Exchange-Based Diffusion in Hb-Graphs: Highlighting Complex Relationships

Xavier Ouvrard
University of Geneva / CERN
Meyrin, Switzerland
xavier.ouvrard@cern.ch

Jean-Marie Le Goff
CERN
Meyrin, Switzerland
jean-marie.le.goff@cern.ch

Stéphane Marchand-Maillet
CUI
University of Geneva
Carouge, Switzerland
Stephane.Marchand-Maillet@unige.ch

Abstract—Most networks tend to show complex and multiple relationships between entities. Networks are usually modeled by graphs or hypergraphs; nonetheless a given entity can occur many times in a relationship: this brings the need to deal with multisets instead of sets or simple edges. Diffusion processes are useful to highlight interesting parts of a network: they usually start with a stroke at one vertex and diffuse throughout the network to reach a uniform distribution. Several iterations of the process are required prior to reaching a stable solution. We propose an alternative solution to highlighting main components of a network using a diffusion process based on exchanges; it is an iterative two-phase step exchange process. This process allows to evaluate the importance not only at the vertices level but also at the regrouping level. To model the diffusion process, we extend the concept of hypergraphs that are family of sets to family of multisets, that we call hb-graphs.

Index Terms—exchange, hypergraph, hb-graph, visualisation

I. Introduction

Many relationships are more than pairwise relations: entities are often grouped into sets, corresponding to n-adic relationships. Each of these sets can be viewed as a collaboration between entities. Hypergraphs naturally represent n-adic relations. It has been shown that facets of an information space can be modeled by hypergraphs [1]: each facet corresponds to a type of metadata. The different facets are then linked by reference data attached to hyperedges within that face. The step forward is to highlight important information: it is commonly achieved in hypergraphs using random walks [2], [3]. Reference [3] shows that the weighting of vertices at the level of the hyperedges in a hypergraph allows better information retrieval. These two approaches - [2], [3] - mainly focus on vertices; but as hyperedges are linked to references that can be used as pivots in between the different facets [1], [4], it is also interesting to highlight important hyperedges. For instance, in a document database, different metadata can be used to label authors, author keywords, processed keywords, categories, added tags: the pivots between the different facets of this information space correspond to the documents themselves. In the specific case of tags, it can be important to have weights attached to them if the users are allowed to attach tags to documents.

Hyperedge-based weighting of vertices is easier to achieve through multisets: multisets store information on multiplicity of elements. We use multisets family over a set of vertices, called hyper-bag graph - hb-graph for short - as an extension of hypergraphs. Hb-graph multisets play the role of the hyperedges in hypergraph: they are called hb-edges. We want to answer the following research questions: "Can we find a network model and a diffusion process that not only rank vertices but also rank hb-edges in hb-graphs?". We develop an iterative exchange approach in hb-graphs with two-phase steps that allows to extract information not only at the vertex level but also at the hb-edge level.

We validate our approach by using randomly generated hb-graphs. The hb-graph visualisation highlights not only vertices but also hb-edges using the exchange process. We show that the exchange-based diffusion process allows proper coloring of vertices with high connectivity and highlights hb-edges with a normalisation approach - allowing small hb-edges to have a chance to be highlighted.

This paper contributes to present an exchange-based diffusion process that allows not only the ranking of vertices but also of hb-edges. It formalizes exchanges by using hb-graphs that can naturally cope for elements multiplicity. It contributes also to a novel visualisation of this kind of network included in each facet of the information space.

In Section II, the related art is listed and the mathematical background is given in Section III. The construction of the formalisation of the exchange process is presented in Section IV. Results and evaluation are given in Section V and future work and conclusion are addressed in Section VI.

II. Related work

A hypergraph $\mathcal{H} = (V, E)$ on a finite set of vertices (or vertices) $V = \{v_1 \, ; \, v_2; \, ... \, ; \, v_n\}$ is defined in [5] as a family of hyperedges $E = (e_1, e_2, ..., e_p)$ where each hyperedge is a non-empty subset of V and such that $\bigcup_{i=1}^{p} e_i = V$. A hypergraph $\mathcal{H}_w = (V, E, w_e)$ is said edge-weighted if there exists an application $w_e : E \to \mathbb{R}^{+*}$.

