Texture classification of photographic papers: improving spectral clustering using filterbanks on graphs

Nicolas TREMBLAY[1], Stéphane G. ROUX[1], Pierre BORGNAT[1], Patrice ABRY[1], Herwig WENDT[2], Paul MESSIER[3]

[1]Laboratoire de Physique, ENS Lyon, CNRS 46 alle d'Italie, 69007, Lyon, France

[2]CNRS, IRIT-ENSHEEIT, Université de Toulouse, France

[3]Paul Messier LLC, Boston, USA

{prenom.nom}@ens-lyon.fr; herwig.wendt@irit.fr; admin@paulmessier.com

Résumé – Du point de vue du traitement du signal sur graphe, la méthode classique de classification dite de "clustering spectral" apparaît comme un banc de filtres passe-bas idéal. Intégrant une adaptation de la détection multiéchelle de communautés de [11] au concept de cœurs de communautés [8], nous proposons une méthode de classification basée sur d'autres bancs de filtres plus adaptés aux données. Dans le cadre d'une classification de textures de papiers photos utile en histoire de l'art, les résultats de cette méthode s'avèrent plus riches et aisément interprétables.

Abstract – From the point of view of graph signal processing, we show that spectral clustering is equivalent to an ideal low-pass filterbank. Building upon previous multiscale community detection ideas [11], and integrating the concept of community cores [8], we propose a data-driven filterbank-based classification method. We apply this method to the texture classification of photographic papers useful to art historians, and we show that it provides a richer and more informative description of the data's structure in clusters.

1 Introduction

An element for historical study of photographic papers is to carefully look at their surface texture as this provides interesting facts about working practices of manufacturers as well as photographs printed on them [5]. To analyze raking light images of photographic papers, the method in [1] combines texture analysis using Hyperbolic Wavelet Transforms and spectral clustering, and shows the relevance of the approach to gain insights into these photographic prints for, e.g., datation, authentification, or questioning stylistic issues.

In this context, we show how to improve unsupervised spectral clustering, used in [1], for a finer grouping of photographic prints in clusters. A shortcoming of spectral clustering is that the number K of clusters is arbitrary, and one needs to use external criteria (e.g., AIC, BIC) to estimate it [3]. Also, it outputs a strict partition where outliers, for instance, are necessarily associated to the closest cluster instead of (ideally) staying unlabelled. We propose an alternative to these shortcomings by revisiting spectral clustering as being an ideal low-pass filtering on graphs and considering other multiscale lowpass filterbanks instead, following what has been done for community detection in networks [11]. We show that a measure of cluster stability combined with a notion of cluster core provide us with tools to estimate the relevant number(s) of clusters as well as the relevance of a given print's attribution to a cluster. This is applied to a dataset of photographic papers.

Work supported by the GRAPHSIP project, ANR-14-CE27-0001-02.

2 Photographic paper characterization

Within the *Historic Photographic Paper Classification Challenge* [5] developed by P. Messier and C.R. Johnson, a dataset of 120 non-printed photographic paper samples has been collected, and the images of their textures were made publicly available (`http://papertextureid.org`). These samples cover a wide range of photographic papers, in terms of manufacturer, brand, texture, reflectance and year of production (all that being the metadata provided). It consists in: (1) samples from one same paper sheet (groups G_0 to G_2); (2) samples from paper sheets with the same metadata (G_3 to G_5); (3) samples that vary in production date (G_6 to G_8); (4) a collection of 30 other sheets (labelled o for 'others') representing a variety of photographic textures and metadata. An imaging system, name TextureScope and fully described in [5], uses a raking light and a microscope to depict 1.00×1.35 cm^2 of a paper surface. The images are digitized in images of 1536×2080 pixels. This imaging method is apt at capturing both microscopic features and macroscopic (ir)regularities.

