypergraphs

Hypergraphs fit to model multi-adic relationships in structures where the traditional pairwise relationship of graphs is insufficient: they are used in many areas such as social networks in particular in collaboration networks—Newman [2001b,a]—, co-author networks—Grossman and Ion [1995], Taramasco et al. [2010]—, chemical reactions—Temkin et al. [1996]—, genome—Chauve et al. [2013]—, VLSI design—Karypis et al. [1999]. Hypergraphs are also used in information retrieval for different purposes such as query formulation in text retrieval—Bendersky and Croft [2012]— or in music recommendation—Bu et al. [2010]. Several applications of hypergraphs exist based on the diffusion process firstly developed by Zhou et al. [2007]. Gao et al. [2012] uses Zhou et al. [2007] for 3D-object retrieval and recognition by building multiple hypergraphs of objects based on their 2D-views that are analyzed using the same approach. In Zhu et al. [2015], multiple hypergraphs are constructed to characterize the complex relations between landmark images and are gathered into a multi-modal hypergraph that allows the integration of heterogeneous sources providing content-based visual landmark searches. Hypergraphs are also used in multi-feature indexing to help image retrieval Xu et al. [2016]. For each image, a hyperedge gathers the first n most similar images based on different features. Hyperedges are weighted by average similarity. A spectral clustering algorithm is then applied to divide the dataset into k sub-hypergraphs. A random walk on these sub-hypergraphs allows to retrieve significant images: they are used to build a new inverted index, useful to query images. In Wang et al. [2018], a joint-hypergraph learning is achieved for image retrieval, combining efficiently a semantic hypergraph based on image tags with a visual hypergraph based on image features.

16.1 VLSI design

Very Large Scale Integration—VLSI for short—aims at integrating numerous transistors and devices into a single chip to create an integrated circuit. In VLSI design, hypergraphs are intensively used for placement of chipsets: in this case, the important thing is to achieve an efficient clustering and partitioning of the hypergraph in order to minimize connections between elements of the circuit; more details are given in Papa and Markov [2007]. In this case, the targeted layout is an orthogonal layout. Dedicated algorithms have been implemented to achieve this—Eschbach et al. [2006], Karypis et al. [1999], Karypis and Kumar [2000].

16.2 Collaboration networks

The comments of this Section are directly cited from Ouvrard et al. [2017].

As mentioned in Newman [2001a], a collaboration is a multi-adic relationship between occurrences, and, therefore, the proper modeling is done by hypergraphs. Nonetheless, in Newman [2001a], this multi-adic relationship is approximated by a 2-adic relationship in between pairs of collaborators when it comes to be studied. The same approximation is made in many other studies such as Ramasco et al. [2004].

The reasons for this approximation are numerous. It enables the use of classical graph techniques and properties when studying the different characteristics of collaboration networks, such as degree distribution, clustering coefficient and also when applying quantifying metrics. Today, many different techniques that help with the retrieval of information from graphs are available.

This 2-adic relationship approximation has been developed in many articles, where even if the multi-adic relationship of the data was pointed to be more pertinent, this multi-adic relationship was not used when getting to clustering. Since the end of the year 2000s, the limitations of the 2-adic approach are more and more challenged, as it leads to a partial loss of the information contained in the multi-adic relationship. As a result, in Estrada and Rodriguez-Velazquez [2005] complex networks are modeled by hypergraphs.

In Taramasco et al. [2010], the authors study the academic team formation using epistemic hypergraphs where hyperedges are subsets of unions of a set of agents and a set of concepts. They introduce new features to characterize the evolution of collaboration networks taking into account the hypergraphic nature of networks. This paper brings keystones in the study of a bi-dimensional hypergraph and shows how the keeping of multi-adic relationships can help to gain in the understanding of the evolution of a network.

16.3 Recommender systems and hypergraphs

Recommender systems contain two big methods of recommendation: the item-item recommendation and the user-item recommendation. Both of them require to make groups of items that have been bought for either the purchase of the same item or for the same profile of user.

In Bu et al. [2010], the authors use a unified hypergraph, i.e. a hypergraph that has multiple types of vertices to perform music recommendation that takes not only into account the preference of other users and similar music to the ones they listen. They start by learning on the hypergraph using a cost function with a regularization term that takes the vector of ranking scores as variable and the query vector. The optimal solution for the ranking vector is found when the gradient of the cost function is zero leading to an explicit expression of the recommendation vector, based on the invert of the α-Laplacian matrix of the hypergraph, where α is learned off-line. In Lü et al. [2012], news recommendation is achieved by using a hypergraph partition and a recommendation via ranking on the hypergraph similar to the previous approach on a hybrid hypergraph that contains seven different implicit relations with different objects.

In Zhu et al. [2016], a heterogeneous hypergraph is built regrouping the information on users, tags and documents for document recommendation. They consider different cost functions, one for annotation relations, one for tag relations and one for document relations, based on the Laplacian of the corresponding sub-hypergraphs. The final cost relation is a linear combination of the three cost functions that is optimized using the first k eigenvectors corresponding to the smallest eigenvalues of the total cost matrix.

In Zheng et al. [2018], a social network hybrid recommender system is proposed based on hypergraph topological structure, that combined with a hybrid matrix factorization model helps to enhance the description of the interior relationships of a social network; in particular the neighborhood of the users and items is used.

Additional references can be found to related work in the previous references and in Lü et al. [2012].

16.4 Hypergraph, machine learning and neural networks

In graph neural networks—GNN for short—, the goal is to learn a graph through a neural network, i.e. captures as mentioned in the review of GNN in Zhou et al. [2018] the dependence of graphs via message passing between the nodes of a graph. Recently, hypergraph neural networks that encode higher order relationships are proposed in Feng et al. [2019] and are shown to overtake classical methods.

16.5 Hypergraph grammar

Hypergraph grammars are extension of graph grammars. Graph grammars, also known as graph rewriting systems, aim at enumerating all the possible graphs from a starting graph, by considering a set of graph rewriting rules, searching an occurrence of the pattern graph (i.e. the current graph) and replacing it by an instance of the replacement graph. Graph rewriting is intensively used in software engineering for construction and verification, in layout algorithms, picture generation but also in chemistry for the search of new molecules.

Several studies exist on hypergraph grammars. Particularly, in Drewes et al. [1997], a handle hyperedge replacement is proposed: a handle is a sub-hypergraph constituted of a hyperedge with its incident vertices, extending the hyperedge replacement proposed in Habel and Kreowski [1987].

Application of hypergraph grammars to molecular hypergraphs is still an active subject of research—Kajino [2018] for instance. Hypergraph rewriting is also used for name binding in Yasen and Ueda [2018], and also for graph design in Luerssen and Powers [2007].

