

# Non-Linear Book Manifolds: Learning from Associations the Dynamic Geometry of Digital Libraries

Richard Nock  
CEREGMIA — Univ.  
Antilles-Guyane  
97259 Schoelcher, France  
rnock@martinique.univ-ag.fr

Frank Nielsen  
SONY Computer Science  
Laboratories, Inc.  
Tokyo 141-0022, Japan  
nielsen@csl.sony.co.jp

Eric Briys  
CEREGMIA — Univ.  
Antilles-Guyane  
97259 Schoelcher, France  
eric.b@cyberlibris.com

## ABSTRACT

Mainstream approaches in the design of virtual libraries basically exploit the same ambient space as their physical twins. Our paper is an attempt to rather capture automatically the actual space on which the books live, and *learn* the virtual library as a non-linear book manifold. This tackles tantalizing questions, chief among which whether modeling should be static and book focused (*e.g.* using bag of words encoding) or dynamic and user focused (*e.g.* relying on what we define as a *bag of readers* encoding). Experiments on a real-world digital library display that the latter encoding is a serious challenger to the former. Our results also show that the geometric layers of the manifold learned bring sizeable advantages for retrieval and visualization purposes. For example, the topological layer of the manifold allows to craft *Manifold* association rules; experiments display that they bring dramatic improvements over conventional association rules built from the discrete topology of book sets. Improvements embrace *each* of the following major standpoints on association rule mining: computational, support, confidence, lift, and leverage standpoint.

## Keywords

Non-linear manifold learning; Pattern mining; Visualization

## 1. INTRODUCTION

Digital libraries are on the brink of Big Data for the man in the street [9]. Scientific analyses at this size level are recent [15], but another challenge is being faced for digital libraries, that will necessitate a leap forward on tools to crunch information, to find and visualize its useful core for the end readers. Ideally, a working digital library should stimulate serendipitous browsing: one finds some book at some place in the digital library, and an even more interesting book is sitting right next to it. In the physical library, centuries of optimization of arrangements have led to *Classifications* like the Library of Congress and Dewey Decimal Classifications (resp. LCC and DCC). These are finely-tuned ways to cope

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM or the author must be honored. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

JCDL'13, July 22–26, 2013, Indianapolis, Indiana, USA.

Copyright 2013 ACM 978-1-4503-2077-1/13/07 ...\$15.00.

with the discrete topology on a set of books  $\mathbb{B}$ , and find the right partition of  $\mathbb{B}$  that fits a set of physical shelves. This is a common-point physical libraries share with their digital twins: digital libraries also adopt a direct standpoint on their content's *ambient space* [3, 7, 18, 20].

In this paper, we claim that perhaps this approach is not optimal when it comes to fit to one's palm — or tablet — a complete digital library. We tackle the question of whether the digital library's organization can be *learned* as a non-linear *submanifold* of its books contents, or as a *submanifold* of the traces that leave readers wandering round the virtual shelves. Our answer, obtained over a real-world experiment for a major international actor of digital libraries, is clear-cut affirmative. It brings new materials to the table of manifold learning algorithms, and a new application of these algorithm to craft association rules. *Manifold association rules* exploit the topological layer of the learned manifold. From both the computational and accuracy standpoints, they are found to dramatically improve upon techniques relying on the first layer of data, the discrete topology of  $\mathbb{B}$ , that are mainstream in association rule mining [12]. Our contributions also include (i) the design and test of an encoding for books, *bag of readers* (BOR), which is shown to be a serious challenger to the popular bag of words encoding (BOW) to capture the library's content; (ii) an improvement of a manifold learning algorithm [25] (iii) the design of information-geometric interfaces to browse through the library's content. We end up with a seamless and scalable integration of manifold-based techniques to digital libraries. Other approaches, pioneered in Latent Semantic Indexing [4], also use a similar algebraic toolbox as ours — singular value decomposition — to come up with a reduced representation of data, and many of them have been used in document or text related works [13]. A crucial difference with our work, however, is that such approaches rely on linear manifolds, like *e.g.* PCA in statistics, and thus fit projections of the ambient space and fall short of the discriminative power of (arbitrary) non-linear manifolds, that are important to capture fine grained structures in the data [23].

The following section defines and explain how to craft book manifolds and manifold association rules. Two experimental sections follows, and a last section concludes with perspectives on non-linear book manifolds.

## 2. DIGITAL LIBRARIES AS NON-LINEAR BOOK MANIFOLDS

Capitalized bold letters like  $\mathbf{M}$  denote matrices, and italicized bold letters like  $\mathbf{v}$  denote vectors.  $m_{ij}$  and  $v_i$  re-

---

**Algorithm 1:**  $\text{learn-}f(\mathbb{B}, d')$ 

---

**Input:**  $\mathbb{B} = \{\mathbf{b}_i, i = 0, 1, \dots, m-1\}$ ,  $d' \in \mathbb{N}_*$ ;

**Output:** Classification function  $f : \mathbb{B} \rightarrow \mathbb{R}^{d'}$ ;

1. Compute  $\mathbf{W} \in \mathbb{R}_+^{m \times m}$  with  $w_{ij} \doteq s(\mathbf{b}_i, \mathbf{b}_j) = w_{ji}$ ;
  2. Normalize  $\mathbf{W}$  to obtain matrix  $\mathbf{N}$  with  $\text{spec}(\mathbf{N}) \subset \mathbb{R}$ ;
  3. Diagonalize:  $\mathbf{N} = \mathbf{P}^{-1}\mathbf{D}\mathbf{P}$ , where  $d_{ii} \geq d_{jj}, \forall i \leq j$ ;
  4. Finish-up:  $f(\mathbf{b}_i) \doteq (p_{i1}, p_{i2}, \dots, p_{id'})$ ,  $\forall \mathbf{b}_i \in \mathbb{B}$ ;
- 

spectively denote coordinate  $(i, j)$  of  $\mathbf{M}$ , and coordinate  $i$  of  $\mathbf{v}$  (coordinates start from zero). Blackboard notations like  $\mathbb{S}$  denote subsets of (tuples of, matrices of) reals, and  $|\mathbb{S}|$  their cardinal. The identity matrix is denoted  $\mathbf{I}$ , and the all-1 vector is denoted  $\mathbf{1}$ . Matrix  $\mathbf{M}$  is (row) stochastic iff  $m_{ij} \geq 0, \forall i, j$  and  $\mathbf{M}\mathbf{1} = \mathbf{1}$ .  $\mathbf{M}$  is doubly stochastic iff both  $\mathbf{M}$  and  $\mathbf{M}^\top$  are stochastic, where “ $\top$ ” denotes transpose. The spectrum of  $\mathbf{M}$  is denoted  $\text{spec}(\mathbf{M})$ .

## 2.1 Learning the digital library

We let  $\mathbb{B} \subset \mathbb{R}^d$  denote a set of books,  $\mathbb{B} = \{\mathbf{b}_i, i = 0, 1, \dots, m-1\}$ , for some respective dimension and size  $d, m \in \mathbb{N}_*$ . Without additional assumption, the *ambient* space of books is a  $d$ -dimensional real-valued *feature space*. Prior to defining this space, we first focus on the way we use it to *learn* the actual geometry of the library. Assume that the library we seek lives on a *non-linear low-dimensional manifold* of  $\mathbb{R}^d$ . We chose to rely mainly on three prominent historical approaches to learn this manifold [14, 19, 25]. We also crafted a novel variant of [25] for the task at hand, which makes overall four tested approaches to learn the library’s geometry. Assume we have a symmetric similarity function between books,  $s : \mathbb{B} \times \mathbb{B} \rightarrow \mathbb{R}_+$ . From this, we can *learn* from book similarities the *classification* function<sup>1</sup>  $f : \mathbb{B} \rightarrow \mathbb{R}^{d'}$  which places the books on the  $d' \ll d$  dimensional manifold we seek. Algorithm 1 ( $\text{learn-}f$ ) presents the four main steps used to learn  $f$ . Its basic ingredient is a *similarity matrix*  $\mathbf{W}$  whose coordinates are the similarities between books.

