→ CROVELLA'S SLIDES

(PAT KELLY – PCA) { MACHINE LEARNING
YUNG-CHIH → PCA TUTORIAL

Coifman / Maggioni

- RR Coifman and M Maggioni, **"Diffusion wavelets,"** In *Appl. Comp. Harm. Anal.*, 21(1):53–94, July 2006. (pdf)

- Mauro Maggioni and Ronald R Coifman, **"Multiscale Analysis of Data Sets with Diffusion Wavelets,"**, Proc. *Data Mining for Biomedical Informatics*. (pdf)
  Analysis of functions of manifolds and graphs is essential in many tasks, such as learning, classification, clustering. The construction of efficient decompositions of functions has till now been quite problematic, and restricted to few choices, such as the eigenfunctions of the Laplacian on a manifold or graph, which has found interesting applications. In this paper we propose a novel paradigm for analysis on manifolds and graphs, based on the recently constructed diffusion wavelets. They allow a coherent and effective multiscale analysis of the space and of functions on the space, and are a promising new tool in classification and learning tasks. In this paper we overview the main motivations behind their introduction, their properties, and sketch an application to multiscale document corpora analysis

YUNG-CHIH

EDMUNDO
LUMPABILITY.

- R.R. Coifman, S Lafon, M Maggioni, Y Keller, AD Szlam, F J Warner, S W Zucker, **"Geometries of sensor outputs, inference and information processing,"**. In *Proc. SPIE*, Vol. 6232, 623204, May 2006. (pdf)
  We describe signal processing tools to extract structure and information from arbitrary digital data sets. In particular heterogeneous multi-sensor measurements which involve corrupt data, either noisy or with missing entries present formidable challenges. We sketch methodologies for using the network of inferences and similarities between the data points to create robust nonlinear estimators for missing or noisy entries. These methods enable coherent fusion of data from a multiplicity of sources, generalizing signal processing to a non linear setting. Since they provide empirical data models they could also potentially extend analog to digital conversion schemes like "sigma delta"

- M Maggioni, AD Szlam, RR Coifman, and JC Bremer Jr., **"Diffusion-driven multiscale analysis on manifolds and graphs:top-down and bottom-up constructions,"** In *Proc. SPIE Wavelet XI*, Vol 5914, 59141D, Sep 2005. (pdf)
  Classically, analysis on manifolds and graphs has been based on the study of the eigenfunctions of the Laplacian and its generalizations. These objects from differential geometry and analysis on manifolds have proven useful in applications to partial differential equations, and their discrete counterparts have been applied to optimization problems, learning, clustering, routing and many other algorithms.1–7 The eigenfunctions of the Laplacian are in general global: their support often coincides with the whole manifold, and they are affected by global properties of the manifold (for example certain global topological invariants). Recently a framework for building natural multiresolution structures on manifolds and graphs was introduced, that greatly generalizes, among other things, the construction of wavelets and wavelet packets in Euclidean spaces.8,9 This allows the study of the manifold and of functions on it at different scales, which are naturally induced by the geometry of the manifold. This construction proceeds bottom-up, from the finest scale to the coarsest scale, using powers of a diffusion operator as dilations and a numerical rank constraint to critically sample the multiresolution subspaces. In this paper we introduce a novel multiscale construction, based on a top-down recursive partitioning induced by the eigenfunctions of the Laplacian. This yields associated local cosine packets on manifolds, generalizing local cosines in Euclidean spaces.10 We discuss some of the connections with the construction of diffusion wavelets. These constructions have direct applications to the approximation, denoising, compression and learning of functions on a manifold and are promising in view of applications to problems in manifold approximation, learning, dimensionality reduction.

Other

- Boaz Nadler, Meirav Galun, **"Fundamental Limitations of Spectral Clustering,"** *NIPS* 2006
  Spectral clustering methods are common graph-based approaches to clustering of data. Spectral

clustering algorithms typically start from local information encoded in a weighted graph on the data and cluster according to the global eigenvectors of the corresponding (normalized) similarity matrix. One contribution of this paper is to present fundamental limitations of this general local to global approach. We show that based only on local information, the normalized cut functional is not a suitable measure for the quality of clustering. Further, even with a suitable simi- larity measure, we show that the first few eigenvectors of such adjacency matrices cannot successfully cluster datasets that contain structures at different scales of size and density. Based on these findings, a second contribution of this paper is a novel diffusion based measure to evaluate the coherence of individual clusters. Our measure can be used in conjunction with any bottom-up graph-based cluster- ing method, it is scale-free and can determine coherent clusters at all scales. We present both synthetic examples and real image segmentation problems where var- ious spectral clustering algorithms fail. In contrast, using this coherence measure finds the expected clusters at all scales