For texture characterization, we follow strictly the method proposed in [1] to obtain a distance between the textures of any two photographic prints. The idea is to make use of the Hyperbolic Wavelet Transform (HWT), that allows us to capture both anisotropy in the textures and their scale-invariant behaviours [7]. The output is a cepstral-type distance matrix $\mathbf{C}_{ab}$ between the multiscale representations (in terms of HWT) of papers a and b. A non-linear transformation is applied to change the distance matrix $\mathbf{C}$ into a similarity matrix of ele-

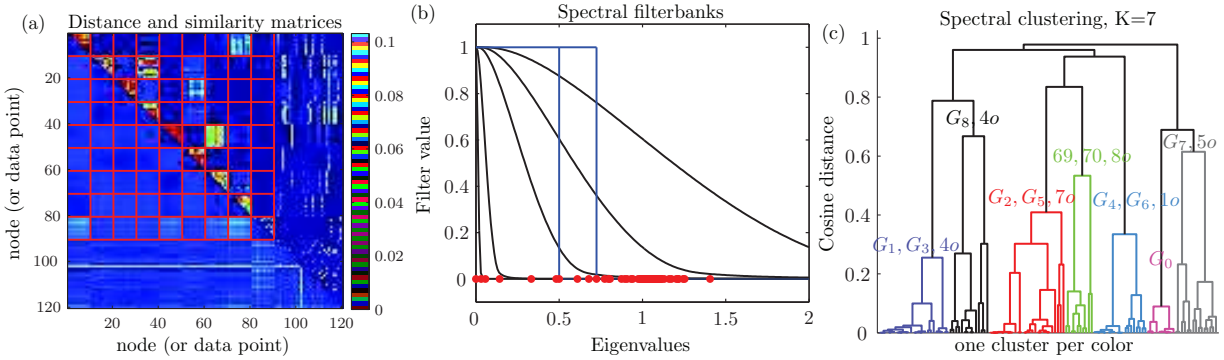

FIG. 1: (a) Distance matrix ($\mathbf{C}$) in lower triangle and similarity matrix ($\mathbf{W}$) in upper triangle; each triangle is normalized to unity. Groups of similar papers G_0 to G_8 are separated by red lines. (b) Filterbanks for spectral clustering (ideal low-pass l_K in blue, for $K = 7, 10$) and for scaling functions (smooth low pass $\boldsymbol{h}_s$ in black, for various values of s). Red dots represent $\boldsymbol{\mathcal{L}}_{rw}$'s eigenvalues. (c) Dendrogram computed with spectral clustering, with $K = 7$ clusters identified.

ments $\mathbf{W}_{ab} = \exp(-\mathbf{C}_{ab}/\epsilon)$, where ϵ is a constant assessing the typical closeness between images on the dataset. This is illustrated in Fig. 1 (a) and $\mathbf{W}$ appears to have larger values within groups, than between them.

3 Spectral clustering revisited

3.1 Background: spectral clustering

Unsupervised clustering can be interpreted, thanks to the similarity matrix $\mathbf{W}$, as a partition problem of an undirected weighted graph where each node is one of the N data samples. Let $\mathcal{G}$ be this graph and $\mathbf{W}$ the weighted and symmetric adjacency matrix; $\mathbf{S}$ is a diagonal matrix with $\mathbf{S}_{aa} = \sum_{b \neq a} \mathbf{W}_{ab}$. The random walk Laplacian matrix is $\boldsymbol{\mathcal{L}}_{rw} = \mathbf{I}_N - \mathbf{S}^{-1}\mathbf{W}$, where $\mathbf{I}_N$ is the identity matrix of size N. We recall (see, e.g. [6]) that $\boldsymbol{\mathcal{L}}_{rw}$ is diagonalizable, with eigenvalues sorted such that: $0 = \lambda_1 \leq \lambda_2 \leq \lambda_3 \leq \ldots \leq \lambda_N \leq 2$; and normalized eigenvectors, here collected in a matrix: $\boldsymbol{\chi} = (\boldsymbol{\chi}_1 | \boldsymbol{\chi}_2 | \ldots | \boldsymbol{\chi}_N)$.