The four approaches we use to learn the geometry of the library differ on the way the normalization is performed in step 2. of  $\text{learn-}f$ . The two main normalizations are the normalized Laplacian and the Markov chain normalization, respectively denoted [14, 19]:

$$\mathbf{N} \doteq \mathbf{D}'^{-\frac{1}{2}} \mathbf{W} \mathbf{D}'^{-\frac{1}{2}}, \quad (1)$$

$$\mathbf{N} \doteq \mathbf{D}'^{-1} \mathbf{W}, \quad (2)$$

where  $\mathbf{D}'$  is diagonal, with  $d'_{jj} \doteq \sum_k w_{jk} = \sum_k w_{kj}$ . The probabilistic interpretation of (2) in terms of percolation between states of Markov chains is particularly interesting for our task [14]. Assume  $s(\mathbf{b}_i, \mathbf{b}_j)$  is roughly proportional to the probability of being jointly interested by books  $\mathbf{b}_i$  and  $\mathbf{b}_j$ . In this case, books that are brought close by function  $f$  tend to maximize this joint interest, acting somehow in favor of serendipitous browsing (See Section 1).

One can also remark that both (1) and (2) are each the first step — of different algorithms — which iteratively approximate  $\mathbf{W}$  by a doubly stochastic matrix. It was remarked by [21] for (2) and much later for (1) [25]. Since  $\mathbf{W}$  is symmetric, [25] propose to explicitly seek the symmet-

<sup>1</sup>We can think of it as a virtual equivalent of DDC or LCC.

ric doubly stochastic matrix  $\mathbf{N}$  which best approximates  $\mathbf{W}$ , according to the minimization of the Frobenius norm, or the entropy. We abstract this problem as find:

$$\arg \min_{\mathbf{N} \in \mathbb{R}_+^{m \times m}} d_r(\mathbf{N} || \mathbf{W}), \quad \text{s.t. } \mathbf{N} = \mathbf{N}^\top, \mathbf{N}\mathbf{1} = \mathbf{1}, \quad (3)$$

where  $d_r(\mathbf{N} || \mathbf{W}) \doteq \sum_{ij} d_r(n_{ij} || w_{ij})$  and

$$d_r(x || y) = r(x) - r(y) - (x - y)r'(y) \quad (4)$$

is a Bregman divergence with generator  $r$ , a strictly convex differentiable function [2] with  $\text{dom}r \supseteq [0, 1]$ , and  $r'$  is the derivative of  $r$ . Frobenius norm and the entropy used in [25] are particular cases of Bregman divergences [2]. The third algorithm we have used to find  $\mathbf{N}$  is the iterative procedure for the Frobenius norm of [25] to solve (3).

Compared with this approach, (1) and (2) achieve a compromise between improving the doubly stochastic approximation of  $\mathbf{N}$  with respect to  $\mathbf{W}$  while limiting the drift between the eigensystems of  $\mathbf{N}$  and  $\mathbf{W}$ , since a single step of the algorithm is performed. To what extent do these two concurrent objectives play in the accuracy of the final manifold is an interesting question. Our fourth approach adds a preprocessing *scaling* step to (3) in which we replace  $\mathbf{W}$  by a matrix which is the closest to a doubly stochastic matrix under the constraint that its eigenvectors are the *same* as  $\mathbf{W}$ . If the closedness is evaluated using the same divergence as (3), then this amounts to finding  $\mathbf{W}' = u\mathbf{W}$  with:

$$u = \arg \min_{v \in \mathbb{R}_+} D_r(v\mathbf{W}\mathbf{1} || \mathbf{1}). \quad (5)$$

Through closed-forms solutions to (5) are not always available, their fast approximation is trivial, as explained below.

LEMMA 1.  $u \in [(\max_i \mathbf{1}_i^\top \mathbf{W}\mathbf{1})^{-1}, (\min_i \mathbf{1}_i^\top \mathbf{W}\mathbf{1})^{-1}]$ , with  $\mathbf{1}_i$  the all-0 vector with a 1 in coordinate  $i$ .

(proof omitted) Hence, a fast dichotomic search is enough to approximate  $u$ . In the case where we consider the Frobenius norm ( $r(x) = x^2$ ), the solution to (5) is explicit as we have  $u = \mathbf{1}^\top \mathbf{W}\mathbf{1} / \mathbf{1}^\top \mathbf{W}^2 \mathbf{1}$ . Let us summarize the four different normalizations performed in step 2. of  $\text{learn-}f$ : (NL) normalized Laplacian (1); (MC) Markov chain normalization (2); (DS) doubly-stochastic approximation (3) with Frobenius norm; (SDS) doubly stochastic approximation preceded by scaling. We have also tested, for each of these normalizations, an additional preprocessing step to step 2. which consists in replacing  $\mathbf{W}$  by a sparse symmetric matrix, as follows (for some  $k \in \mathbb{N}_*$ ):  $w_{ij} = 1$  iff  $\mathbf{b}_i \in k\text{-SNN}(\mathbf{b}_j)$  and 0 otherwise.  $k\text{-SNN}(\mathbf{b}_j)$  is the set of  $k$  *symmetric* nearest neighbors of book  $\mathbf{b}_j$  [17]. The sparsified matrix is a truncated version of the geodesic distance matrix of ISOMAP [23], but in which  $k$ -NNs are replaced by  $k$ -SNNs, guaranteeing to keep the symmetry.

## 2.2 Books and similarities

### 2.2.1 Books: bag of words vs bag of readers

The first solution to code the ambient space of books is to consider the books’ content, and more precisely treat books as *bag of words* (BoW) [10]. Experimentally, rather than parsing words from the books’ contents — which would make a very large  $d$  —, we use the books’ XML descriptions of their titles and abstracts. Vectors are post-processed in the usual way [10], with TF-IDF weighting and normalization to

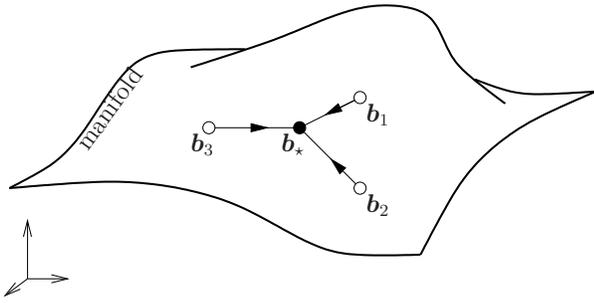


Figure 1: A bookshelf  $\{b_*, b_1, b_2, b_3\}$  obtained after affinity propagation. The star shows that book  $b_*$  is the exemplar of the bookshelf: it may be thought as its most accurate representative (see text).

unit norm. There is however another way to represent books, which consists in taking the binary Users $\times$ Books matrix that would be used to describe the users as a function of the books with which they have interacted<sup>2</sup>, and then transpose it to reveal a description of books as a simple function of the users who have interacted with the books, considering that similar books should be the subject of joint interactions for many readers. This is what we call a *bag of readers* (BoR) encoding. Though this encoding is accessible only once a sufficiently large number of interactions have occurred, it is perhaps more appealing than BoW, as it is dynamic and evolves with the readers’ habits, and it makes also possible to reveal associations between books that would have been missed from the books’ contents alone.

### 2.2.2 Book similarities

We have considered three kinds of similarities, each of which falls in  $[0, 1]$ , “1” being the maximal similarity always achieved for  $s(\mathbf{b}_i, \mathbf{b}_j)$ . The first one, derived from the Heat kernel, belongs to the most popular in manifold learning [1]:

$$s_{\text{hea}}(\mathbf{b}_i, \mathbf{b}_j) = \exp(-\|\mathbf{b}_i - \mathbf{b}_j\|_2^2 / T), T > 0. \quad (6)$$

The second one is built from the cosine similarity:

$$s_{\text{cos}}(\mathbf{b}_i, \mathbf{b}_j) = \frac{1}{2} \left( 1 + \frac{\mathbf{b}_i^\top \mathbf{b}_j}{\|\mathbf{b}_i\|_2 \|\mathbf{b}_j\|_2} \right). \quad (7)$$

The last one is the Jaccard index, that we note by a minor abuse of notation as:

$$s_{\text{jac}}(\mathbf{b}_i, \mathbf{b}_j) = \frac{\mathbb{1}(\mathbf{b}_i)^\top \mathbb{1}(\mathbf{b}_j)}{\|\mathbb{1}(\mathbf{b}_i + \mathbf{b}_j)\|_2^2}. \quad (8)$$

Here,  $\mathbb{1} : \mathbb{R}^d \rightarrow \{0, 1\}^d$  denotes the Heaviside function which replaces  $> 0$  coordinates by 1.