In a weighted hypergraph the degree $\deg(v_i)$ of a vertex v_i is defined as:

$$d_i = \deg(v_i) = \sum_{e_k \in E : v_i \in e_k} w_e(e_k).$$

The volume of $S \subseteq V$ is defined as:

$$\text{vol}(S) = \sum_{v_i \in S} \deg(v_i).$$

The incident matrix of a hypergraph is the matrix $H = [h_{kl}]_{\substack{1 \leqslant k \leqslant n \\ 1 \leqslant l \leqslant p}}$ of $M_{n \times p}(\{0\,;\,1\})$, where $h_{kl} = \begin{cases} 1 & \text{if } v_k \in e_l \\ 0 & \text{otherwise} \end{cases}$.

Random walks are largely used to evaluate the importance of vertices. In [2], a random walk on a hypergraph is defined as choosing a hyperedge e_k with a probability proportional to $w_e(e_k)$; in a given hyperedge a vertex is uniformly randomly chosen. The probability transition from a vertex v_i to a vertex v_j is:

$$p(v_i, v_j) = \sum_{k=1}^{p} w_e(e_k) \frac{h_{ik}}{d_i} \times \frac{h_{jk}}{\delta_k},$$

where $\delta_k = \deg(e_k)$ is the degree of a hyperedge defined in [2] as its cardinality. This random walk has a stationary state which is shown to be $\pi_i = \dfrac{d_i}{\text{vol} V}$ for $1 \leqslant i \leqslant n$ [6]. The process differs from the one we propose: our diffusion process is done by successive steps from a random initial vertex on vertices and hyperedges.

Reference [3] defines a random walk to allow the use of weighted hypergraph with weight functions both on hyperedges and on vertices: a vector of weights is built for each vertex making weights of vertices hyperedge-based; a random walk similar to the one above is built that takes into account the weight of the vertices. The evaluation is done using a hypergraph built from a public dataset of computer science conference proceedings; each document is seen as a hyperedge that contains keywords; hyperedges are weighted by citation score and vertices of a hyperedge are weighted with a tf-idf score. Reference [3] shows that a random walk on the (double-) weighted hypergraph enables vertex ranking with higher precision than with unweighted vertices random walk. The process differs again from our proposal: our process not only enables simultaneous alternative updates of vertices and hb-edges values but also allow the ranking of hb-edges. We also introduce a new theoretical framework to achieve this.

Random walks are related to diffusion processes. [7] use random walks in hypergraph to do image matching. [8] builds higher order random walks in hypergraph and constructs a generalised Laplacian attached to graphs generated by their random walks.

Hypergraphs are used in multi-feature indexing to help the retrieval of images [9]. 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 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.

III. MATHEMATICAL BACKGROUND

A. Multisets

Our definitions on multisets are mainly based on [10]. A **multiset** - or **mset** or **bag** - is a pair $A_m = (A, m)$ where A is a set of distinct objects and m is an application from A to $\mathbb{W} \subseteq \mathbb{R}$ or $\mathbb{N}$. A is called the **universe** of the multiset A_m, m is called the **multiplicity function** of the multiset A_m. $A_m^\star = \{x \in A : m(x) \neq 0\}$ is called the **support** of A_m. The elements of the support of an mset are called its **generators**. A multiset where $\mathbb{W} \subseteq \mathbb{N}$ is called a **natural multiset**. The **m-cardinality** of A_m written $\#_m A_m$ is defined as:

$$\#_m A_m = \sum_{x \in A} m(x).$$

Considering $\mathcal{A} = \Omega_{m_\mathcal{A}}$ and $\mathcal{B} = \Omega_{m_\mathcal{B}}$ two msets on the same universe Ω, we define the empty mset, written $\emptyset$ the set of empty support. $\mathcal{A}$ is said to be **included** in $\mathcal{B}$ - written $\mathcal{A} \subseteq \mathcal{B}$ - if for all $x \in \Omega$: $m_\mathcal{A}(x) \leqslant m_\mathcal{B}(x)$. In this case, $\mathcal{A}$ is called a **submset** of $\mathcal{B}$. The power multiset of A, written $\widetilde{\mathcal{P}}(A)$, is the multiset of all submsets of A. Different operations can be defined on multisets of same universe as union, intersection and sum [10].

B. Hb-graphs

Hb-graphs are introduced in [11]. A **hb-graph** is a family of multisets with same universe V and support a subset of V. The msets are called the **hb-edges** and the elements of V the **vertices**. We consider for the remainder of the article a hb-graph $\mathcal{H} = (V, E)$, with $V = \{v_1, ..., v_n\}$ and $E = \{e_1, ..., e_p\}$ the family of its hb-edges.