To partition the graph, spectral clustering [3, 6] computes the first K eigenvectors and creates a feature vector $\mathbf{f}_{K,a} \in \mathbb{R}^K$ for each node a: $\forall k \in [1, K] \quad \mathbf{f}_{K,a}(k) = \boldsymbol{\chi}_k(a)$, which is in turn used to obtain K clusters thanks to, e.g., K-means or hierarchical clustering.

3.2 Filterbanks-based spectral graph clustering

Spectral clustering's feature vectors. An analogy between signals on graphs and usual signals [9] suggests to interpret the spectrum of $\boldsymbol{\mathcal{L}}_{rw}$ as a Fourier domain for graphs, hence defining filters on graphs as diagonal operators after change of basis with $\boldsymbol{\chi}^{-1}$. It turns out that the features $\mathbf{f}_{K,a}$ can be obtained by ideal low-pass filtering of the Delta function δ_a (localized at node a). Indeed, let l_K be the step function where $l_K(\lambda) = 1$ if $\lambda \leq \lambda_K$ and 0 otherwise. We define $\mathbf{L}_K$ the diagonal matrix for which $\mathbf{L}_K(i,i) = l_K(\lambda_i)$. Then:

$$\mathbf{f}_{K,a} = \mathbf{L}_K \boldsymbol{\chi}^{-1} \delta_a \in \mathbb{R}^N, \tag{1}$$

when filling the last $N - K$ values with 0's. Henceforth, spectral clustering is equivalent to clustering using low-pass filtering of the local descriptors δ_a of each node a of the graph $\mathcal{G}$.

Scaling function filterbanks' feature vectors. This analogy opens the way to use other (than l_K) filter kernels to define feature vectors. Especially, one can rely on a notion of scale as in [11], where it is shown that a wavelet filterbank on graphs, as defined in [2], behaves well to detect communities in complex networks. To stay in this paper's scope, we limit the analysis to scaling function filterbanks, defined from a low-pass filter kernel function h designed in the "Fourier" space $[0, 2]$, as in [10]. At every scale $s \in \mathbb{R}^{+*}$, a discrete filter vector $\boldsymbol{h}_s$ is $\forall i \in [1, N] \quad \boldsymbol{h}_s(i) = h(s\lambda_i)$, and is used to define a new feature vector: $\mathbf{f}_{s,a} = \mathbf{H}_s \boldsymbol{\chi}^{-1} \delta_a$ where $\mathbf{H}_s = \mathrm{diag}(\mathbf{h}_s)$. To explain our terminology, note that $\boldsymbol{\chi}\mathbf{f}_{s,a}$ may be interpreted as the scaling function at scale s centered around node a (see [10]). These filterbanks are illustrated in Fig. 1 (b).

Distance matrix. To compute the distance matrix $\boldsymbol{D}_s$ between the scaling function's feature vectors, we use the cosine distance, as it was shown in [11] to be a good measurement of the similarity between two data a and b at scale s :

$$\boldsymbol{D}_s(a,b) = 1 - \frac{\boldsymbol{f}_{s,a}^\top \boldsymbol{f}_{s,b}}{||\boldsymbol{f}_{s,a}||_2 \, ||\boldsymbol{f}_{s,b}||_2}. \tag{2}$$

In spectral clustering, there is no consensus on which distance to choose. Here, for a fair comparison with the scaling function method, the cosine distance is also used.

Clustering In both cases, we use a hierarchical agglomerative clustering with average-linkage [3], that outputs a dendrogram from the distance matrix. In the case of spectral clustering (resp. scaling function filterbanks), the dendrogram is cut in order to create K clusters (resp. at the largest global gap as in [11]). A multiscale analysis in the case of spectral clustering (resp. scaling function filterbanks) is obtained by varying K (resp. $s \in \mathbb{R}^{+*}$). Fig. 1 (c) shows the result for $K = 7$.