## 2.3 Bookshelves

So far, we have described the way we shape the digital library as a submanifold of  $\mathbb{R}^d$ . The next step is the building of virtual *bookshelves*, that is, clusters of books. There exists an important literature on clustering in the context of digital libraries [11]. A natural approach to crafting bookshelves consists in performing a (hard) clustering of the manifold,

<sup>2</sup>In our experiments, users may not just read the books online: they can also place them on public or private bookshelves, or buy physical copies. Hence, “interacted” is more suited than simply “read”.

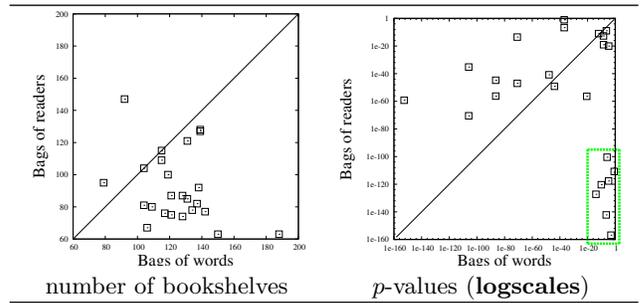
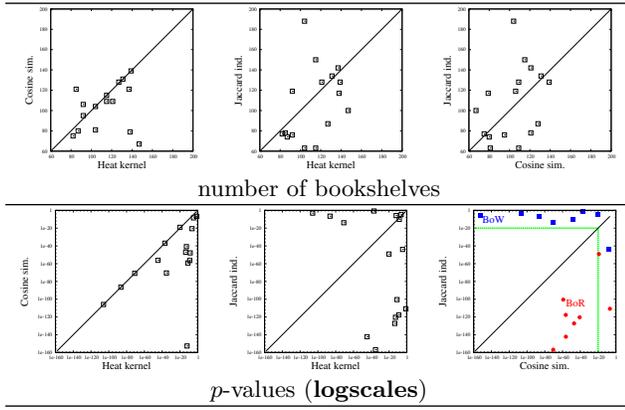


Figure 2: Scatterplot comparing BoW encoding versus BoR encoding, according to the number of bookshelves (left) and  $\chi^2$   $p$ -values (right, see text for details about the implementation). On both plots, the diagonal is the  $y = x$  line. Points below the  $y = x$  line represent smaller numbers of bookshelves for BoR (left) and smaller (hence, better)  $p$ -values for BoR (right). The green thick dashed rectangle (right plot) on the lower-right distinguishes a region for which (i)  $\log(p_{\text{BoW}}/p_{\text{BoR}}) \in [80, 140]$  — *i.e.* BoR is much more accurate than BoW, and (ii)  $p$ -values for BoW are “close” to 1 relatively to the other  $p$ -values.

under the constraint that the number of bookshelves is sufficiently (but not too) large. Because of this constraint, algorithms that are known to be subject to trapping in local minima — like  $k$ -means — gave poor results. Even improvements that guarantee on average to approach global minima [16] gave better but still unsatisfactory results, and were outperformed by an algorithm which comes in the nick of time for such an application: affinity propagation [6]. Affinity propagation is a message passing algorithm which ends with a hard clustering solution whose number of clusters is not fixed in advance; it depends on a user-fixed preference parameter  $P$ . Affinity propagation brings, from its ancestry in median-based clustering methods a rare plus: each bookshelf contains an *exemplar* [6], that is, a book which is supposed to embody the best all books in the bookshelf. Figure 1 provides a schematic view of what are exemplars and bookshelves. One can provide a high-level description of the algorithm as a process alternating between exchanges of two types of messages between books until convergence. In a first round, each book  $\mathbf{b}_i$  sends a real-valued *responsibility* message to each book  $\mathbf{b}_k$ , telling to which extent  $\mathbf{b}_k$  may serve as an exemplar for  $\mathbf{b}_i$ . Then, Each  $\mathbf{b}_k$  sends a real-valued *availability* message to each  $\mathbf{b}_i$ , telling to which extent  $\mathbf{b}_i$  may choose  $\mathbf{b}_k$  as exemplar. Before a new round of message exchanges begins, each book elects its exemplar on the basis of the current availabilities and responsibilities [6]. Each bookshelf is thus a set of books having chosen the same exemplar.

## 2.4 MARs: Manifold association rules

Most of the theoretical works on association rule mining work on the discrete topology on  $\mathbb{B}$ . This is natural: itemsets are the basic building blocks of association rules, and they are also the open sets of the discrete topology. This is, however, everything but efficient from the computational standpoint, and finding the best rules resembles finding a needle in a haystack as  $\mathbb{B}$  gets larger [8, 24]. The library’s



**Figure 3:** Scatterplots comparing the results of the three similarities tested in (6), (7) and (8). The top row displays the number of bookshelves, and the bottom row the  $p$ -values. Conventions follow Figure 2. On the lower right plot, the green thick dashed lines delineate a region in which at least one of the  $p$ -values is “large” ( $\geq 10^{-20}$ ); this scatterplot also depicts the results of BoW encoding (blue squares) and BoR encoding (red circles, see text for details).

manifold provides us with a topological layer which comes in handy to reduce the number of candidates.

An *association rule* is a logical rule over the discrete topology of  $\mathbb{B}$ , which can be written  $\mathbb{B}_l \Rightarrow \mathbb{B}_r$ , where  $\mathbb{B}_l \cup \mathbb{B}_r \subseteq \mathbb{B}, \mathbb{B}_l \cap \mathbb{B}_r = \emptyset$  [22]. In the context of the digital library, it can be interpreted as: if a reader has interacted with the books of  $\mathbb{B}_l$ , then he has also interacted — or will likely interact — with the books from  $\mathbb{B}_r$ . Set  $\mathbb{B}_l \cup \mathbb{B}_r$  is called the *itemset* of the association rule. We have chosen to rely on four of the most popular criteria to evaluate association rules: the *support* of the rule  $\mathbb{B}_l \Rightarrow \mathbb{B}_r$  is the support of its itemset, noted  $\text{SUPP}(\mathbb{B}_l \cup \mathbb{B}_r)$ , equal to the proportion of readers having interacted with all the corresponding books. The *confidence* of the association rule is:

$$\text{CONF}(\mathbb{B}_l \Rightarrow \mathbb{B}_r) \doteq \frac{\text{SUPP}(\mathbb{B}_l \cup \mathbb{B}_r)}{\text{SUPP}(\mathbb{B}_l)}. \quad (9)$$

The *lift* of the association rule is:

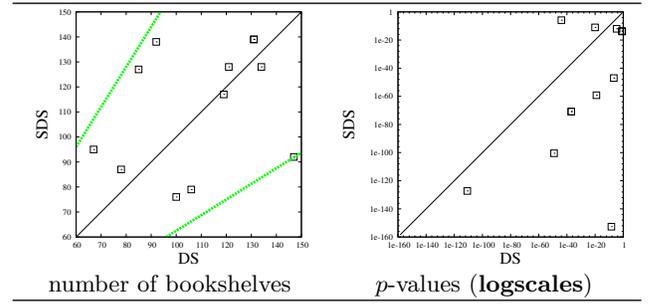
$$\text{LIFT}(\mathbb{B}_l \Rightarrow \mathbb{B}_r) \doteq \frac{\text{SUPP}(\mathbb{B}_l \cup \mathbb{B}_r)}{\text{SUPP}(\mathbb{B}_l)\text{SUPP}(\mathbb{B}_r)}. \quad (10)$$

Finally, the *leverage* is defined as:

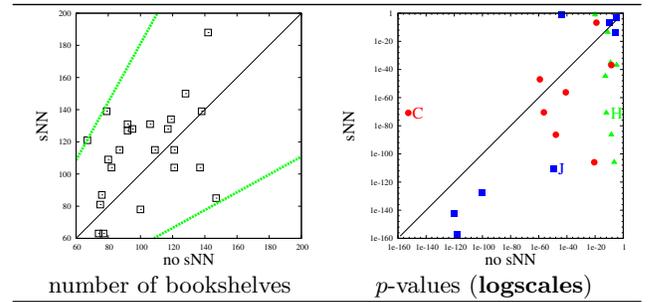
$$\text{LEVE}(\mathbb{B}_l \Rightarrow \mathbb{B}_r) \doteq \text{SUPP}(\mathbb{B}_l \cup \mathbb{B}_r) - \text{SUPP}(\mathbb{B}_l)\text{SUPP}(\mathbb{B}_r). \quad (11)$$

Remark that the support, lift and leverage are invariant to the permutation of the left-hand sides and right-hand sides of an association rule.