16.6 Hypergraphs and linear algebra calculus

In Catalyurek and Aykanat [1999], the authors use a hypergraph-partitioning of the rows (or columns) of a sparse matrix that has to be multiplied by a vector to be used by an iterative solver in order to achieve parallel computation. The partitioning aims at minimizing the total communication volume required between its different segments. The authors use two hypergraphs: one in which the vertices represent the rows of the matrix and the hyperedges represent the columns that have a nonzero intersection with the rows and the other, its dual model.

There are several methods of matrix factorization that use hypergraph regularization. They are all based on the fact that a hypergraph can represent a matrix, by considering either the rows as vertices and columns as hyperedges containing vertices of nonzero elements or reversely using the dual hypergraph. This approach is applied to propose a hypergraph-based non-symmetrical nested dissection ordering algorithm for LU decomposition of sparse matrices in Grigori et al. [2010].

In Zeng et al. [2014], the authors propose a Hypergraph regularized Non-negative Matrix Factorization (HNMF for short) to refine the classical NMF approach and apply it to image clustering.

In Jin et al. [2015], the authors propose a multiple hypergraph regularized low-rank matrix factorization, where the pool of hypergraphs used for the regularization comes from the selection of neighbors in the manifold through the bandwidth parameter used in the heat kernel. The hypergraph Laplacian matrices are combined together allowing the construction of a multi-hypergraph regularization that is added to the regularization term of the original Truncated Singular Value Decomposition—TSVD for short.

In Wu et al. [2018], the authors propose a mixed hypergraph regularized non-negative matrix factorization framework that constrains the vertices of a hyperedge to be projected onto the same latent subspace. The heterogeneity comes from the fact that hyperedges are built using two kind of neighbors, one based on topological information and the other on similarity information.

17 Some additional comments on hypergraphs

A common objection to hypergraphs is: "O.K., hypergraphs extend graphs, but the same can be done with bipartite graphs." To answer this objection different arguments can be developed. We regroup here some of the common arguments found either on Internet[6][7] or in the literature.

First, graphs are particular case of hypergraphs: every graph is a hypergraph, but not all hypergraphs are graphs. Bipartite graphs are particular case of graphs: not every graph is a bipartite graph, but all bipartite graphs are graphs. Hence:

$$\mathcal{B} \subsetneq \mathcal{G} \subsetneq \mathcal{H}$$

writing $\mathcal{B}$ the set of all bipartite graphs, $\mathcal{G}$ the set of all graphs and $\mathcal{H}$ the set of all hypergraphs.

Second, it is true that to a given hypergraph $\mathcal{H} = (V, E)$ corresponds a bipartite graph $G = (V \cup V', E_G)$. The bipartite graph is obtained by considering for each hyperedge $e_j \in E$ an additional vertex v_{e_j}. $V' = \left\{v_{e_j} : e_j \in E\right\}$ and that there is an edge between $v \in V$ and $v_{e_j} \in V'$ if and only if $v \in e_j$. The bipartite graph is then called the incidence graph of the hypergraph $\mathcal{H}$. Conversely, only bipartite graphs with no isolated vertex in any of their vertex part can represent a hypergraph with non-empty hyperedge, as required in Berge [1967, 1973].

Third, from the previous argument, we can develop the argument of factorization. A hypergraph is a factorized version of its incidence graph: a hyperedge e_j is defined as a collection of distinct vertices: $e_j = \left\{v_{j_i} : i \in [\![n_j]\!]\right\}$,with n_j representing the cardinality of e_j. The associated bipartite graph, also called incidence graph, is the developed version: there is an edge—i.e. a pair (v_{j_i}, e_j)—per vertex membership of hyperedge and the vertex set is extended to hold a vertex representation of the original hyperedge. Hence, the vertex set of the incidence graph is varying depending on the hyperedges content, which is not the case for hypergraphs. Hypergraphs bring the power of sets: operations on sets such as union, intersection, complementation or subsets are well defined.

Nonetheless, one has to remark that a hypergraph and its dual have the same corresponding bipartite graph: the bipartite graph confuses the hypergraph with its dual Dörfler and Waller [1980].

[6] https://cs.stackexchange.com/questions/12769/how-is-a-hypergraph-different-from-a-bipartite-graph
[7] http://en.wikipedia.org/wiki/Hypergraph#Bipartite_graph_model

Fourth, bipartite graphs correspond exactly to 2-colorable graphs. Coloring a hypergraph is defined in Berge [1967, 1973] as assigning a color to each vertex such that no hyperedge, differing from a singleton, is monochromatic. In Seymour [1974], a hypergraph that is not 2-colourable but where every proper subset is 2-colourable is called a condenser. Condensers are shown to be minimal hypergraphs that are not 2-colourable. Examples of such condensers are: $\{\{1,2\},\{2,3\},\ldots,\{n-1,n\},\{n,1\}\}$ with n odd or $\{\{1,2,\ldots,n\}\}\cup\{\{0,i\}:1\leqslant i\leqslant n\}$ with $n\geqslant 2$.

Fifth, in Zykov [1974], the author insists on the fact that while the isomorphism between a hypergraph and its incident bipartite representation, called the König graph, is indubitable, some of the statements made with hypergraphs have a very "artificial and cumbersome character in terms of König representation which obscures the point". The author takes the example of the chromatic number definition of an ordinary graph, which is a very particular case of hypergraph, in terms of its König representation.

Sixth, in Berge [1967, 1973], the author focuses on some problems that have solutions only with hypergraphs: the representative graph or line graph is the 2-section of the dual hypergraph. Given a representative graph G, the question raised is to know if there exists a hypergraph $\mathcal{H}$ having G as its representative graph. The problem is shown feasible for general hypergraphs, but Berge [1967, 1973] shows that this problem has no solution when it is required that $\mathcal{H}$ is r-uniform and that a vertex belongs to more than r cliques. Hence, the graph shown in Figure 10 cannot be the representative graph of a 2-uniform hypergraph.

18 Hyper-bag-graphs: a generalization of hypergraphs

Limitations of hypergraphs occur when appears the need to weight vertices independently at the level of each hyperedge, what could be named a hyperedge-based weighting of vertices, as well when the need of repetition of some vertices inside the same hyperedge takes place. This kind of approach is used several time in the literature as mentioned in Ouvrard et al. [2020], for instance in Bellaachia and Al-Dhelaan [2013] where the authors show that retrieving information with a "hyperedge-based" weighting improve the retrieval compared to the binary—in the sense of belong or not to the hyperedge—approach.