Each hb-edge $e_i \in E$ has V as universe and a multiplicity function associated to it: $m_{e_i} : V \to \mathbb{W}$ where $\mathbb{W} \subset \mathbb{R}^+$. For a general hb-graph, each hb-edge has to be seen as a weighted system of vertices, where the weights of each vertex are hb-edge dependent.

A hb-graph where the multiplicity range of each hb-edge is a subset of $\mathbb{N}$ is called a **natural hb-graph**. A **hypergraph** is a natural hb-graph where the hb-edges have multiplicity one for every vertex of their support.

The **order** of a hb-graph $\mathcal{H}$ - written $O(\mathcal{H})$ - is:

$$O(\mathcal{H}) = \sum_{j=1}^{n} \max_{e \in E} (m_e(v_j)).$$

The **support hypergraph** of a hb-graph $\mathcal{H} = (V, E)$ is the hypergraph whose vertices are the ones of the hb-graph and whose hyperedges are the support of the hb-edges in a one-to-one way. We write it $\underline{\mathcal{H}} = (V, \underline{E})$, where $\underline{E} = \{e^\star : e \in E\}$.

The **hb-star** of a vertex $x \in V$ is the multiset - written $H(x)$ - defined as:

$$H(x) = \left\{ e_i^{m_{e_i}(x)} \,:\, \forall 1 \leqslant i \leqslant p \colon e_i \in E \wedge x \in e_i^* \right\}.$$

The **m-degree of a vertex** $x \in V$ of a hb-graph $\mathcal{H}$ - written $\deg_m(x) = d_m(x)$ - is defined as:
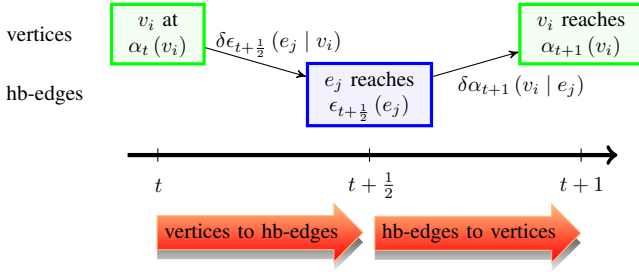
Figure 1. Diffusion by exchange: principle

$$\deg_m(x) = \#_m H(x).$$

The degree of a vertex $x \in V$ of a hb-graph $\mathcal{H}$ - written $\deg(x) = d(x)$ - corresponds to the degree of this vertex in the support hypergraph $\underline{\mathcal{H}}$.

The matrix $H = [m_j(v_i)]_{\substack{1 \leqslant i \leqslant n \\ 1 \leqslant j \leqslant p}}$ is called the **incident matrix** of the hb-graph $\mathcal{H}$.

A **weighted hb-graph** $\mathcal{H}_w = (V, E, w_e)$ is a hb-graph $\mathcal{H} = (V, E)$ where the hb-edges are weighted by $w_e : E \to \mathbb{R}^{+*}$. An unweighted hb-graph is then a weighted hb-graph with $w_e(e) = 1$ for all $e \in E$.

A **strict m-path** $x_0 e_1 x_1 ... e_s x_s$ in a hb-graph from a vertex x to a vertex y is a vertex / hb-edge alternation with hb-edges e_1 to e_s and vertices x_0 to x_s such that $x_0 = x$, $x_s = y$, $x \in e_1$ and $y \in e_s$ and that for all $1 \leqslant i \leqslant s - 1$, $x_i \in e_i \cap e_{i+1}$.

A strict m-path $x_0 e_1 x_1 ... e_s x_s$ in a hb-graph corresponds to a unique path in the hb-graph support hypergraph called the **support path**. In this article we abusively call it a path of the hb-graph. The **length of a path** corresponds to the number of hb-edges it is going through.

Representation of hb-graphs can be achieved either by using sub-mset representation or by using edge representation. In this article we use the extra-vertex representation of the support hypergraph of the hb-graph: an extra-vertex is added for each hb-edge. Each hb-edge is represented by enabling a link in between each vertex of the hb-edge support and the hb-edge extra-vertex.