We let *Manifold association rules* (MARs) denote association rules built from the topological layer of the book manifold itself, and not directly from the discrete topology of  $\mathbb{B}$ . Typically, this means building itemsets out of the geometric proximities between books on the manifold. We give a simple and efficient example of MARs: the candidate itemsets are restricted to be subsets of the bookshelves, and thus subsets of the clusters found by affinity propagation. In Figure 1, bookshelf  $\{\mathbf{b}_x, \mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3\}$  would thus generate 6



**Figure 4:** Scatterplots comparing the results of DS and our scaled modification SDS, according to the number of bookshelves (left) and the  $\chi^2$   $p$ -values (right). Conventions follow Figure 2; on the left plot, the green thick dashed lines are lines of equation  $y = (147/92)x, y = (92/147)x$  (see text for details).



**Figure 6:** Scatterplots evaluating the influence of sparsifying  $\mathbf{W}$  with symmetric nearest neighbors, according to the number of bookshelves (left) and  $p$ -values (right). Conventions follow Figure 2; on the left plot, the green thick dashed lines are the lines of equation  $y = (121/67)x, y = (67/121)x$ ; the right plot distinguishes the results of the different similarity functions: green triangles are the (H)eat kernel (6), red circles are the (C)osine similarity (7) and blue squares are the (J)accard index (8) (see text).

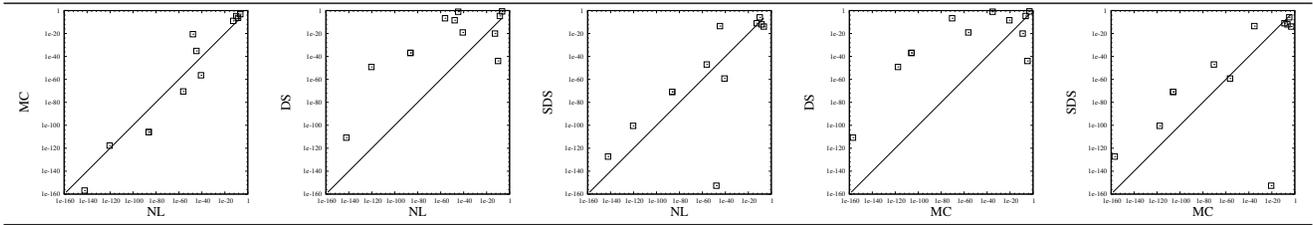
candidate itemsets of size 2. To save space, we shall work in this paper only, but extensively, on association rules of size two, with a single book on the left- and right-hand sides.

This size constraint leaves us space to study a very interesting particular case of MARs. Because affinity propagation makes a clearcut distinction between the exemplar of a bookshelf and its other books, it is interesting to define MARs in which either the left- or the right-hand side of the association rule is an exemplar. For example,  $\mathbf{b}_x \Rightarrow \mathbf{b}_1$  is an example of such a rule from Figure 1;  $\mathbf{b}_2 \Rightarrow \mathbf{b}_x$  is another example. We denote the subset of MARs in which  $\mathbf{b}_x$  is in the left-hand side as F-MARs, for centriFuge-MARs, while those in which  $\mathbf{b}_x$  is in the right-hand side shall be denoted P-MARs, for centriPete-MARs. For example, the bookshelf of Figure 1 would yield 12 MARs, 3 F-MARs and 3 P-MARs.

### 3. EXPERIMENTS

#### Domain.

Experiments related in this Section have been pursued during the deployment of a digital library, built for individ-



**Figure 5: Logscale scatterplots comparing  $p$ -values for the four different types of normalizations (Subsection 2.1). Conventions follow Figure 2 (see text for details).**

ual readers by Europe’s leader on digital books rental<sup>3</sup> over a year in 2010-2011. We recorded the complete behavior of the first seven thousands readers of the digital library during the first year of their membership. Any such user can perform three types of actions on books: (i) read books online, (ii) place books of his/her choice on a virtual bookshelf which can be shown to other users, (iii) buy physical copies of the books of his/her choice. We filtered out the books on which no user had interacted, to ensure that BoR encoding was meaningful; the resulting set of books (more than eight thousands) was also used for BoW encoding for the purpose of manifold comparisons.

The remaining parameters still not fixed in Section 2 have been chosen as follows:  $k = 5$  in  $k$ -SNN,  $T = 1000$  in (6),  $P$  in affinity propagation was fixed to be the 90<sup>th</sup> percentile of similarities, and finally  $d' = 3$  for visualization purposes.

### Evaluation metrics to rank manifolds.

We have computed and visualized all manifolds for every possible choice of parameters described in Section 2. This represents 4 (normalization)  $\times$  2 (sparsification with -SNN)  $\times$  2 (books encoding)  $\times$  3 (similarities) = 48 manifolds, and as many different digital libraries. This task is unsupervised, yet there is a possibility to compare between each others the manifolds obtained, *via* the bookshelves computed. Intuitively the set of bookshelves should be sufficiently correlated with the high-leveled topics of the books, and it turns out that each book of the digital library is classified according to one of six primary topics (cooking, home, family, money, leisure, health) — obviously, informations about topics are not used to learn whichever of the manifolds. Out of the set of  $b$  bookshelves obtained, we compute a  $6 \times b$  contingency table and compute the  $p$ -value of a  $\chi^2_{5(b-1)}$  independence test for the observed contingency. The primary goal of  $p$ -values is not to make inference, but rather rank manifolds: manifolds are indeed compared on the basis of the log-ratio of their associated  $p$ -values. Still, inference would be possible on each manifold with the  $\chi^2$  test of independence, the minute  $p$ -values obtained in general and the small number of bookshelves observed (given the number of books used) raising little possibilities of type II errors.

### 3.1 Books encoding: BoR challenges BoW

Figure 2 summarizes the  $p$ -values that were obtained for the two different encodings of books. Since primary topics are heavily correlated to the books’ titles and abstracts, one would expect BoW encoding to beat hands down BoR encoding. Figure 2 displays that it is not the case, and it is even the opposite which happens sometimes, as witnessed

<sup>3</sup><http://www.cyberlibris.com>

by the green dashed rectangle of dots, for which BoW performs quite “poorly” ( $p \approx .2$  for the rightest point), while  $p$ -values of BoR are all of minute order ( $\leq 10^{-100}$ ). Thus, choosing the dynamic information contained in the user’s traces in lieu of the static information of the books contents yields in some cases dramatic improvements in the library’s manifold. The number of bookshelves is also significantly smaller for BoR encoding (sign test  $p$ -value  $\approx 10^{-12}$ ), but the ratio with BoW encoding is almost always in the interval  $[.5, 2]$ , which denotes a remarkable stability in the results of affinity propagation.

### 3.2 Books similarity: Cosine and Jaccard similarities perform the best

Comparing the three similarities in terms of number of bookshelves gives no clear-cut pattern as to whether one similarity would bring different results than the others, as evidenced by the top row of Figure 3. When it comes to comparing  $p$ -values, the observations are much different. Indeed, the heat kernel similarity performs extremely poorly compared with the cosine similarity — a single point is (slightly) above the  $y = x$  line —, and quite poorly, even when it is less noticeable, compared with Jaccard index. It is important to emphasize the fact that these poor results should not stem from a bad choice of parameter  $T$  of the heat kernel, as it was tuned to get the best results.