In this case, one needs multiset to formalize this concept, implying to have another algebra than the one that takes place with sets. Families of subset of a vertex set—i.e. hypergraphs—are then generalized to families of multisets of same universe called the vertex set, called hyper-bag-graphs (or hb-graphs for short). It is worth mentioning that at the same time we were introducing the hb-graphs in Ouvrard et al. [2018a,b], the authors in Kamiński et al. [2019] independently consider a—abusively named—hypergraph where the hyperedges are multisets to achieve clustering via hypergraph modularity. It is worth also mentioning after these two publications, a hypergraph with hyperedge-dependent vertex weights in Chitra and Raphael [2019]. The usage of multisets is implicit, but without considering the related algebra. In a recent article Pedronette et al. [2019], a continuous incident matrix for multimedia retrieval is used that is no more that our hb-graph incident matrix.

One can consult several applications of hypergraphs in the Thesis of the present author Ouvrard et al. [2020]. We just introduce here the basics of hb-graphs that have been given through the cited articles.

A multiset $\mathfrak{A}_m=(A,m)$ is a couple where A is a set and $m:A\mapsto\mathbb{R}^+$. The set $A^\star=\{x\in A:m(x)\neq 0\}$ is called the support of $\mathfrak{A}_m$. It is not the place to make a full introduction on multisets, the interested reader can refer to Syropoulos [2001] for a full introduction on multisets, and particularly on the way the union and intersection of two multisets is defined. With multisets the Morgan laws do not hold. Multisets have been largely used in the literature with name such as bags, collections,... Some authors have started to consider some sets of multisets—Pearson and Zhang [2014]—or collection of multisets for defining a multiset topology by considering a collection of multisets in the power set of a given multiset—Girish and Jacob [2012]: this latest is a strong background to our work, but have to be seen as a potential extension of simplicial complexes, but remains as a particular case of hb-graphs like simplicial complexes are particular cases of hypergraphs.

A **hyper-bag-graph or hb-graph** for short over $V=\{v_i:i\in[\![n]\!]\}$ is defined in Ouvrard et al. [2018b] as a family of multisets of universe V and support a subset of V. The multisets are called the **hb-edges** and the elements of V the **vertices**.

Each hb-edge $\mathfrak{e}_j\in\mathfrak{E}$ is of universe V and has a multiplicity function associated to it: $m_{\mathfrak{e}_j}:V\to\mathbb{W}$ where $\mathbb{W}\subseteq\mathbb{R}^+$. When the context is clear the notation m_j is used for $m_{\mathfrak{e}_j}$ and m_{ij} for $m_{\mathfrak{e}_j}(v_i)$. Most of the definitions extend with no difficulty to $\mathbb{W}\subseteq\mathbb{R}$.

A hb-graph having each of its multiplicity function range a subset of the non-negative integers is called a **natural hb-graph**.

A hypergraph appears as a very particular case of natural hb-graph where the multiplicity function has its range in $\{0, 1\}$.

As a first immediate application, we consider the set of prime numbers. The prime dividers of a positive integer can be gathered using a subset of the prime numbers; so given a list of positive integers and their associate prime dividers, we have a hypergraph where the vertices are the prime numbers that are divider of one of the number in the list and the hyperedges are the set of dividers of a number in the list. This hypergraph illustrates the common dividers of two numbers. But taking the same list of numbers, if you want to have the exact decomposition in primes of each number, it will require a hb-graph to store the information, where the universe or vertex set is the set of primes, and where a decomposition of a given number is a hb-edge. By this means, each number can be decomposed in its prime decomposition and have a nice representation of that decomposition using hb-graph representations. For a full approach on the subject, one can refer to Ouvrard et al. [2019], Ouvrard [2020], where gcd and lcm are associated with operations on hb-graphs related to operations that can be done with multisets.

Hb-graphs can be represented with a similar representation to hypergraphs, particularly the extra-node representation in the case of natural hb-graph, but this time the underlying graph is a bipartite multi-graph, i.e. vertices and hb-edges are represented by nodes of different shapes and edges can be multiple between a vertex and a hb-edge. For general hb-graphs, the thickness of the edge will be made proportional to the multiplicity the vertex has in the corresponding hb-edge.

To a hb-graph is always associated a unique support hypergraph with same vertex set that is the hypergraph of the support of its hb-edges. A given hypergraph is in contrast the support of an infinity of hb-graphs.

As mentioned previously, the incident matrix of a hypergraph is a binary matrix where rows represent vertices and columns are associated to hyperedges and there is a value of 1 at an intersection of a row and a column if the vertex is in the hyperedge. With hb-graphs, the incident matrix has at the intersection of a row and a column the multiplicity of the vertex inside the hb-edge. Hypergraphs are bijectively in link with the set of binary matrices, hb-graphs are bijectively in link with the set of real matrices.

Hb-graphs open the door to refinements of results that are obtained with hypergraphs, particularly in machine learning; we have shown a few applications, using diffusion over such structure. Other approach can be thought such as using hb-graphs with neural networks, extending approaches taken in GNN—graph neural networks—and HNN—hypergraph neural networks.

19 Acknowledgments

This work was mostly done during the writing of the Thesis of Xavier OUVRARD, defended successfully in March 2020 at the University of Geneva. The author is really thankful to Pr. Stéphane MARCHAND-MAILLET and Dr. Jean-Marie LE GOFF, who have co-directed the PhD, that had been funded by a doctoral position in the Collaboration Spotting team at CERN IPT department.

References

Jiří Adámek, Horst Herrlich, and George E Strecker. Abstract and concrete categories. The joy of cats. 2004. URL http://katmat.math.uni-bremen.de/acc/acc.pdf.

Bilal Alsallakh, Luana Micallef, Wolfgang Aigner, Helwig Hauser, Silvia Miksch, and Peter Rodgers. The State-of-the-Art of Set Visualization: The State-of-the-Art of Set Visualization. *Computer Graphics Forum*, 35(1):234–260, February 2016. ISSN 01677055. doi: 10.1111/cgf.12722. URL https://doi.org/10.1111/cgf.12722.

Grey Ballard, Alex Druinsky, Nicholas Knight, and Oded Schwartz. Hypergraph Partitioning for Sparse Matrix-Matrix Multiplication. *ACM Transactions on Parallel Computing*, 3(3):1–34, December 2016. ISSN 23294949. doi: 10.1145/3015144. URL https://doi.org/10.1145/3015144.

Abdelghani Bellaachia and Mohammed Al-Dhelaan. Random walks in hypergraph. In *Proceedings of the 2013 International Conference on Applied Mathematics and Computational Methods, Venice Italy*, pages 187–194, 2013.

Michael Bendersky and W. Bruce Croft. Modeling higher-order term dependencies in information retrieval using query hypergraphs. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval - SIGIR '12*, page 941, Portland, Oregon, USA, 2012. ACM Press. ISBN 978-1-4503-1472-5. doi: 10.1145/2348283.2348408. URL https://doi.org/10.1145/2348283.2348408.