IV. EXCHANGE-BASED DIFFUSION IN HB-GRAPHS

In diffusion processes and random walks an initial vertex is chosen. The diffusion process leads to homogenising the information over the structure. Random walks in hypergraphs rank vertices by the number of times they are reached and this ranking is related to the structure of the network itself. Several random walks with random choices of the starting vertex can be needed to achieve ranking by averaging.

We consider a weighted hb-graph $\mathcal{H} = (V, E, w_e)$ with $|V| = n$ and $|E| = p$; we write H the incident matrix of the hb-graph.

At time t we set a distribution of values over the vertex set:

$$\alpha_t : \begin{cases} V \to \mathbb{R} \\ v \mapsto \alpha_t(v) \end{cases}.$$

and a distribution of values over the hb-edge set:

$$\epsilon_t : \begin{cases} E \to \mathbb{R} \\ e \mapsto \epsilon_t(e) \end{cases}.$$

$P_{V,t} = (\alpha_t(v_i))_{1 \leqslant i \leqslant n}$ is the row state vector of the vertices at time t and $P_{E,t} = (\epsilon_t(e_j))_{1 \leqslant j \leqslant p}$ is the row state vector of the hb-edges.

We consider an iterative process with two-phase steps as illustrated in Figure 1. At every time step: the first phase starts at time t and ends at $t + \frac{1}{2}$ followed by the second phase between time $t + \frac{1}{2}$ and $t + 1$.

The initialisation sets $\alpha_0(v) = \frac{100}{O(\mathcal{H})}$ for every vertex $v \in V$ and $\epsilon_0(e) = 0$ for every hb-edge $e \in E$.

During the first phase between time t and $t + \frac{1}{2}$, each vertex v_i of the hb-graph shares the value $\alpha_t(v_i)$ it holds at time t with the hb-edges it is connected to.

In an unweighted hb-graph, the fraction of $\alpha_t(v_i)$ given by v_i of m-degree $d_{v_i} = \deg_m(v_i)$ to each hb-edge is $\frac{m_j(v_i)}{\deg_m(v_i)}$, which corresponds to the ratio of multiplicity of the vertex v_i due to the hb-edge e_j over the total m-degree of hb-edges that contains v_i in their support.

In a weighted hb-graph, each hb-edge has a weight $w_e(e_j)$. The value $\alpha_t(v_i)$ of a vertex v_i has to be shared by taking not only the multiplicity of the vertices in the hb-edge but also the weight $w(e_j)$ of a hb-edge e_j into account.

The weights of the hb-edges are stored in a column vector $w_E = (w(e_j))_{1 \leqslant j \leqslant p}^\top$. We also consider the weight diagonal matrix $W_E = \operatorname{diag}(w(e_j))_{1 \leqslant j \leqslant p}$.

We introduce the weighted m-degree matrix:

$$D_{w,V} = \operatorname{diag}(d_{w,v_i})_{1 \leqslant i \leqslant n} = H w_E.$$

where d_{w,v_i} is called the weighted m-degree of the vertex v_i. It is:

$$d_{w,v_i} = \deg_{w,m}(v_i) = \sum_{1 \leqslant j \leqslant p} m_j(v_i) w(e_j).$$

The contribution to the value $\epsilon_{t+\frac{1}{2}}(e_j)$ attached to hb-edge e_j of weight $w(e_j)$ from vertex v_i is:

$$\delta\epsilon_{t+\frac{1}{2}}(e_j \mid v_i) = \frac{m_j(v_i) w(e_j)}{d_{w,m}(v_i)} \alpha_t(v_i).$$

It corresponds to the ratio of weighted multiplicity of the vertex v_i in e_j over the total weighted m-degree of the hb-edges where v_i is in the support.

And the value $\epsilon_{t+\frac{1}{2}}(e_j)$ is calculated by summing over the vertex set:

$$\epsilon_{t+\frac{1}{2}}(e_j) = \sum_{i=1}^n \delta\epsilon_{t+\frac{1}{2}}(e_j \mid v_i).$$

Hence, we obtain:

$$P_{E,t+\frac{1}{2}} = P_{V,t} D_{w,V}^{-1} H W_E \qquad (1)$$

During the second phase that starts at time $t+\dfrac{1}{2}$, the hb-edges share their values between the vertices they hold taking into account the multiplicity of the vertices in the hb-edge. Every value is modulated by the weight $w\left(e_j\right)$ of the hb-edge e_j it comes from.