We believe that there is a technical rationale to this phenomenon — which represents good news for manifold association rules. Both BoW and BoR encoding, with TF-IDF weighting and normalization, approximate maximum likelihood fitting, incorporating knowledge about all books into the modeling of each of them [5]. The normalizations of  $\mathbf{W}$  are also related to a probabilistic modeling of transitions between books. It turns out that (7) and (8) still follow this probabilistic modeling, while (6) does not. It is quite obvious for Jaccard index (8), which, in the BoR encoding, estimates the probability that a user interacts with two books:

$$s_{\text{jac}}(\mathbf{b}_i, \mathbf{b}_j) \doteq \hat{\text{Pr}}[\mathbf{b}_i \cap \mathbf{b}_j]. \quad (12)$$

The cosine similarity (7), on the other hand, relies on a Laplacian normalization of joint marginals. Whenever  $\mathbf{b}_i, \mathbf{b}_j \in \{0, 1\}^d$ , it is indeed not hard to see that up to an additive constant,

$$s_{\text{cos}}(\mathbf{b}_i, \mathbf{b}_j) \propto \hat{\text{Pr}}[\mathbf{b}_i \cap \mathbf{b}_j] \left( \hat{\text{Pr}}[\mathbf{b}_i] \hat{\text{Pr}}[\mathbf{b}_j] \right)^{-\frac{1}{2}}.$$

Figure 3 (lower-right plot) also displays that the accuracy of these two similarities heavily relies on book encodings: while cosine similarity works better with BoW encoding, Jaccard index favors BoR encoding.

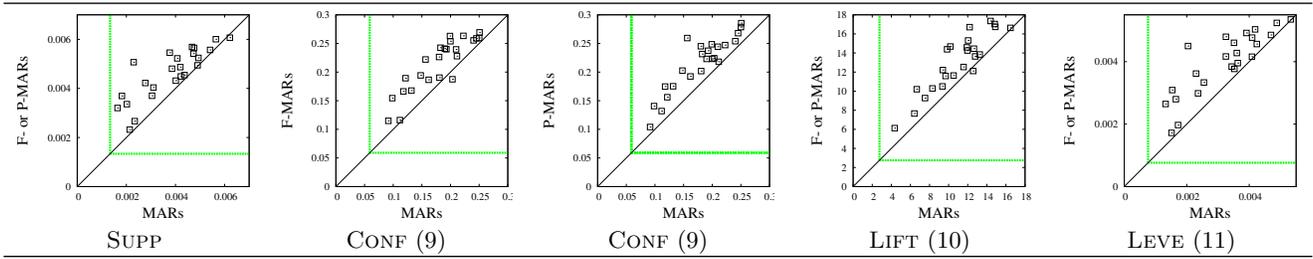


Figure 7: Scatterplots comparing the results of MARs vs P- and/or F-MARs, from the support, confidence, lift and leverage standpoints (from left to right). Each point is the average (over all manifolds) of the parameters computed over *all* manifold association rules obtained. General conventions follow Figure 2. On each scatterplot, the green thick dashed lines represents the average performances of size-two association rules (See text for details); in the squares delineated by these green lines, any point denotes manifold association rules performing on average *better* than the whole set of association rules.

### 3.3 Doubly stochastic approximation: SDS beats DS

Recall that DS may be understood as a generalization of the Laplacian and Markov chain normalizations, in which rather than making a single iteration towards doubly stochastic approximation, one performs all necessary iterations [21, 25]. SDS is a variant of DS in which DS is given a similarity matrix already “close” to a doubly stochastic matrix *while* keeping the same eigenvectors as the initial similarity matrix  $\mathbf{W}$ . DS and SDS were compared on our domain, and the results obtained, both in terms of the number of bookshelves and  $p$ -values, are displayed in Figure 4. Conclusions from both scatterplots are quite easy to draw: first, none of the algorithms yields a significant increase (or decrease) in the number of bookshelves: ratios do not exceed  $147/92 \approx 1.6$  (Figure 4). Second,  $p$ -values tend to display that SDS performs better than DS: all but two points are below the  $y = x$  line (sign test  $p$ -value  $\approx 0.016$ ), with log ratios that can be very large:  $\log(p_{\text{DS}}/p_{\text{SDS}}) \approx 144$  for the lower right point.

### 3.4 Manifold normalization: SDS is competitive against NL and MC

We have compared the four different techniques to normalize  $\mathbf{W}$  in `learn-f`. Figure 5 presents the scatterplots obtained. The first conclusion that can be drawn is that DS performs poorly with respect to all other approaches. MC, SDS and NL produce results that are quite equivalent. The slight advantage for NL and MC observed from the points that are slightly above the  $y = x$  line (on scatterplots 3 and 5, starting from the left) is dampened by the significant outlier on which SDS clearly beats the two other approaches, with  $\log(p_{\text{NL}}/p_{\text{SDS}}) > 100$  and even  $\log(p_{\text{MC}}/p_{\text{SDS}}) > 130$ . This outlier is however important as it corresponds to mixing popular settings to learn the manifold: BoW encoding, cosine similarity (7) and no sparsification of  $\mathbf{W}$ . This might indicate that SDS may be a valuable candidate to compete against standard normalization techniques on other domains as well. This may also indicate that the normalization step in manifold learning has to achieve a compromise between finding a doubly stochastic approximation of  $\mathbf{W}$ , while staying close to the original  $\mathbf{W}$  from the spectral standpoint.

### 3.5 Sparsifying $\mathbf{W}$ may improve the results

An important question in the context of the digital library is the way sparsifying  $\mathbf{W}$  works in manifold learning, and

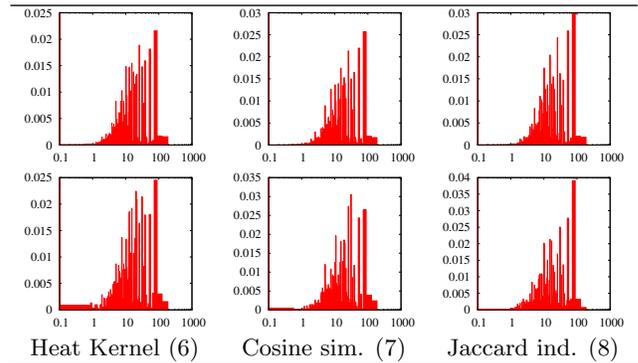
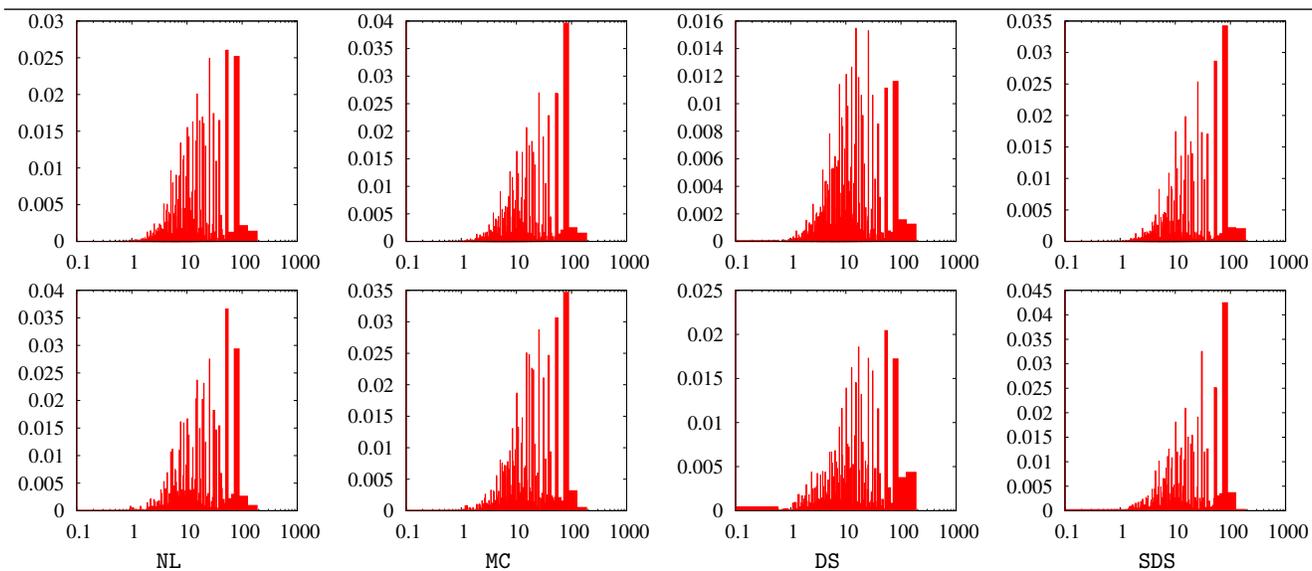