Austin R. Benson, Rediet Abebe, Michael T. Schaub, Ali Jadbabaie, and Jon Kleinberg. Simplicial closure and higher-order link prediction. *Proceedings of the National Academy of Sciences*, 115(48):E11221–E11230, November 2018. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1800683115. URL https://doi.org/10.1073/pnas.1800683115.

Claude Berge. Graphes et hypergraphes. 1970. *Dunod, Paris*, 1967.

Claude Berge. *Graphs and hypergraphs*, volume 7. North-Holland publishing company Amsterdam, 1973.

François Bertault and Peter Eades. Drawing Hypergraphs in the Subset Standard (Short Demo Paper). In *Graph Drawing*, volume 1984, pages 164–169. Springer Berlin Heidelberg, Berlin, Heidelberg, 2001. ISBN 978-3-540-41554-1 978-3-540-44541-8. doi: 10.1007/3-540-44541-2_15. URL https://doi.org/DOI:10.1007/3-540-44541-2_15.

Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, October 2008. ISSN 1742-5468. doi: 10.1088/1742-5468/2008/10/P10008. URL https://doi.org/10.1088/1742-5468/2008/10/P10008.

Alain Bretto. *Hypergraph Theory*. Mathematical Engineering. Springer International Publishing, Heidelberg, 2013. ISBN 978-3-319-00079-4 978-3-319-00080-0. doi: 10.1007/978-3-319-00080-0. URL https://doi.org/10.1007/978-3-319-00080-0.

Jiajun Bu, Shulong Tan, Chun Chen, Can Wang, Hao Wu, Lijun Zhang, and Xiaofei He. Music recommendation by unified hypergraph: combining social media information and music content. In *Proceedings of the international conference on Multimedia - MM '10*, page 391, Firenze, Italy, 2010. ACM Press. ISBN 978-1-60558-933-6. doi: 10.1145/1873951.1874005. URL https://doi.org/10.1145/1873951.1874005.

Csilla Bujtás, Zsolt Tuza, and Vitaly Voloshin. Hypergraph colouring. In *Topics in Chromatic Graph Theory*, pages 230–254. Cambridge University Press, Cambridge, 2015. ISBN 978-1-139-51979-3. doi: 10.1017/CBO9781139519793.014. URL https://doi.org/10.1017/CBO9781139519793.014.

U.V. Catalyurek and C. Aykanat. Hypergraph-partitioning-based decomposition for parallel sparse-matrix vector multiplication. *IEEE Transactions on Parallel and Distributed Systems*, 10(7):673–693, July 1999. ISSN 10459219. doi: 10.1109/71.780863. URL https://doi.org/10.1109/71.780863.

Cedric Chauve, Murray Patterson, and Ashok Rajaraman. Hypergraph Covering Problems Motivated by Genome Assembly Questions. In *Combinatorial Algorithms*, volume 8288, pages 428–432. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013. ISBN 978-3-642-45277-2 978-3-642-45278-9. doi: 10.1007/978-3-642-45278-9_37. URL https://doi.org/10.1007/978-3-642-45278-9_37.

B. Chazelle and J. Friedman. A deterministic view of random sampling and its use in geometry. In *[Proceedings 1988] 29th Annual Symposium on Foundations of Computer Science*, pages 539–549, White Plains, NY, USA, 1988. IEEE. ISBN 978-0-8186-0877-3. doi: 10.1109/SFCS.1988.21970. URL https://doi.org/10.1109/SFCS.1988.21970.

Uthsav Chitra and Benjamin J Raphael. Random Walks on Hypergraphs with Edge-Dependent Vertex Weights. *arXiv preprint arXiv:1905.08287*, 2019.

K.D. Devine, E.G. Boman, R.T. Heaphy, R.H. Bisseling, and U.V. Catalyurek. Parallel hypergraph partitioning for scientific computing. In *Proceedings 20th IEEE International Parallel & Distributed Processing Symposium*, page 10 pp., Rhodes Island, Greece, 2006. IEEE. ISBN 978-1-4244-0054-6. doi: 10.1109/IPDPS.2006.1639359. URL https://doi.org/10.1109/IPDPS.2006.1639359.

W. Dörfler and D. A. Waller. A category-theoretical approach to hypergraphs. *Archiv der Mathematik*, 34(1):185–192, December 1980. ISSN 0003-889X, 1420-8938. doi: 10.1007/BF01224952. URL https://doi.org/10.1007/BF01224952.

F. Drewes, H.-J. Kreowski, and A. Habel. Hyperedge replacement graph grammars. In *Handbook of Graph Grammars and Computing by Graph Transformation*, pages 95–162. WORLD SCIENTIFIC, February 1997. ISBN 978-981-02-2884-2 978-981-238-472-0. doi: 10.1142/9789812384720_0002. URL https://doi.org/10.1142/9789812384720_0002.

Thomas Eschbach, Wolfgang Guenther, and Bernd Becker. Orthogonal Hypergraph Drawing for Improved Visibility. *Journal of Graph Algorithms and Applications*, 10(2):141–157, 2006. ISSN 1526-1719. doi: 10.7155/jgaa.00122. URL https://doi.org/10.7155/jgaa.00122.

Ernesto Estrada and Juan A Rodriguez-Velazquez. Complex networks as hypergraphs. *arXiv preprint physics/0505137*, 2005.

Yifan Feng, Haoxuan You, Zizhao Zhang, Rongrong Ji, and Yue Gao. Hypergraph Neural Networks. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33:3558–3565, July 2019. ISSN 2374-3468, 2159-5399. doi: 10.1609/aaai.v33i01.33013558. URL https://doi.org/10.1609/aaai.v33i01.33013558.

Yue Gao, Meng Wang, Dacheng Tao, Rongrong Ji, and Qionghai Dai. 3-D Object Retrieval and Recognition With Hypergraph Analysis. *IEEE Transactions on Image Processing*, 21(9):4290–4303, September 2012. ISSN 1057-7149, 1941-0042. doi: 10.1109/TIP.2012.2199502. URL https://doi.org/10.1109/TIP.2012.2199502.

Michael R Garey and David S Johnson. *Computers and intractability*, volume 29. wh freeman New York, 2002.

Reza Ghaemi, Md Nasir Sulaiman, Hamidah Ibrahim, Norwati Mustapha, and others. A survey: clustering ensembles techniques. *World Academy of Science, Engineering and Technology*, 50:636–645, 2009.

KP Girish and John Sunil Jacob. On multiset topologies. *Theory and Applications of Mathematics & Computer Science*, 2(1):37–52, 2012.

Peter Gould and Anthony Gatrell. A structural analysis of a game: The Liverpool v Manchester united cup final of 1977. *Social Networks*, 2(3):253–273, January 1979. ISSN 03788733. doi: 10.1016/0378-8733(79)90017-0. URL https://doi.org/10.1016/0378-8733(79)90017-0.