The contribution to $\alpha_{t+1}\left(v_i\right)$ given by a hb-edge e_j of weight $w\left(e_j\right)$ to the vertex v_i of multiplicity $m_j\left(v_i\right)$ is:

$$\delta\alpha_{t+1}\left(v_i \mid e_j\right) = \frac{m_j\left(v_i\right)w\left(e_j\right)}{\#_m e_j}\epsilon_{t+\frac{1}{2}}\left(e_j\right).$$

The value $\alpha_{t+1}\left(v_i\right)$ is then obtained by:

$$\alpha_{t+1}\left(v_i\right) = \sum_{j=1}^{p}\delta\alpha_{t+1}\left(v_i \mid e_j\right).$$

By writing $d_{m,j} = \#_m e_j$ for $e \in E$ and writing $D_E = \operatorname{diag}\left(d_{m,j}\right)_{1\leqslant j\leqslant p}$ the diagonal matrix of size $p \times p$, it comes:

$$P_{E,t+\frac{1}{2}}D_E^{-1}W_E H^\top = P_{V,t+1}. \tag{2}$$

Gathering 1 and 2:

$$P_{V,t+1} = P_{V,t}D_{w,V}^{-1}HW_E D_E^{-1}W_E H^\top. \tag{3}$$

It is valuable to keep a trace of the intermediate state $P_{E,t+\frac{1}{2}} = P_{V,t}D_{w,V}^{-1}HW_E$ as it records the importance of the hb-edges.

Writing $T = D_{w,V}^{-1}HW_E D_E^{-1}W_E H^\top$, it follows from 3:

$$P_{V,t+1} = P_{V,t}T. \tag{4}$$

V. RESULTS AND EVALUATION

This diffusion by exchange process has been validated with two experiments: the first experiment generates a random hb-graph to validate our approach and the second compares the results to a classical random walk on the hb-graph.

We built a random unweighted hb-graph generator. The generator allows to construct single-component hb-graphs or hb-graphs with multiple connected components. A single connected component is built by choosing the number of intermediate vertices that link the different components to ensure that a single component hb-graph is obtained. We generate $N_{\max}$ vertices. We start by building each component and then interconnect them. Let k be the number of components. A first set V_0 of interconnected vertices is built by choosing N_0 vertices out of the $N_{\max}$. The remaining $N_{\max} - N_0$ vertices are then separated into k groups. In each of these k groups we generate two groups of vertices: a first set $V_{j,1}$ of $N_{j,1}$ vertices and a second set $V_{j,2}$ of N_2 vertices with $N_{j,1} \ll N_{j,2}$, $1 \leqslant j \leqslant k$. The number of hb-edges to be built is adjustable and shared between the different groups. The m-cardinality $\#_m\left(e\right)$ of a hb-edge is chosen randomly below a maximum tunable threshold. The $V_{j,1}$-vertices are considered as important vertices and must be present in a certain amount of hb-edges per group; the number of important vertices in a hb-edge is randomly fixed below a maximum number. The completion of the hb-edge is done by choosing randomly vertices in the $V_{j,2}$ set. The random choice made into

this two groups is tuned to follow a power law distribution: it implies that some vertices occur more often than others. Interconnection between the k components is achieved by choosing vertices in V_0 and inserting them randomly into the hb-edges built.

We apply our diffusion process on these generated hb-graphs: after a few iterations we visualize the hb-graphs to show the evolution of the vertex value with a gradient coloring scale. We also take advantage of the half-step to highlight hb-edges in the background to show important hb-edges with an other gradient coloring scale.

To get proper evaluation and show that vertices with the highest α-values correspond to vertices that are important in the network - in the way they are central for the connectivity - we compute the eccentricity of vertices from a subset S of the vertices V to the remaining $V \backslash S$ of the vertices. Eccentricity of a vertex in a graph is the length of a maximal shortest path between this vertex and the other vertices of the graph: extending this definition to hb-graphs is straightforward. If the graph is disconnected then each vertex has infinite eccentricity.

For the purpose of evaluation, in this article, we define a **relative eccentricity** as the length of a maximal shortest path starting from a given vertex in S and ending with any vertices of $V \backslash S$; the relative eccentricity is calculated for each vertex of S provided that it is connected to vertices of $V \backslash S$; otherwise it is set to $-\infty$.