Figure 9: Truncated distributions of non-zero lifts for MARs (upper row) and F- or P-MARs (bottom row), as a function of the similarity  $s(\cdot, \cdot)$  (Subsection 2.2).  $x$ -scales are *logscales* (see text for details).

whether it allows to keep or improve the results. This question is important from two standpoints, the first of which being the computational complexity standpoint: a sparse  $\mathbf{W}$  opens possibilities of faster algorithms to diagonalize  $\mathbf{N}$ ; the second standpoint is robustness: if sparsification works, it indicates that the digital library’s organization does not degrade with local information loss. We have thus compared the manifolds obtained with and without sparsification by  $k$ -SNN. Fixing  $k = 5$  results in a harsh sparsification: in our case, since  $2kn$  coordinates of  $\mathbf{W}$  remain non-zero, we end up with less than 1% non-zero values on the sparsified  $\mathbf{W}$ . Figure 6 displays the way sparsification influences manifold learning. Sparsification has a small, but significant, qualitative influence, as it tends to increase the number of bookshelves (sign test  $p$ -value  $\approx 0.02$ ). From the quantitative standpoint, the increase is in fact quite small as seen from the two lines in between which the data is sandwiched (See Figure 6), and from the barycenter of data, accounting for only roughly 10% increase in average.  $p$ -values display that sparsification tends to improve the results (sign test  $p$ -value  $\approx 0.02$ ), but its positive influence is the most dramatic when  $\mathbf{W}$  is the least sparse: using the heat kernel as similarity (6) brings a similarity matrix with no zero coordinate. Without sparsification,  $p$ -values are always  $\geq 10^{-20}$ . Sparsification helps to decrease  $p$  to minute values  $< 10^{-100}$ . The impact of sparsification may explain that DS performs poorly



**Figure 8:** Truncated distributions of non-zero lifts for MARs (upper row) and F- or P-MARs (bottom row), as a function of the normalization of  $\mathbf{W}$  (Subsection 2.1).  $x$ -scales are *logscales* (see text for details).

in part because it tends to de-sparsify  $\mathbf{W}$ . Among all runs, the two poorest results ( $p$ -values = 0.09, 0.19) are obtained when running DS after having sparsified  $\mathbf{W}$ . In these particular cases, a *single* iteration of DS yields a matrix already without zeroes, as this matrix is  $\mathbf{W} + (1/n)\mathbf{1}\mathbf{1}^\top$  [25].

### 3.6 Manifold association rules always improve upon association rules

#### General patterns on MARs, F- and P-MARs.

We have first compared the average parameters (support, confidence, lift, leverage) of the sets of MARs for each possible way of building the manifold with BOR encoding. This represents 24 ways of building the manifold, on each of which we computed the averages of the four parameters — support, confidence, lift, leverage — for the three different types of MARs: general MARs, F-MARs and P-MARs. For each type of MARs, we generated *every* association rule following the setting of Subsection 2.4, thus without minimal support requirement. We compared the results obtained with Apriori generating *all* association rules with two books — thus, still without minimal support constraint. Manifold association rules obtained are thus a subset of the association rules output by Apriori, and our first objective is to assess the quality of the subsets with respect to the whole set.

The results are displayed on Figure 7. All comparisons are clear-cut: regardless of the combination of parameters to build the manifold, regardless of the criterion used to compare the sets of association rules obtained (average support, confidence, lift or leverage), manifold association rules *systematically* improve upon the whole set of association rules on average, and most of the times by orders of magnitude. This implies that the manifold geometrically captures the most prominent associations between books. This is not surprising: learning the manifold relies on basic building blocks that call to a probabilistic modeling of data: book similarities (Subsection 2.2), normalization of  $\mathbf{W}$  (Subsection 2.1), etc. . Since the criteria used to evaluate association rules

have probabilistic backgrounds (9—11), it is not surprising that the best association rules “pop out” on the manifold.

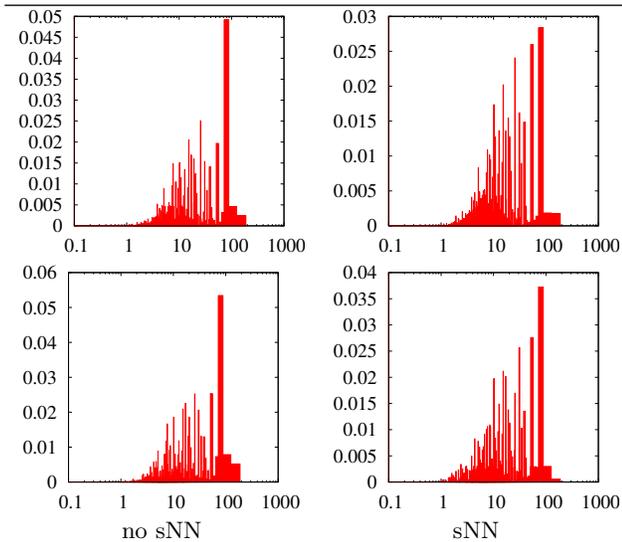
#### Improvements on F- and P-MARs.

There is more: drilling down into MARs displays that affinity propagation succeeds in pinpointing important books as exemplars. For example, the average support of F- or P-MARs is greater than MARs’ for all but one manifolds (Figure 7, left plot). Both F-MARs and P-MARs also improve confidences compared to MARs, but there is no difference on average between F-MARs and P-MARs from the confidence standpoint (scatterplot not shown to save space). Results from the lift and leverage criteria confirm the conclusion that MARs select very accurate subsets of association rules, an improvement which is even better for F- and P-MARs.

These conclusions are to be read in the light of the number of association rules selected: the number of MARs ranged from 4.8% of the total association rules to 12.9% of the total, and the numbers of F- or P-MARs were remarkably stable, ranging from 0.22% to 0.26% of the total association rules *regardless of the manifold*. Once again, this stability of affinity propagation is remarkable.

#### Criteria at a glance: the case of the lifts.

We have also drilled down into the influences of the manifold construction’s parameters on association rules, to see to what extent the choices made upstream on the manifold construction also impact downstream, on the association rules built from the manifold itself. We focus on the lifts of the association rules, and first consider the influence of the normalizations of  $\mathbf{W}$ , as displayed in Figure 8 for the lifts’ distributions obtained. To keep the best viewing of plots, we filtered out MARs whose left- or right-hand side has zero support, showing only the truncated part of distributions for MARs with non zero lift. Notice that for more than 95% of the curves, the remaining truncated part still represented a mass  $> .5$ , sometimes even exceeding  $.75$ . Three notable facts emerge: (i) DS performs the worst among all four nor-



**Figure 10: Truncated distributions of non-zero lifts for MARs (upper row) and F- or P-MARs (bottom row), as a function of the sparsification of  $\mathbf{W}$  (Subsection 2.1).  $x$ -scales are *logscales* (see text).**

malizations; (ii) the strictly positive lifts for MC (MARs, F- or P-MARs) and SDS (MARs) are, for more than 99.9% of the association rules, *strictly greater than one* — once again, this accounts for the fact that the manifold is built around the most prominent dependences between books; (iii) MC and SDS peak on MARs with very large lifts ( $\geq 100$ ), and seem to be the best normalizations from this standpoint. To summarize, the choice of the normalization of  $\mathbf{W}$  impacts quite similarly on the quality of the association rules (from the lift standpoint), and on the quality of the book manifold.

Drilling down into the global influence of the similarity function on the lift of the association rules somehow confirms an intuition, namely that the choice of Jaccard index provides the best lifts on whichever of MARs, F-MARs or P-MARs, as displayed in Figure 9 (right column). In Subsection 3.2, we emphasized the quality of the manifold associating BOR encoding and Jaccard index. Figure 9 somehow confirms the positive impact downstream of this combination, but the intuition also suggests that the positive impact may be boosted by the fact that Jaccard index already encodes in  $\mathbf{W}$  a quality (support) of association rules with two books (12). We also remark that Cosine similarity still wins — but this time by a whisker — over the Heat kernel. This, again, goes hand in hand with remarks in Subsection 3.2 according to which the Cosine similarity is seemingly a better associate to BoW encoding.