Laura Grigori, Erik G. Boman, Simplice Donfack, and Timothy A. Davis. Hypergraph-Based Unsymmetric Nested Dissection Ordering for Sparse LU Factorization. *SIAM Journal on Scientific Computing*, 32(6):3426–3446, January 2010. ISSN 1064-8275, 1095-7197. doi: 10.1137/080720395. URL https://doi.org/10.1137/080720395.

Jerrold W Grossman and Patrick DF Ion. On a portion of the well-known collaboration graph. *Congressus Numerantium*, pages 129–132, 1995.

András Gyárfás and Gábor N Sárközy. Monochromatic path and cycle partitions in hypergraphs. *the electronic journal of combinatorics*, 20(1):18, 2013.

Annegret Habel and Hans-Jörg Kreowski. Some structural aspects of hypergraph languages generated by hyperedge replacement. In Franz J. Brandenburg, Guy Vidal-Naquet, and Martin Wirsing, editors, *STACS 87*, volume 247, pages 207–219. Springer-Verlag, Berlin/Heidelberg, 1987. ISBN 978-3-540-17219-2. doi: 10.1007/BFb0039608. URL https://doi.org/10.1007/BFb0039608.

Allen Hatcher. *Algebraic topology*. 2005.

Ahmed Isah and Y. Tella. The Concept of Multiset Category. *British Journal of Mathematics & Computer Science*, 9(5):427–437, January 2015. ISSN 22310851. doi: 10.9734/BJMCS/2015/18415. URL https://doi.org/10.9734/bjmcs/2015/18415.

Mathieu Jacomy, Tommaso Venturini, Sebastien Heymann, and Mathieu Bastian. ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software. *PLoS ONE*, 9(6):e98679, June 2014. ISSN 1932-6203. doi: 10.1371/journal.pone.0098679. URL https://dx.plos.org/10.1371/journal.pone.0098679.

Taisong Jin, Jun Yu, Jane You, Kun Zeng, Cuihua Li, and Zhengtao Yu. Low-rank matrix factorization with multiple Hypergraph regularizer. *Pattern Recognition*, 48(3):1011–1022, March 2015. ISSN 00313203. doi: 10.1016/j.patcog.2014.09.002. URL https://doi.org/10.1016/j.patcog.2014.09.002.

D. S. Johnson and H. O. Pollak. Hypergraph planarity and the complexity of drawing venn diagrams. *Journal of Graph Theory*, 11(3):309–325, 1987. ISSN 03649024, 10970118. doi: 10.1002/jgt.3190110306. URL http://doi.wiley.com/10.1002/jgt.3190110306.

Martin Junghans. *Visualization of hyperedges in fixed graph layouts*. PhD Thesis, Master\rqs thesis, Brandenburg University of Technology, Cottbus, 2008.

Hiroshi Kajino. Molecular hypergraph grammar with its application to molecular optimization. *arXiv preprint arXiv:1809.02745*, 2018.

Bogumił Kamiński, Valérie Poulin, Paweł Prałat, Przemysław Szufel, and François Théberge. Clustering via hypergraph modularity. *PLOS ONE*, 14(11):e0224307, November 2019. ISSN 1932-6203. doi: 10.1371/journal.pone.0224307. URL https://doi.org/10.1371/journal.pone.0224307.

G. Karypis, R. Aggarwal, V. Kumar, and S. Shekhar. Multilevel hypergraph partitioning: applications in VLSI domain. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 7(1):69–79, March 1999. ISSN 1063-8210, 1557-9999. doi: 10.1109/92.748202. URL https://doi.org/10.1109/92.748202.

George Karypis and Vipin Kumar. Multilevel k-way Hypergraph Partitioning. *VLSI Design*, 11(3):285–300, January 2000. ISSN 1065-514X, 1563-5171. doi: 10.1155/2000/19436. URL https://doi.org/10.1155/2000/19436.

Michael Kaufmann, Marc van Kreveld, and Bettina Speckmann. Subdivision Drawings of Hypergraphs. In David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Friedemann Mattern, John C. Mitchell, Moni Naor, Oscar Nierstrasz, C. Pandu Rangan, Bernhard Steffen, Madhu Sudan, Demetri Terzopoulos, Doug Tygar, Moshe Y. Vardi, Gerhard Weikum, Ioannis G. Tollis, and Maurizio Patrignani, editors, *Graph Drawing*, volume 5417, pages 396–407. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009. ISBN 978-3-642-00218-2 978-3-642-00219-9. doi: 10.1007/978-3-642-00219-9_39. URL https://doi.org/10.1007/978-3-642-00219-9_39.

Andreas Kerren and Ilir Jusufi. A novel radial visualization approach for undirected hypergraphs. *Eurographics Association*, pages 25–29, 2013.

H.A. Kierstead and V. Rodl. Applications of hypergraph coloring to coloring graphs not inducing certain trees. *Discrete Mathematics*, 150(1-3):187–193, April 1996. ISSN 0012365X. doi: 10.1016/0012-365X(95)00187-2. URL https://doi.org/10.1016/0012-365X(95)00187-2.

John Lee. *Introduction to topological manifolds*, volume 202. Springer Science & Business Media, 2010.

Chi-Jen Lu. Deterministic hypergraph coloring and its applications. *SIAM Journal on Discrete Mathematics*, 18(2): 320–331, 2004.

Linyuan Lü, Matúš Medo, Chi Ho Yeung, Yi-Cheng Zhang, Zi-Ke Zhang, and Tao Zhou. Recommender systems. *Physics reports*, 519(1):1–49, 2012.

Martin H Luerssen and David MW Powers. Graph design by graph grammar evolution. In *2007 IEEE Congress on Evolutionary Computation*, pages 386–393. IEEE, 2007.

Erkki Mäkinen. How to draw a hypergraph. *International Journal of Computer Mathematics*, 34(3-4):177–185, 1990. doi: 10.1080/00207169008803875. URL https://doi.org/10.1080/00207169008803875.

M. E. J. Newman. Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. *Physical Review E*, 64(1):016132, June 2001a. ISSN 1063-651X, 1095-3787. doi: 10.1103/PhysRevE.64.016132. URL https://link.aps.org/doi/10.1103/PhysRevE.64.016132.

M. E. J. Newman. Scientific collaboration networks. I. Network construction and fundamental results. *Physical Review E*, 64(1):016131, June 2001b. ISSN 1063-651X, 1095-3787. doi: 10.1103/PhysRevE.64.016131. URL https://doi.org/10.1103/PhysRevE.64.016131.

Alexey Nikolaev. Modeling and analysis of higher-order interactions in collaborative networks. 2016.