For the vertices set V, the subset is built by using a threshold value s_V: vertices with α value above this threshold are gathered into a subset $A_V\left(s_V\right)$ of V. We consider $B_V(s) = V \backslash A_V\left(s_V\right)$ the set of vertices with α values below the threshold. We evaluate the relative eccentricity of each vertex of $A_V\left(s_V\right)$ to vertices of $B_V\left(s_V\right)$ in the support hypergraph of the corresponding hb-graph.

Assuming that we stop iterating at time T, we let s_V vary from 0 to the value $\alpha_{T,\max} = \max_{v\in V}\left(\alpha_T\left(v\right)\right)$ - obtained by iterating the algorithm on the hb-graph - by incremental steps and until the eccentricity is kept above 0, first of the two achieved. In order to have a ratio we calculate:

$$r_V = \frac{s_V}{\alpha_{\text{ref}}}$$

where α_{ref} is the reference normalised value, defined as $\alpha_{\text{ref}} = \dfrac{100}{O\left(\mathcal{H}\right)}$ for the hb-graph $\mathcal{H}$. This ratio has values increasing by steps from 0 to $\dfrac{\alpha_{T,\max}}{\alpha_{\text{ref}}}$.

We show the results obtained in Figure 3: we plot two curves. The first plot corresponds to the maximal length of the path between vertices of $A_V\left(s_V\right)$ and vertices of $B_V\left(s_V\right)$ that are connected in function of the value of $r_V = \dfrac{s_V}{\alpha_{\text{ref}}}$: the length of the path corresponds to the half of the length of the path observed in the extra-vertex graph representation of the hb-graph support hypergraph as in between two vertices of V there is an extra-vertex that represent the hb-edge (or the support hyperedge). The second curve plots the percentage of vertices that are in $A_V\left(s_V\right)$ over the vertex set V in function of r_V. When r_V increases the number of elements in $A_V\left(s_V\right)$
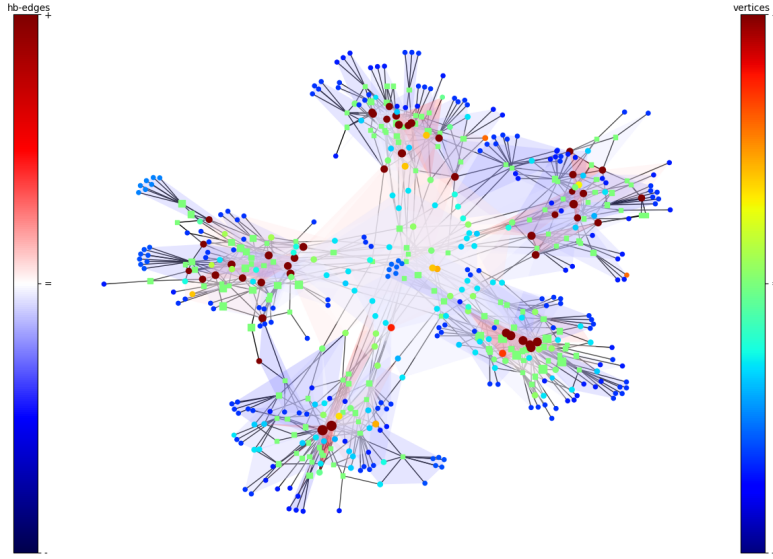
Figure 2. Exchange-based diffusion in hb-graphs: highlighting important hb-edges. Simulation with 548 vertices (chosen randomly out of 10 000) gathered in 5 groups of vertices (with 6, 16, 12, 18 and 2 important vertices and 2 important vertex per hb-edge), 300 hb-edges (with cardinality of support less or equal to 15), 10 vertices in between the 5 groups. Extra-vertices are colored in green and have square shape.
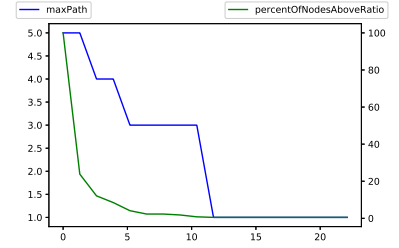

Figure 3. Path maximum length and percentage of vertices in $A_V(s)$ over vertices in V vs ratio r_V.
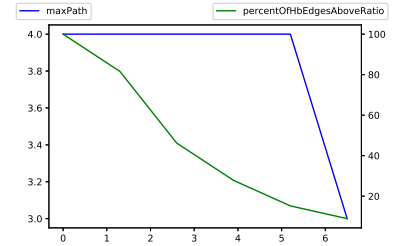
Figure 4. Path maximum length and percentage of vertices in $A_E(s)$ vs ratio.

naturally decreases while they are closer to the elements of $B_V(s_V)$, marking the fact that they are central.