Last, but not least, Figure 10 depicts the influence of sparsification on the lifts obtained for MARs, F- and P-MARs. While all curves peak on association rules with large lifts, the plots clearly display that the peaks are more pronounced when  $\mathbf{W}$  is not sparsified. The plots display that the difference is much more pronounced for MARs: lifts  $\approx 100$  peak round 5% without sparsification, while they peak approximately at 2.8% with sparsification. This general picture hides in fact significant discrepancies appearing as one drills down into the influences of the choice of the similarity function and the normalization of  $\mathbf{W}$ . Figure 11 reveals the

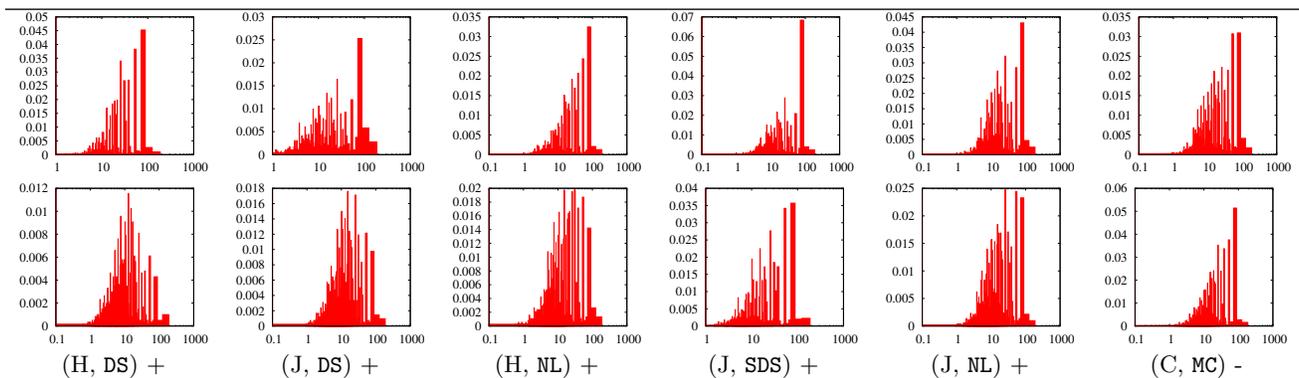
most significant differences between this “reference peak” (lift  $\approx 100$ ) for MARs taken without and with sparsification of  $\mathbf{W}$ . The first pattern is that DS normalization is in fact the worst setting for sparsifying  $\mathbf{W}$ . This parallels the results of Subsection 3.5, according to which sparsification before using DS already produces the worst manifolds, and should thus reasonably produce association rules of lower quality as well. We also remark that the normalized Laplacian NL provides pattern which is similar in all but one aspect: sparsifying  $\mathbf{W}$  still produces good results according to the reference peak. Hence, the loss of information due to sparsification does not prevent to obtain good association rules. When we look at the best results obtained, including results not shown to save space, the most dramatic density is the one shown in Figure 11 for Jaccard index, SDS normalization and no sparsification of  $\mathbf{W}$ . The peak of lifts  $\approx 100$  is the tallest (it peaks at  $\approx 7\%$ ) of all experiments, by far. Once again, this sounds like a confirmation that selecting good parameters to build the manifold also impact positively on the selection of manifold association rules: in particular, recall that Jaccard index favors BoR encoding (Subsection 3.2). We also think that this result may be interpreted as the matrix  $\mathbf{W}$  should be sparsified, but not too much to produce good results from the manifold association rules standpoint. Indeed, Jaccard index already produces a sparse matrix  $\mathbf{W}$  given that few users had more than a dozen books in their traces, from which it comes that roughly 3% of the matrix  $\mathbf{W}$  is potentially non-zero. Finally, we notice from the rightmost column of Figure 11 that in some cases, sparsifying  $\mathbf{W}$  improves the results as witnessed by MC’s normalization with Cosine similarity.

#### *Other criteria: summary.*

We lack space to provide results as detailed for the support, confidence and leverage parameters, but we can sketch general patterns: results of MARs, F- and P-MARs follow similar trends. Without sparsification, the best results are obtained in general for MC, while they are obtained in general for SDS with sparsification. The poorest normalization is always, and sometimes by far, DS. Hence, we see once again that the best parameters to build the manifold also yield the most accurate manifold association rules.

### 3.7 Manifold visualization

A very important perspective on manifold-based learning is the way we map the manifold on end-user devices. From the geometrical standpoint, conventional Euclidean embeddings become suboptimal with respect to more sophisticated embeddings: for example, making all data fit into the display may reduce the readability of data (Figure 12, (a)). This problem does not hold for the Poincaré and Beltrami-Klein disks (resp. (b) and (c) in Figure 12; manifold computed using Jaccard index (8), and SDS normalization of  $\mathbf{W}$  without sparsification). These mappings keep invariant the Voronoi diagram, and hence the wings of the digital library observed in (b), from the organization of colors, actually exist in the manifold. Here, the east green part of the picture is a wing of cooking; it is noticeable that a lot of health books (deep blue) are attached to cooking books (green), and two of the health books that are the closest to cooking’s turn out to be diet books. Such topics did not belong to the data used: they have been brought together and define clear-cut wings in the digital library only because manifold learning was



**Figure 11: Truncated distributions of non-zero lifts for MARs without (top row) and with (bottom row) sparsification of  $W$ . Plots displayed are those for which we observe the most significant differences between the situation with and without sparsification. This difference is taken as the ratio between the proportion of very large lifts ( $\approx 100$ ). The bottom line gives the indication of the similarity function ((H)eat kernel, (C)osine similarity, (J)accard index) and the normalization of  $W$ , along with a symbol indicating whether the proportion of large lifts is larger without (+) or with (-) sparsification.  $x$ -scales are *logscales* (see text).**

successful in capturing, from the local behaviours of users, a global contextual structure meaningful for the librarian or the expert. This map is a concrete example of the power a careful application of non-linear manifold learning techniques. In (c), lines denote itemsets that belong to the same association rules. Remark that this geometric visualization of association rules offers, on a single picture, a general view which carries out more informations than the equivalently long lists of association rules that would be displayed in a more conventional way.

#### 4. SCALING UP TO A GROWING LIBRARY

Armed with the best set of parameters from Section 3, the digital library is currently being scaled up to its real size since the end of 2011. Figure 12 (d, e) presents 2D and 3D focus+context visualizations that were specifically developed from hyperbolic geometric visualizations (Subsection 3.7), to fit rectangle displays like TVs or tablets. The dataset now contains 60 000+ readers (mainly scholars, students and library patrons, mostly from Europe and Maghreb countries) and containing 9 980 books dealing with science, travel, business and cooking<sup>4</sup>. In the same way as was remarked the correlation between the manifold learned and the book topics — not included in data — in subsection 3.7, we can remark, from (e), the strong correlation between Dewey classification subjects — not included in data — and the organization of the manifold on this larger domain.

#### 5. CONCLUSION AND PERSPECTIVES

When it comes to digital libraries, the Big Data phenomenon reminds the magical realistic tale of the (huge) Library of Babel of J.-L. Borges, and in particular the metaphor that the library’s random content is so big that any book ever written must be hidden somewhere, but the human population roaming the library — including the librarians — is left in a state of complete despair. Digital libraries face the challenge of lifting their tools towards an always improved meaningful handling of their content for the end-user. Our

paper is a step towards a roadmap to handle this problem with non-linear manifold learning algorithms. Its solution is not straightforward, yet experimental clues provide us with sets of working parameters that seem to perform well, among which the emphasis on a “bag of users” encoding allowing to learn the dynamics of users wandering around virtual bookshelves. The current scaling up to a whole digital library, and the feedback being received from readers comfort us in this opinion. From a research standpoint, the topological layer of the manifold provides us with a material to build association rules with striking results. More than just our application to the digital library, computational complexity is perhaps the strongest advocate for the application of manifold association rules to other domains [24].