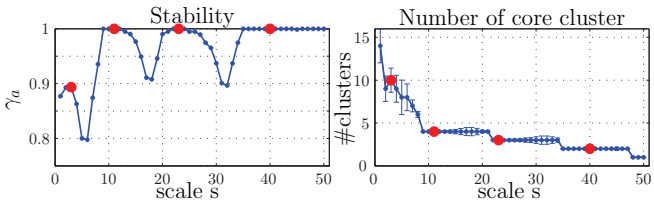

FIG. 2: Stability (left) and number of clusters (right), vs. scale. The red points are local maxima of the stability, hence the scales of interest. The confidence intervals around the number of clusters are from the J (possibly different) partitions $\{P_s^j\}$.

3.3 Stochastic filterbanks-based clustering

Following Section 5 of [11], we take advantage of the fast graph wavelet transform of η random vectors to directly estimate $\boldsymbol{D}_s(a,b)$ knowing $\boldsymbol{\mathcal{L}}_{rw}$ and $\boldsymbol{h}_s$, without computing each feature vector $\mathbf{f}_{s,a}$. A proof is explicit in [11] when considering wavelets. For scaling functions, the results hold if one removes the zero frequency. This method is now stochastic: at every scale s, one may synthesize J sets of η random vectors, and obtain J partitions P_s^j in clusters. This stochasticity allows us to develop two helpful notions: the **stability** of partitions [11], telling which scales (and henceforth which number of clusters) are relevant; and the **relevance of the attribution of a node to a cluster**, probed using the cluster cores developed in [8].

Stability $\gamma_a(s)$ is defined for each scale s as the mean of the similarity between all pairs of partitions of $\{P_s^j\}_{j\in[1,J]}$:

$$\gamma_a(s) = \frac{2}{J(J-1)} \sum_{(i,j)\in[1,J]^2, i\neq j} \texttt{ari}(P_s^i, P_s^j) \qquad (3)$$

where the function ari is the Adjusted Rand Index, measuring the similarity between partitions [4]. The more stable is the partition at scale s, i.e. the more interesting this scale is, the closer to 1 will be $\gamma_a(s)$. Only the most stable scales provides partitions that are worth retaining.

We scan a sampled set of scales, typically 50 logarithmically spaced scales between two boundaries automatically detected by the algorithm [11] and we output the most stable scales s^* (local maxima of $\gamma_a(s)$) and their associated partitions. This also gives an estimation of the number K of clusters that are relevant to keep. It bypasses the issue of choosing or estimating K in classical spectral clustering. This is illustrated in Fig. 2.

Detecting cluster cores. Following [8], two nodes a and b are in a cluster core if, in each of the J partitions, they are always classified in the same cluster. Clusters of size 1 are not counted as cores. Given $\{P_{s^*}^j\}_{j\in[1,J]}$, we obtain, for each stable scale s^*, a list of cluster cores $\{C_z\}_{z\in[1,Z]}$ and the association of nodes to them returned as $(P_{s^*}^c, \rho_{s^*}) \in (\mathbb{R}^N)^2$:

- if a belongs to core C_z, then $P_{s^*}^c(a) = z$ and $\rho_{s^*}(a) = 1$.
- else, we compute the core $C_{z'}$ in which it is most often classified and write $P_{s^*}^c(a) = z'$, and $\rho_{s^*}(a) = n_{z'}/J$ the relative frequency with which it is associated to z'. In some cases, a is always in his own disregarded cluster of size 1 and kept unclassified ($P_s^c(a) = 0$ and $\rho_s(a) = 0$).