Figure 5 shows that high values of $\alpha_T(v)$ correspond to vertices that are highly connected either by degree or by m-degree. Hence vertices that are in the positive side of the scale color in Figure 2 correspond to highly connected vertices: the closer to red on the right scale they are, the higher the value of $\alpha_T(v)$ is.

Figure 5. Alpha value of vertices at step 5 and (m-)degree of vertices.

A similar approach is taken for the hb-edges: assuming that the diffusion process stops at time T, we use the $\epsilon_{T-\frac{1}{2}}$ function to partition the set of hb-edges into two subsets for a given threshold s_E: $A_E(s_E)$ of the hb-edges that have ϵ values above the threshold and $B_E(s_E)$ the one gathering hb-edges that have ϵ values below s_E.

s_E varies from 0 to $\epsilon_{T-\frac{1}{2},\max} = \max\limits_{e \in E}\left(\epsilon_{T-\frac{1}{2}}(e)\right)$ by incremental steps while keeping the eccentricity is kept above

0, first of the two conditions achieved. In the hb-graph representation, each hb-edge corresponds to an extra-vertex. Each time we evaluate the length of the maximal shortest path linking one vertex of $A_E(s_E)$ to one vertex of $B_E(s_E)$ that are connected in the hb-graph support hypergraph extra-vertex graph representation: the length of the path corresponds to the half of the one obtained from the graph for the same reason than before. In Figure 4 we observe for the hb-edges the same trend than the one observed for vertices: the length of the maximal path between two hb-edges decreases as the ratio r_E increases while the percentage of vertices in $A_E(s_E)$ over V decreases.

Figure 6 shows on the left figure the high correlation between the value of $\epsilon(e)$ and the cardinality of e; the right figure shows that the correlation between value of $\epsilon(e)$ and the m-cardinality of e is even stronger.

The results obtained after five iterations on hb-graphs with different configurations show that we always retrieve the important vertices as the most highlighted. The diffusion by exchange process also highlights additional vertices that were not in the first group but that are at the confluence of different hb-edges. The results on the hb-edges show that the value obtained is highly correlated to the m-cardinality of the hyperedges. To color the hb-edges as it is done in Figure 2 we calculate the ratio $r_{T-\frac{1}{2}}(e) = \dfrac{\epsilon_{T-\frac{1}{2}}(e)}{\epsilon_{\text{norm}}(e)}$, where $\epsilon_{\text{norm}}(e) = \sum\limits_{v \in e^\star} \dfrac{m_e(v)}{\deg_m(v)} v_{\text{ref}}$ corresponds to the value
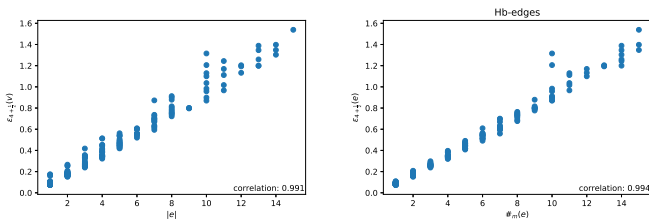
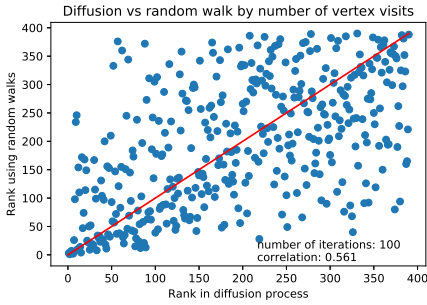Figure 6. Epsilon value of hb-edge at stage $4+\frac{1}{2}$ and (m-)cardinality of hb-edge.

Figure 7. Comparison of the rank obtained by 100 random walks after total discovery of the vertices in the hb-graph and rank obtained in the exchange-based diffusion process.

obtained from the vertices of the hb-edge support by giving to each of them the reference value. Hb-edges are colored using $r_{T-\frac{1}{2}}(e)$, the higher its value, the closer to red the color is: we use the left gradient color bar for it.