DON-CROWCROFT PRESENTATION
BUBBLE-RAP IN 3 WEEKS ($\pm$1)
BRUNO

- Ann B. Lee and Larry Wasserman, **"Spectral Connectivity Analysis"**, pre-print, Nov. 2008

  Spectral kernel methods are techniques for transforming data into a coordinate system that efficiently reveals the geometric structure— in particular, the "connectivity"—of the data. These methods depend on certain tuning parameters. We analyze the dependence of the method on these tuning parameters. We focus on one particular technique—diffusion maps—but our analysis can be used for other methods as well. We identify the population quantities implicitly being estimated, we explain how these methods relate to classical kernel smoothing and we define an appropriate risk function for analyzing the estimators. We also show that, in some cases, fast rates of convergence are possible even in high dimensions.

PNAS PAPER (WEI)BO. POINT TO EDMUNDO

- Stephane Lafon and Ann B. Lee, **"Diffusion Maps and Coarse-Graining: A Unified Framework for Dimensionality Reduction, Graph Partitioning, and Data Set Parameterization"**, - *IEEE Transactions On Pattern Analysis And Machine Learning*, 2006

  We provide evidence that nonlinear dimensionality reduction, clustering, and data set parameterization can be solved within one and the same framework. The main idea is to define a system of coordinates with an explicit metric that reflects the connectivity of a given data set and that is robust to noise. Our construction, which is based on a Markov random walk on the data, offers a general scheme of simultaneously reorganizing and subsampling graphs and arbitrarily shaped data sets in high dimensions using intrinsic geometry. We show that clustering in embedding spaces is equivalent to compressing operators. The objective of data partitioning and clustering is to coarse-grain the random walk on the data while at the same time preserving a diffusion operator for the intrinsic geometry or connectivity of the data set up to some accuracy. We show that the quantization distortion in diffusion space bounds the error of compression of the operator, thus giving a rigorous justification for k-means clustering in diffusion space and a precise measure of the performance of general clustering algorithms.

NILANJAN

- M Coates, Y Pointurier, M Rabbat, **"Compressed network monitoring for ip and all-optical networks"** , *IMC* 2007,

  We address the problem of efficient end-to-end network monitoring of path metrics in communication networks. Our goal is to minimize the number of measurements or monitors required to maintain an acceptable estimation accuracy. We present a framework based on diffusion wavelets and nonlinear estimation. Our procedure involves the development of a diffusion wavelet basis that is adapted to the monitoring problem. This basis exploits spatial and temporal correlations in the measured phenomena to provide a compressible representation of the path metrics. The framework employs nonlinear estimation techniques using 1 minimization to generate estimates for the unmeasured paths. We describe heuristic approaches for the selection of the paths that should be monitored, or equivalently, where hardware monitors should be located. We demonstrate how our estimation framework can improve the efficiency of end-to-end delay estimation in IP networks and reduce the number of hardware monitors required to track bit-error rates in all-optical networks (networks with no electrical regenerators).

BRUNO

- D Rincón, M Roughan, W Willinger, **"Towards a Meaningful MRA of Traffic Matrices"**, *IMC* 2008

Most research on traffic matrices (TM) has focused on ¯nd- ing models that help with inference, but not with other im- portant tasks such as synthesis of TMs, tra±c prediction, or anomaly detection. In this paper we approach the problem of a general model for traffic matrices, and argue that such a model must be sparse, i.e., have a small number of parame- ters in comparison to the size of the TM. A Multi-Resolution Analysis (MRA) of TMs can provide such a sparse repre- sentation. The Di®usion Wavelet (DW) transform is a good choice as a MRA tool here, because it inherently adapts to the structure of the underlying network. The paper de- scribes our construction of the two-dimensional version of the DW transform and shows how to use it for our proposed MRA of TMs. The results obtained with operational networks con¯rm the sparseness of the DW-based TM analysis approach and its applicability to other TM-related tasks.

SUPP

- Pascal Pons and Matthieu Latapy, "**Computing communities in large networks using random walks,**" In *Journal of Graph Algorithms and Applications* vol. 10, no. 2, pages 191-218, 2006. (pdf)
  Dense subgraphs of sparse graphs (communities), which appear in most real-world complex networks, play an important role in many contexts. Computing them however is generally expensive. We propose here a measure of similarities between vertices based on random walks which has several important advantages: it captures well the community structure in a network, it can be computed efficiently, it works at various scales, and it can be used in an agglomerative algorithm to compute efficiently the community structure of a network. We propose such an algorithm which runs in time $O(mn^2)$ and space $O(n^2)$ in the worst case, and in time $O(n^2 \log n)$ and space $O(n^2)$ in most real-world cases (n and m are respectively the number of vertices and edges in the input graph). Experimental evaluation shows that our algorithm surpasses previously proposed ones concerning the quality of the obtained community structures and that it stands among the best ones concerning the running time. This is very promising because our algorithm can be improved in several ways, which we sketch at the end of the paper.