Xavier Ouvrard, Jean-Marie Le Goff, and Stéphane Marchand-Maillet. Networks of Collaborations: Hypergraph Modeling and Visualisation. *arXiv preprint arXiv:1707.00115*, 2017.

Xavier Ouvrard, Jean-Marie Le Goff, and Stéphane Marchand-Maillet. Adjacency and Tensor Representation in General Hypergraphs. Part 2: Multisets, Hb-graphs and Related e-adjacency Tensors. *arXiv preprint arXiv:1805.11952*, 2018a.

Xavier Ouvrard, Jean-Marie Le Goff, and Stéphane Marchand-Maillet. Diffusion by Exchanges in HB-Graphs: Highlighting Complex Relationships. *CBMI Proceedings*, 2018b. doi: 10.1109/CBMI.2018.8516525. URL https://doi.org/10.1109/CBMI.2018.8516525.

Xavier Ouvrard, Jean-Marie Le Goff, and Stéphane Marchand-Maillet. On Hb-graphs and their Application to General Hypergraph e-adjacency Tensor. *MCCCC32 Special Volume of the Journal of Combinatorial Mathematics and Combinatorial Computing, to be published*, 2019.

Xavier Ouvrard, Jean-Marie Le Goff, and Stéphane Marchand-Maillet. The HyperBagGraph DataEdron: An Enriched Browsing Experience of Datasets. *LNCS, 46th International Conference on Current Trends in Theory and Practice of Computer Science (SOFSEM 2020)*, 2020.

Xavier Eric Ouvrard. *Hyper-bag-graphs and their applications: Modeling, Analyzing and Visualizing Complex Networks of Co-occurrences*. 2020. doi: 10.13097/archive-ouverte/unige:137520. URL https://archive-ouverte.unige.ch/unige:137520.

David Papa and Igor Markov. Hypergraph Partitioning and Clustering. In Teofilo Gonzalez, editor, *Handbook of Approximation Algorithms and Metaheuristics*, volume 20073547, pages 61–1–61–19. Chapman and Hall/CRC, May 2007. ISBN 978-1-58488-550-4 978-1-4200-1074-9. doi: 10.1201/9781420010749.ch61. URL https://doi.org/10.1201/9781420010749.ch61.

Jesse Paquette and Taku Tokuyasu. Hypergraph visualization and enrichment statistics: how the EGAN paradigm facilitates organic discovery from big data. page 78650E, San Francisco Airport, California, USA, February 2011. doi: 10.1117/12.890220. URL https://doi.org/10.1117/12.890220.

Kelly J. Pearson and Tan Zhang. On Spectral Hypergraph Theory of the Adjacency Tensor. *Graphs and Combinatorics*, 30(5):1233–1248, September 2014. ISSN 0911-0119, 1435-5914. doi: 10.1007/s00373-013-1340-x. URL https://doi.org/10.1007/s00373-013-1340-x.

Daniel Carlos Guimaraes Pedronette, Lucas Pascotti Valem, Jurandy Almeida, and Ricardo da S. Torres. Multimedia Retrieval Through Unsupervised Hypergraph-Based Manifold Ranking. *IEEE Transactions on Image Processing*, 28(12):5824–5838, December 2019. ISSN 1057-7149, 1941-0042. doi: 10.1109/TIP.2019.2920526. URL https://doi.org/10.1109/TIP.2019.2920526.

José J. Ramasco, S. N. Dorogovtsev, and Romualdo Pastor-Satorras. Self-organization of collaboration networks. *Physical Review E*, 70(3):036106, September 2004. ISSN 1539-3755, 1550-2376. doi: 10.1103/PhysRevE.70.036106. URL https://doi.org/10.1103/PhysRevE.70.036106.

Vsevolod Salnikov, Daniele Cassese, Renaud Lambiotte, and Nick S. Jones. Co-occurrence simplicial complexes in mathematics: identifying the holes of knowledge. *Applied Network Science*, 3(1):37, December 2018. ISSN 2364-8228. doi: 10.1007/s41109-018-0074-3. URL https://doi.org/10.1007/s41109-018-0074-3.

P. D. Seymour. On the two-colouring of hypergraphs. *The Quarterly Journal of Mathematics*, 25(1):303–311, 1974. ISSN 0033-5606, 1464-3847. doi: 10.1093/qmath/25.1.303. URL https://doi.org/10.1093/qmath/25.1.303.

John G. Stell. Relations on Hypergraphs. In *Relational and Algebraic Methods in Computer Science*, volume 7560, pages 326–341. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012. ISBN 978-3-642-33313-2 978-3-642-33314-9. doi: 10.1007/978-3-642-33314-9_22. URL https://doi.org/10.1007/978-3-642-33314-9_22.

Apostolos Syropoulos. Mathematics of Multisets. In *Multiset Processing*, volume 2235, pages 347–358. Springer Berlin Heidelberg, Berlin, Heidelberg, 2001. ISBN 978-3-540-43063-6 978-3-540-45523-3. doi: 10.1007/3-540-45523-X_17. URL https://doi.org/10.1007/3-540-45523-X_17.

Carla Taramasco, Jean-Philippe Cointet, and Camille Roth. Academic team formation as evolving hypergraphs. *Scientometrics*, 85(3):721–740, December 2010. ISSN 0138-9130, 1588-2861. doi: 10.1007/s11192-010-0226-4. URL https://doi.org/10.1007/s11192-010-0226-4.

Oleg N Temkin, Andrew V Zeigarnik, and DG Bonchev. *Chemical reaction networks: a graph-theoretical approach*. CRC Press, 1996.

John Venn. On the Diagrammatic and Mechanical Representation of Propositions and Reasonings. 1881.

VI Voloshin. The mixed hypergraphs. *Computer Science Journal of Moldova*, 1(1):1, 1993.

Vitaly Voloshin. *Coloring Mixed Hypergraphs: Theory, Algorithms and Applications*, volume 17 of *Fields Institute Monographs*. American Mathematical Society, Providence, Rhode Island, June 2002. ISBN 978-0-8218-2812-0 978-1-4704-3144-0. doi: 10.1090/fim/017. URL https://doi.org/10.1090/fim/017.

Vitaly I Voloshin. *Introduction to graph and hypergraph theory*. Nova Science Publ., 2009. ISBN 978-1-60692-372-6.

Yaxiong Wang, Li Zhu, Xueming Qian, and Junwei Han. Joint Hypergraph Learning for Tag-Based Image Retrieval. *IEEE Transactions on Image Processing*, 27(9):4437–4451, September 2018. ISSN 1057-7149, 1941-0042. doi: 10.1109/TIP.2018.2837219. URL https://doi.org/10.1109/TIP.2018.2837219.