We have generated random walks on the hb-graphs with random choice of hb-edges when the walker is on a vertex v with a distribution of probability $\left(\frac{m_i(v)}{\deg_m(v)}\right)_{1\leqslant i\leqslant p}$ and a random choice of the vertex when the walker is on a hb-edge e with a distribution of probability $\left(\frac{m_e(v_i)}{\#_m(e)}\right)_{1\leqslant i\leqslant n}$. We let the possibility of teleportation to an other vertex from a vertex with a tunable value β: $1-\beta$ represents the probability to be teleported. We choose $\beta = 0.85$. We count the number of passage of the walker through each vertex and each hb-edge. We stop the random walk when the hb-graph is fully explored. We iterate N times the random walk, N varying.
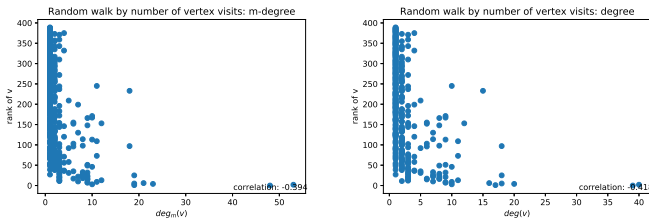

Figure 8. Comparison of the rank obtained by 100 random walks after total discovery of the vertices in the hb-graph and (m-)degree of vertices

Figure 7 shows that after 100 iterations there is weak correlation between the rank obtained by the random walk and our diffusion process. There is no correlation at all with the m-degree of the vertices and the degree of vertices as shown in Figure 8. 100 iterations for the random walk take 6.31 s while it takes 0.009 ms to achieve the 5 iterations needed in the exchange-based approach.

VI. FUTURE WORK AND CONCLUSION

The results obtained by using hb-graph highlight the possibility of using hb-edges for analyzing networks; they confirm that vertices are highlighted due to their connectivity. The highlighting of the hb-edges has been achieved by using the intermediate step of our diffusion process: to achieve it conveniently without having a ranking by hb-edge m-cardinality we normalized it. Different applications can be thought in particular in the search of tagged multimedia documents: sharing of keywords, geographic location, or any valuable information contained in the annotations. Using tagged documents ranking by this means could help in creating summary for visualisation. Our approach is seen as a strong basis to refine the approach of [9].

REFERENCES

[1] X. Ouvrard, J.-M. Le Goff, and S. Marchand-Maillet, "A hypergraph based framework for modelisation and visualisation of high dimension multi-facetted data," *Soon on Arxiv*, 2018.

[2] D. Zhou, J. Huang, and B. Schölkopf, "Learning with hypergraphs: Clustering, classification, and embedding," in *Advances in neural information processing systems*, pp. 1601–1608, 2007.

[3] A. Bellaachia and M. Al-Dhelaan, "Random walks in hypergraph," in *Proceedings of the 2013 International Conference on Applied Mathematics and Computational Methods, Venice Italy*, pp. 187–194, 2013.

[4] M. Dörk, N. H. Riche, G. Ramos, and S. Dumais, "Pivotpaths: Strolling through faceted information spaces," *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 12, pp. 2709–2718, 2012.

[5] C. Berge and E. Minieka, *Graphs and hypergraphs*, vol. 7. North-Holland publishing company Amsterdam, 1973.

[6] A. Ducournau and A. Bretto, "Random walks in directed hypergraphs and application to semi-supervised image segmentation," *Computer Vision and Image Understanding*, vol. 120, pp. 91–102, 2014.

[7] J. Lee, M. Cho, and K. M. Lee, "Hyper-graph matching via reweighted random walks," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pp. 1633–1640, IEEE, 2011.

[8] L. Lu and X. Peng, "High-ordered random walks and generalized laplacians on hypergraphs.," in *WAW*, pp. 14–25, Springer, 2011.

[9] Z. Xu, J. Du, L. Ye, and D. Fan, "Multi-feature indexing for image retrieval based on hypergraph," in *Cloud Computing and Intelligence Systems (CCIS), 2016 4th International Conference on*, pp. 494–500, IEEE, 2016.

[10] D. Singh, A. Ibrahim, T. Yohanna, and J. Singh, "An overview of the applications of multisets," *Novi Sad Journal of Mathematics*, vol. 37, no. 3, pp. 73–92, 2007.

[11] X. Ouvrard, J.-M. L. Goff, and S. Marchand-Maillet, "Adjacency and tensor representation in general hypergraphs. part 2: Multisets, hb-graphs and related e-adjacency tensors," *arXiv preprint arXiv:1805.11952*, 2018.