- L. Grady, "**Random Walks for Image Segmentation**", In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Volume: 28, Issue: 11, Nov 2006.
  A novel method is proposed for performing multilabel, interactive image segmentation. Given a small number of pixels with user-defined (or pre-defined) labels, one can analytically and quickly determine the probability that a random walker starting at each unlabeled pixel will first reach one of the pre-labeled pixels. By assigning each pixel to the label for which the greatest probability is calculated, a high-quality image segmentation may be obtained. Theoretical properties of this algorithm are developed along with the corresponding connections to discrete potential theory and electrical circuits. This algorithm is formulated in discrete space (i.e., on a graph) using combinatorial analogues of standard operators and principles from continuous potential theory, allowing it to be applied in arbitrary dimension on arbitrary graphs.

- S Mahadevan, M Maggioni , "**Value Function Approximation with Diffusion Wavelets and Laplacian Eigenfunctions**", *NIPS*, 2006.
  We investigate the problem of automatically constructing efficient representations or basis functions for approximating value functions based on analyzing the structure and topology of the state space. In particular, two novel approaches to value function approximation are explored based on automatically constructing basis functions on state spaces that can be represented as graphs or manifolds: one approach uses the eigenfunctions of the Laplacian, in effect performing a global Fourier analysis on the graph; the second approach is based on diffusion wavelets, which generalize classical wavelets to graphs using multiscale dilations induced by powers of a diffusion operator or random walk on the graph. Together, these approaches form the foundation of a new generation of methods for solving large Markov decision processes, in which the underlying representation and policies are simultaneously learned

SEND SHRIDAR E-MAIL

Possibly related works

BO

- Jimeng Sun, Dacheng Tao, Christos Faloutsos, "**Beyond Streams and Graphs: Dynamic Tensor Analysis**", In *KDD*, 2006.
  How do we find patterns in author-keyword associations, evolving over time? Or in Data Cubes, with

product-branch-customer sales information? Matrix decompositions, like principal component analysis (PCA) and variants, are invaluable tools for mining, dimensionality reduction, feature selection, rule identification in numerous settings like streaming data, text, graphs, social networks and many more. However, they have only two orders, like author and keyword, in the above example.We propose to envision such higher order data as tensors,and tap the vast literature on the topic. However, these methods do not necessarily scale up, let alone operate on semi-infinite streams. Thus, we introduce the dynamic tensor analysis (DTA) method, and its variants. DTA provides a compact summary for high-order and high-dimensional data, and it also reveals the hidden correlations. Algorithmically, we designed DTA very carefully so that it is (a) scalable, (b) space efficient (it does not need to store the past) and (c) fully automatic with no need for user defined parameters. Moreover, we propose STA, a streaming tensor analysis method, which provides a fast, streaming approximation to DTA.We implemented all our methods, and applied them in two real settings, namely, anomaly detection and multi-way latent semantic indexing. We used two real, large datasets, one on network flow data (100GB over 1 month) and one from DBLP (200MB over 25 years). Our experiments show that our methods are fast, accurate and that they find interesting patterns and outliers on the real datasets.

- Richard G. Baraniuk and Michael B. Wakin, **"Random Projections of Smooth Manifolds"**, *IEEE Intl. Conf. Acoustics, Speech and Signal*", 2006
  We propose a new approach for nonadaptive dimensionality reduction of manifold-modeled data, demonstrating that a small number of random linear projections can preserve key informa- tion about a manifold-modeled signal. Our results bear strong resemblance to the emerging theory of Compressed Sensing (CS), in which sparse signals can be recovered from small numbers of random linear measurements. As in CS, the random measurements we propose can be used to recover the original data in RN. Moreover, like the fundamental bound in CS, our requisite M is linear in the "information level" K and logarithmic in the ambient dimension N; we also identify a logarithmic dependence on the volume and curvature of the manifold. In addition to recovering faithful approximations to manifold-modeled signals, however, the random projections we propose can also be used to discern key properties about the manifold. We discuss connections and contrasts with existing techniques in manifold learning, a setting where dimensionality reducing mappings are typically nonlinear and constructed adaptively from a set of sampled training data.

- James Gary Propp and David Bruce Wilson, **"Exact sampling with coupled Markov chains and applications to statistical mechanics"**. In *Proceedings of the seventh international conference on Random structures and algorithms*, 1996, pp. 223–252. (Mixing time of Markov chains using random walks).