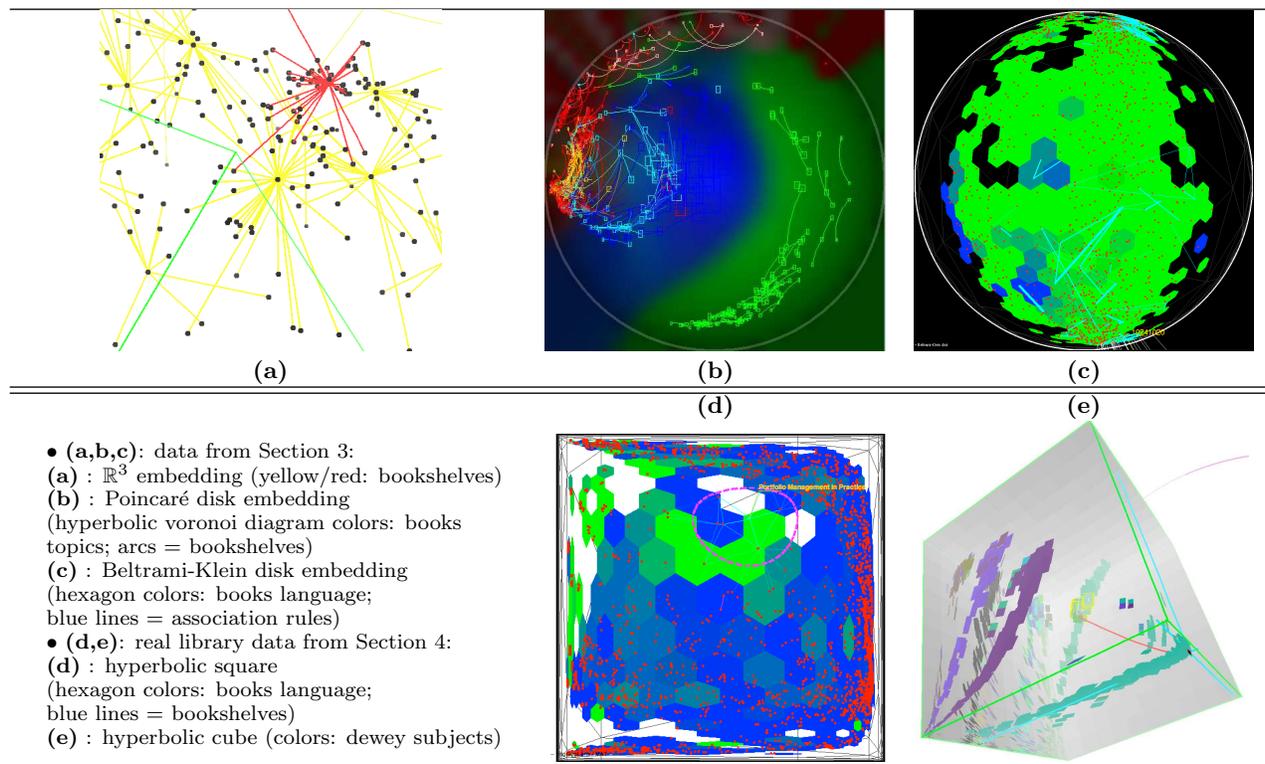
#### 6. ACKNOWLEDGMENTS

The authors wish to thank the reviewers for constructive comments that helped to improve the paper.

#### 7. REFERENCES

- [1] M. Belkhin and P. Niyogi. Laplacian eigenmaps and spectral techniques for embedding and clustering. In T. G. Dietterich, S. Becker, and Z. Ghahramani, editors, *NIPS\*14*. MIT Press, 2001.
- [2] J.-D. Boissonnat, F. Nielsen, and R. Nock. Bregman voronoi diagrams. *DCG*, 44(2):281–307, 2010.
- [3] J. Chim, R.-W.-H. Lau, H.-V. Leong, and A. Si. CyberWalk: a web-based distributed virtual walkthrough environment. *IEEE T. Mult.*, 5:503–515, 2003.
- [4] S. Deerwester, S.-T. Dumais, T.-K. Landauer, G.-W. Furnas, and L. Beck. Improving information retrieval with latent semantic indexing. In *51<sup>st</sup> Ann. Meeting of the Am. Society for Inf. Science*, pages 36–40, 1988.
- [5] C. Elkan. Deriving TF-IDF as a Fisher kernel. In *String Processing and Information Retrieval*, volume 3772 of *LNCS*, pages 295–300. Springer Verlag, 2005.
- [6] B.-J. Frey and D. Dueck. Clustering by passing messages between data points. *Science*, 315:972–976, 2007.

<sup>4</sup><http://www.scholarvox.com/?sitelang=en>



**Figure 12: Different kinds of viewing for the digital library (see Subection 3.7 and Section 4 for details).**

- [7] K. Grønbaek, A. Rohde, B. Sundararajah, and S. Bech-Petersen. Infogallery: informative art services for physical library spaces. In *6<sup>th</sup> JCDL*, pages 21–30, 2006.
- [8] J. Han, J. Pei, Y. Yin, and R. Mao. Mining frequent patterns without candidate generation. *Data Mining and Knowledge Discovery*, 8:53–87, 2004.
- [9] Q. Hardy. Harvard Releases Big Data for Books, 2012. <http://bits.blogs.nytimes.com/2012/04/24/harvard-releases-big-data-for-books/>.
- [10] T. Joachims. Text categorization with support vector machines: Learning with many relevant features. In *Proc. of the 10<sup>th</sup> ECML*, pages 137–142, 1998.
- [11] W. Ke. Information-theoretic term weighting schemes for document clustering. In *13<sup>th</sup> JCDL*, 2013.
- [12] Y. Ke, J. Cheng, and W. Ng. Mining quantitative correlated patterns using an information-theoretic approach. In *12<sup>th</sup> ACM KDD*, pages 227–236, 2006.
- [13] T.-K. Landauer, D.-S. McNamara, S. Dennis, and W. Kintsch. *Handbook of Latent Semantic Analysis*. University of Colorado Institute of Cognitive Science Series, 2007.
- [14] M. Meilă and J. Shi. Learning segmentation by random walks. In T. G. Dietterich, S. Becker, and Z. Ghahramani, editors, *NIPS\*14*. MIT Press, 2001.
- [15] J.-B. Michel, Y.-K. Shen, A.-P. Aiden, A. Veres, M.-K. Gray, The Google Books Team, J.-P. Pickett, D. Hoiberg, D. Clancy, P. Norvig, J. Orwant, S. Pinker, M.-A. Nowak, and E.-L. Aiden. Quantitative analysis of culture using millions of digitized books. *Science*, 331:176–182, 2010.
- [16] R. Nock, P. Luosto, and J. Kivinen. Mixed Bregman clustering with approximation guarantees. In *Proc. of the 19<sup>th</sup> ECML*, pages 154–169, 2008.
- [17] R. Nock, M. Sebban, and D. Bernard. A simple locally adaptive nearest neighbor rule with application to pollution forecasting. *International Journal on Pattern Recognition and Artificial Intelligence*, 17:1–14, 2003.
- [18] P. Serdyukov, V. Murdock, and R. van Zwol. Placing flickr photos on a map. In *33<sup>rd</sup> ACM SIGIR*, pages 484–491, 2009.
- [19] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Trans. PAMI*, 22:888–905, 2000.
- [20] H.-Y. Shiaw, R.-J.-K. Jacob, and G.-R. Crane. The 3D vase museum: a new approach to context in a digital library. In *4<sup>th</sup> JCDL*, pages 125–134, 2004.
- [21] R. Sinkhorn. A relationship between arbitrary positive matrices and doubly stochastic matrices. *Annals of Mathematical Statistics*, 35:876–879, 1964.
- [22] R. Srikant and R. Agrawal. Mining generalized association rules. In *Proc. of the 21<sup>th</sup> VLDB*, pages 407–419, 1995.
- [23] J. B. Tenenbaum, V. de Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290:2319–2323, 2000.
- [24] G. Yang. The complexity of mining maximal frequent itemsets and maximal frequent patterns. In *Proc. of the 10<sup>th</sup> ACM KDD*, pages 344–353, 2004.
- [25] R. Zass and A. Shashua. Doubly-stochastic normalization for spectral clustering. In *NIPS\*19*, 2006.