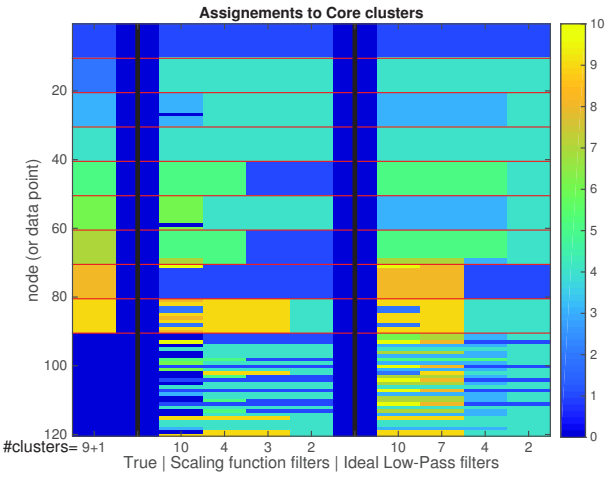

FIG. 3: Estimated clusters (at each scale, a cluster is a color). First column: groups G_i in the dataset (separated by red lines). Middle: cores from scaling functions (one scale s^* per column); unclassified samples are in deep blue (value 0). Right: clusters from spectral clustering (one K per column).

At scale s^*, if $\rho_{s^*}(a) = 1$, then the algorithm's user may confidently classify data a in cluster core $P_{s^*}^c(a)$. If not, then a cannot be classified with full reliability with other points in this dataset, and $P_{s^*}^c(a)$ is only an indication of its closest core.

4 Results

We apply the proposed stochastic scaling function filterbanks-based clustering on the 120 photographic paper samples; with $\eta = 100$, $J = 80$, and compute stability and attribution to cluster cores $(P_{s^*}^c, \rho_{s^*})$ for the most stable scales s^*. In Fig. 3, we compare these results to simple spectral clustering (but stability nor cores) with features from eq. (1), for which we kept $K =$ 2, 4, 7, or 10 as they correspond to eigenvalue gaps of $\boldsymbol{\mathcal{L}}_{rw}$ [1] (this is a usual arbitrary rule of thumb to choose K [6]).

Discussion about the groups of photographic papers. The dataset is structured in 9 groups from G_0 to G_8 of 10 sheets each, plus 30 different samples. Most of these 9 groups in the dataset end up being in core clusters at most of the scales, some groups being very close and usually in the same core: G_1 is close to G_3, because the paper in G_1 is from the same groups as papers in G_3; G_4 to G_6 which are all glossy papers; and (less so) G_2 (Chamois reflectance) to G_5 (Half Mat reflectance).

Other elements could be more surprising. At small scale, group G_8 is separated in 3 cores. After careful inspection, there are indeed visual differences in G_8 and the 3 groups make sense: 83 and 84 (same brands, same date), 82, 85, 87, 89 (adequate visual matches) and 81 and 90 (visually a good match

TAB. 1: Results of classification with 10 clusters.

Cluster =	with G_i's	with others	unclassified
Sp. Cl., G_i	87	3	0
Sp. Cl., others	17	13	0
Sc.Fct, G_i	85	3	2
Sc.Fct, others	16	1	13

with more spiky features than others in G_8). Samples 71 and 69-70 are not classified in their expected group G_7 and G_6. This makes sense as 69-70 are from a different brand than the rest of G_6, and 71 has indeed more rounded features than the generally spiky features of G_7.

Clustering in action: a scanning through scales. The stability of the scaling functions (see Fig. 2) retain $2, 4$ and 10 clusters as relevant partitions, and this is an advantage of the method – providing an estimation of the relevant number of clusters, while the choice for K in spectral clustering is more arbitrary. The stability also tells us that the spectral clustering in 7 groups given in Fig. 1 (c) is not fully reliable. It is obtained for scaling functions yet for a smaller stability. An explanation might be that this scale with 7 clusters allows mostly for: 1) an early splitting of G_8 from the cluster of G_1-G_3, yet they have all Luster reflectance, 2) splitting of G_7 from G_0 yet this separation is, according to the dendrogram, almost equivalent as separating also sample 71 from G_7, and 3) the creation of a group of other samples (in green) that could as well be also split in 2, like G_8. On the dendogram for $K = 7$, the splitting in 10 clusters is actually apparent, and is more relevant.