Stanley Wasserman and Katherine Faust. *Social Network Analysis: Methods and Applications*. Cambridge University Press, 1 edition, November 1994. ISBN 978-0-521-38707-1 978-0-521-38269-4 978-0-511-81547-8. doi: 10.1017/CBO9780511815478. URL https://doi.org/10.1017/CBO9780511815478.

Wenhui Wu, Sam Kwong, Yu Zhou, Yuheng Jia, and Wei Gao. Nonnegative matrix factorization with mixed hypergraph regularization for community detection. *Information Sciences*, 435:263–281, April 2018. ISSN 00200255. doi: 10.1016/j.ins.2018.01.008. URL https://doi.org/10.1016/j.ins.2018.01.008.

Zihang Xu, Junping Du, Lingfei Ye, and Dan Fan. Multi-feature indexing for image retrieval based on hypergraph. In *2016 4th International Conference on Cloud Computing and Intelligence Systems (CCIS)*, pages 494–500, Beijing, China, August 2016. IEEE. ISBN 978-1-5090-1256-5. doi: 10.1109/CCIS.2016.7790309. URL https://doi.org/10.1109/CCIS.2016.7790309.

Wenyin Yang, Guojun Wang, Md Zakirul Alam Bhuiyan, and Kim-Kwang Raymond Choo. Hypergraph partitioning for social networks based on information entropy modularity. *Journal of Network and Computer Applications*, 86:59–71, May 2017. ISSN 10848045. doi: 10.1016/j.jnca.2016.10.002. URL https://doi.org/10.1016/j.jnca.2016.10.002.

Alimujiang Yasen and Kazunori Ueda. Name Binding is Easy with Hypergraphs. *IEICE Transactions on Information and Systems*, E101.D(4):1126–1140, 2018. ISSN 0916-8532, 1745-1361. doi: 10.1587/transinf.2017EDP7257. URL https://doi.org/10.1587/transinf.2017edp7257.

Yuchi Huang, Qingshan Liu, Fengjun Lv, Yihong Gong, and Dimitris N Metaxas. Unsupervised Image Categorization by Hypergraph Partition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(6):1266–1273, June 2011. ISSN 0162-8828, 2160-9292. doi: 10.1109/TPAMI.2011.25. URL https://doi.org/10.1109/TPAMI.2011.25.

Kun Zeng, Jun Yu, Cuihua Li, Jane You, and Taisong Jin. Image clustering by hyper-graph regularized non-negative matrix factorization. *Neurocomputing*, 138:209–217, August 2014. ISSN 09252312. doi: 10.1016/j.neucom.2014.01.043. URL https://doi.org/10.1016/j.neucom.2014.01.043.

Xiaoyao Zheng, Yonglong Luo, Liping Sun, Xintao Ding, and Ji Zhang. A novel social network hybrid recommender system based on hypergraph topologic structure. *World Wide Web*, 21(4):985–1013, July 2018. ISSN 1386-145X, 1573-1413. doi: 10.1007/s11280-017-0494-5. URL https://doi.org/10.1007/s11280-017-0494-5.

Denny Zhou, Jiayuan Huang, and Bernhard Schölkopf. Learning with hypergraphs: Clustering, classification, and embedding. In *Advances in neural information processing systems*, pages 1601–1608, 2007.

Jie Zhou, Ganqu Cui, Zhengyan Zhang, Cheng Yang, Zhiyuan Liu, Lifeng Wang, Changcheng Li, and Maosong Sun. Graph neural networks: A review of methods and applications. *arXiv preprint arXiv:1812.08434*, 2018.

Lei Zhu, Jialie Shen, Hai Jin, Ran Zheng, and Liang Xie. Content-Based Visual Landmark Search via Multimodal Hypergraph Learning. *IEEE Transactions on Cybernetics*, 45(12):2756–2769, December 2015. ISSN 2168-2267, 2168-2275. doi: 10.1109/TCYB.2014.2383389. URL https://doi.org/10.1109/TCYB.2014.2383389.

Yu Zhu, Ziyu Guan, Shulong Tan, Haifeng Liu, Deng Cai, and Xiaofei He. Heterogeneous hypergraph embedding for document recommendation. *Neurocomputing*, 216:150–162, December 2016. ISSN 09252312. doi: 10.1016/j.neucom.2016.07.030. URL https://linkinghub.elsevier.com/retrieve/pii/S0925231216307755.

A A Zykov. Hypergraphs. *Russian Mathematical Surveys*, 29(6):89–156, December 1974. ISSN 0036-0279, 1468-4829. doi: 10.1070/RM1974v029n06ABEH001303. URL https://doi.org/10.1070/RM1974v029n06ABEH001303.

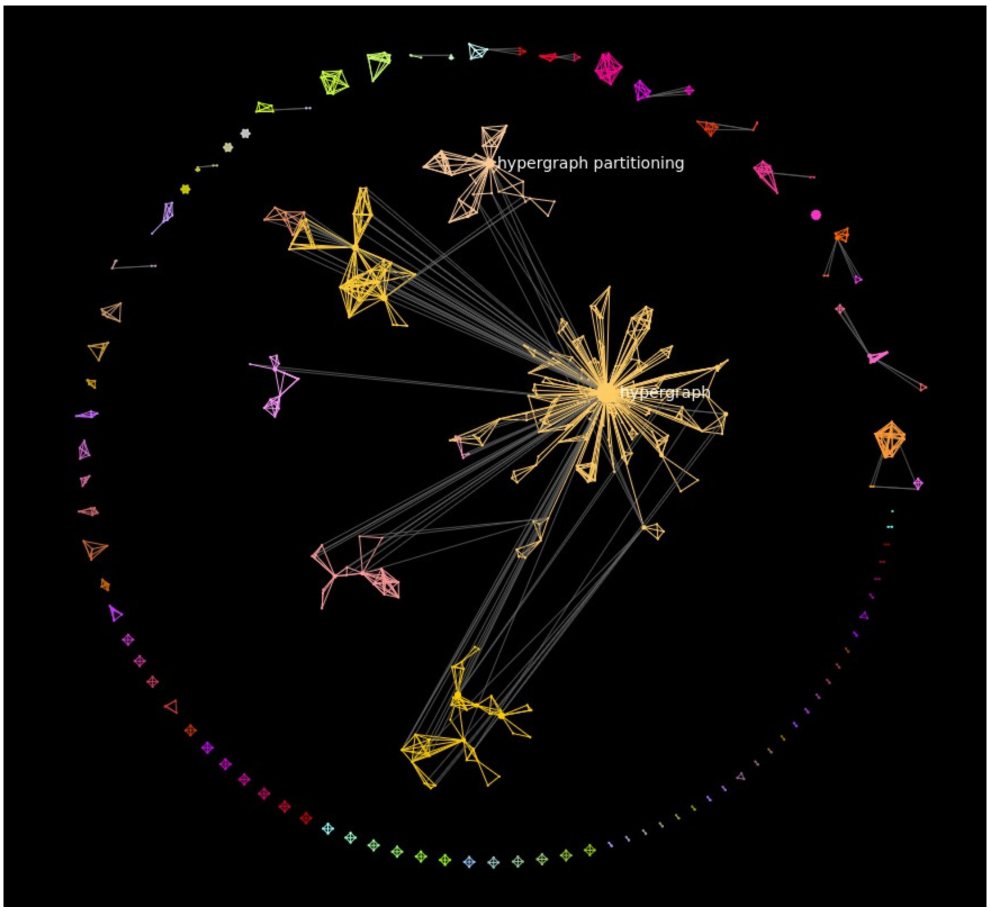


Figure 7: Keyword collaborations from search: “TITLE: hypergraph”.
#publications = 200, #nodes = 707, #edges = 1655, #clusters = 102, #isolated clusters = 67.