Photographic papers not in groups. Table 1 summarizes the classification, for samples in the G_i's and the 30 other samples. A drawback of spectral clustering is that each of these papers necessarily ends up in a cluster (the closest), even though it may hardly share anything with it. For $K = 10$, 17 such samples are put in clusters associated of one (or some) of the 9 groups, while these connections are not supported by the metadata. With scaling functions, a more interesting conclusion is reached: 13 of the other samples are left unclassified at smallest scale (i.e., $\rho_{s^*}(a) < 1$), and only 2 of the samples from the groups are not. This is relevant as there is no way that this small dataset of papers captures 100 years of photo paper manufacture: we expect to have some papers matching none.

We end up with only half of the other samples matching one of the groups. Some are indeed correct matches. For example, associated to cores of G_1 and G_3 (having metadata: Kodak, Kodabromide, Fine Grained, Lustre, 1967), we find the papers 106 (Kodak, Azo, smooth, glossy); 112 (Kodak, Kodabromide, buff luster, 1950); and 117 (Kodak, Kodabromide, smooth, Glossy, 1959) match in manufacturers and paper brand. The interpretation might be that manufactures had similar, and possibly limited means to achieve a smooth, reflective, surface given the realities of the materials and techniques they had available at that time. However, the message is that, if anything in the clustering casts doubts about the association – and the estimate of cores is a way to raise doubts – one should refrain from considering the association valid.

5 Conclusion

We proposed an extension of spectral clustering, taking the point of view of filterbanks in the graph Fourier domain. Instead of an ideal low-pass filterbank, we consider low-pass scaling functions filterbanks, and add to the method measures of stability (to estimate K) and a probability of correct classification for each node. An extension of the work would be to sue wavelet filterbanks on graphs [2] for classification, and we might expect details at finer scales with them.

For art historians, it is of great added value to have a measure of the confidence of classification, and to be able to keep some samples unclassified: this is required so as to avoid drawing conclusion from intempestive associations. As a complement to scholarship about photographic prints, clustering should be used in the most strict manner: not fully reliable associations should not be considered by art historian scholars.

References

[1] P. Abry *et al.*, Multiscale Anisotropic Texture Analysis and Classification of Photographic Prints. *Signal Proc. Mag.*, to be published, July 2015.

[2] D.K. Hammond, *et al.*Wavelets on graphs via spectral graph theory. *Applied and Computational Harmonic Analysis*, vol. 30, no. 2, pp. 129–150, 2011.

[3] T. Hastie, *et al.The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. NY, USA. Springer, 2001.

[4] L. Hubert and P. Arabie, Comparing partitions, *Journal of Classification*, vol. 2, no. 1, pp. 193–218, 1985.

[5] C. R. Johnson *et al.*, Pursuing automated classification of historic photographic papers from raking light photomicrographs. *Journal of the American Institute for Conservation*, vol. 53, no. 3, pp. 159–170, 2014.

[6] U. von Luxburg. A tutorial on spectral clustering. *Statistics and computing*, 17(4):395–416, 2007.

[7] S. G. Roux *et al.*, "Self-Similar Anisotropic Texture Analysis: The Hyperbolic Wavelet Transform Contribution," *IEEE Trans. Image Proc.*, vol. 22, no. 11, pp. 4353–4363, 2013.

[8] M. Seifi, *et al.*Stable community cores in complex networks, *Complex Networks*, pp. 87–98, 2013.

[9] D.I. Shuman, *et al.* The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains, *Signal Processing Magazine*, vol. 30, no. 3, pp. 83–98, 2013.

[10] N. Tremblay and P. Borgnat, Multiscale community mining in networks using spectral graph wavelets, *21st EUSIPCO*, Marrakech, Morocco, Sep. 2013.

[11] N. Tremblay and P. Borgnat, Graph Wavelets for Multiscale Community Mining, *Signal Processing, IEEE Transactions on*, vol. 62, no. 20, pp. 5227–5239, 2014.