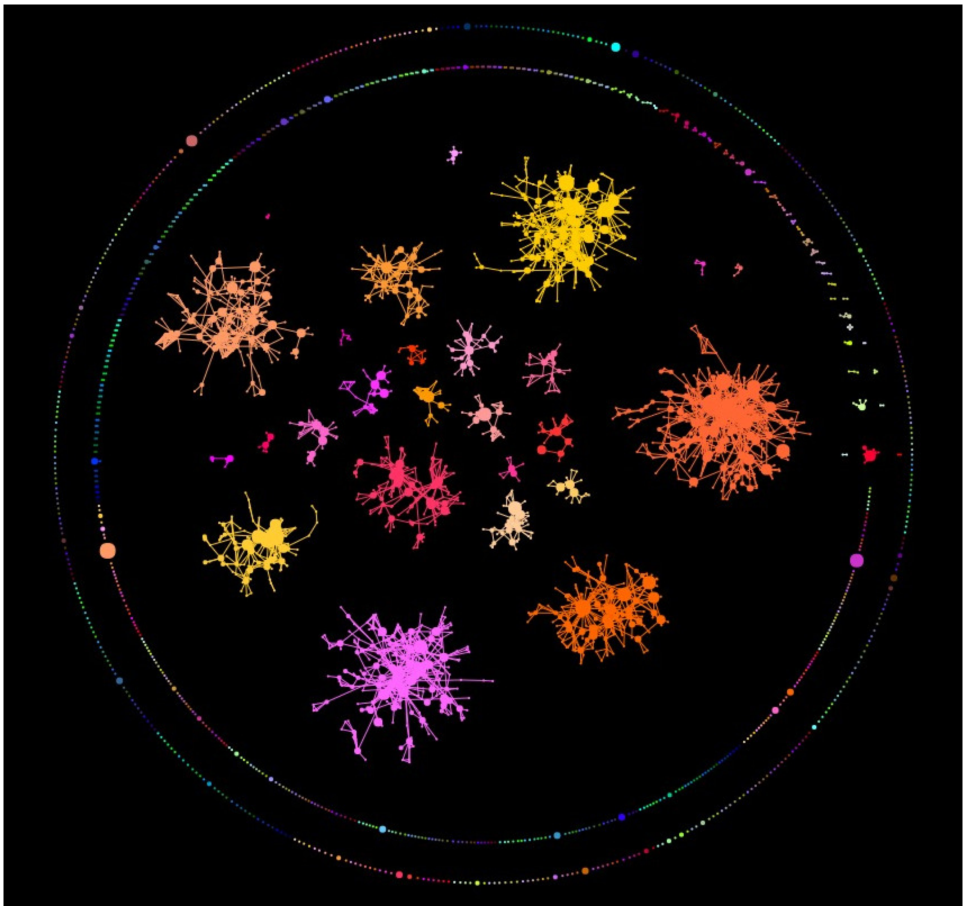


Figure 8: Organization collaborations from search: “TITLE:graph”.
#publications = 3969, #patents=893
#nodes = 2932, #edges = 4731, #clusters = 951, #isolated clusters = 914.

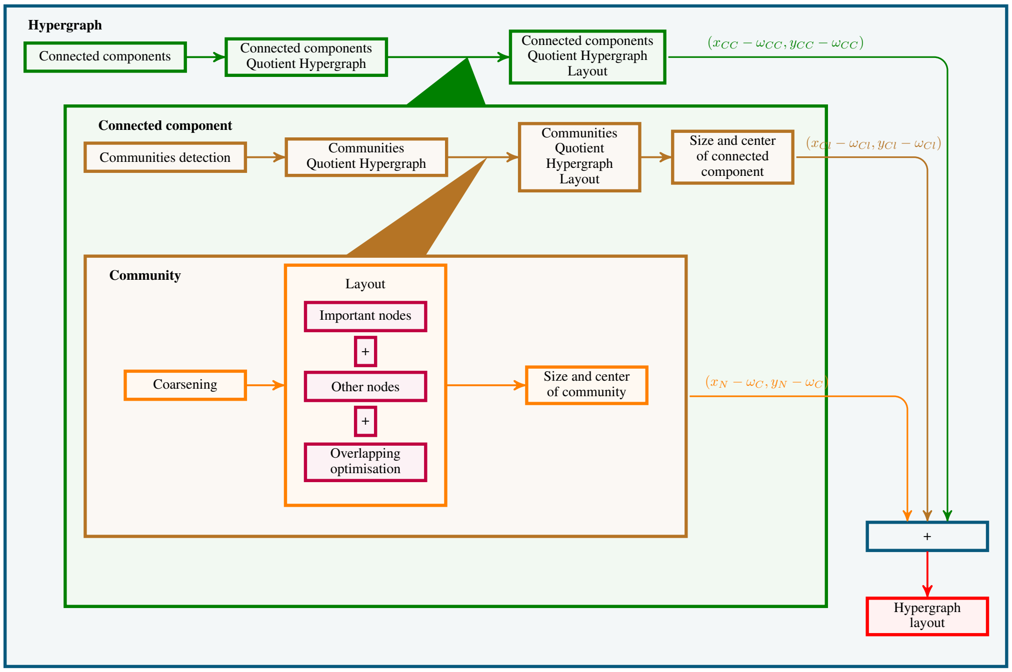


Figure 9: Principle of the calculation of coordinates for large hypergraphs.

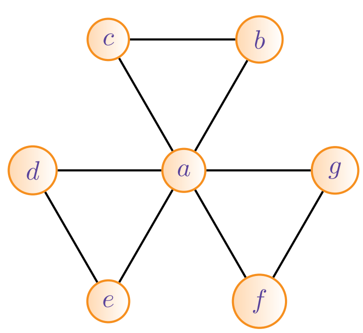


Figure 10: Counter-example of graph that cannot be the representative graph of a 2-uniform hypergraph.