Information Geometric Graph Indexing from Partial Node Coverages

Oral Talk

Francisco Escolano
Boyan Bonev
& Miguel A. Lozano
University of Alicante, Spain

Gbr2011 May 18-20 2011
Müster,GERMANY

Graph Indexing

- Graph Indexing Implies: (i) Specify local-to-global indexes, (ii) Build a suitable representation, and (iii) Define a dissimilarity measure.
- Indexes rely on topological statistics: (e.g. subgraphs – **gIndex** – [Yan et al., SIGMOD'04], summarization graphs – **bags of indexed subgraphs** – [Zhou et al., DASFAA'08], local-to-gobal – **spectral indexation of trees** – [Shokoufandeh et al, PAMI'05]).
- Representations Topological statistics imply a **distributional approach**. Such approaches are currently linked to **embedding**: (e.g. embed vertex-labeled graphs into **normed spaces** and match pairs of weighted distributions using EMD–[Demirci et al., IJCV'06], or embed graph nodes in **manifolds** and deform them for matching–[Escolano, Hancoch & Lozano, CVPR'11].
- Dissimilarities Compare **statistics** vs use **Information-theoretic**

Partial Node Coverages

- Node Histories. Emerge for the natural generalization of the **nesting approach** for trees and from recent approaches in quantifying graph complexity [Escolano et al., SSPR'10].
- Partial Node Coverages. A graph is the **overlap of several subgraphs**, one per each vertex. What is the optimal extent of each coverage? Close to the **graphlet** whose spectrum has been recently described [Kondor et al., ICML'09].
- Spectral Features of Subgraphs. Recent experiments show the **most discriminative spectral features** [Bonev et al.,SSPR'10]: Commute Times, Perron-Frobenius, Fiedler vector and node centrality vector.

Partial Node Coverages

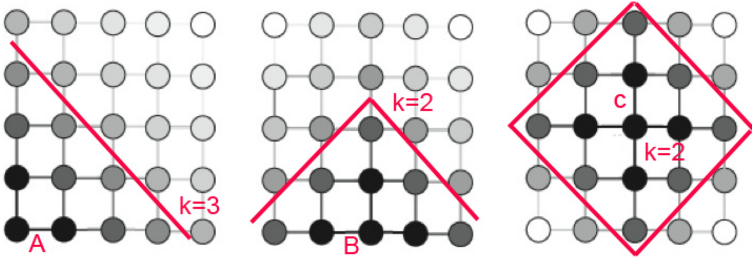
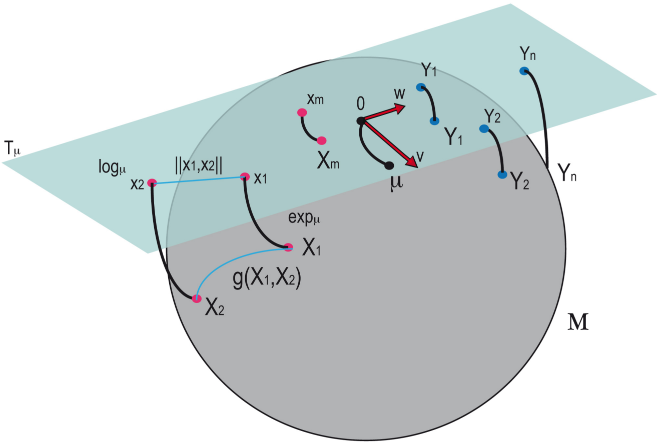


Figure: Several histories and node coverages in a regular grid

Why Information Geometry in Graphs

- Tensors. For each subgraph/partial coverage we can compute a **covariance matrix** from the features measured along the formation of the coverage. Such matrices live in a **Riemannian manifold** [Tuzel et al., PAMI'08], [Porikli, S+SSPR'10].
- Tensor Distributions. Given two graphs to compare, each graph is given by a **population of covariance matrices** (one per node). Although geodesics are defined in the manifold **it is hard to define distributional dissimilarities there**.
- Tangent Space. We define the **tangent space** associated to the **Karcher mean** of the union of both population, deproject tensors and compute distributional dissimilarities in the tangent space.

Information Geometry



Spectral Descriptors

Vectorization

Consider $\Phi(i) = (f_1(i), \ldots, f_d(i))^T$ a vector of spectral descriptors of the partial node coverage $H = e^k(i) \subseteq G$

- For **CTs** we consider both the Laplacian and the normalized Laplacian of H: the elements of the upper off-diagonal elements of the CT kernel are downsampled to select $m = |V_H|$ elements and they are normalized by m^2.
- **Fiedler** and **Perron-Frobenius** vectors have m elements by definition.
- **Node centrality** is more selective than degree. This measure is also normalized by m^2.

Tensors in the Manifold

Tensors

For each partial coverage H we can compute the statistics of d spectral descriptors taking m samples. Such statistics can be easily encoded in a $d \times d$ covariance matrix

$$\vec{X}_i = \frac{1}{n-1} \sum_{i=1}^{m} (\Phi(i) - \vec{\mu})(\Phi(i) - \vec{\mu})^T.$$

lying in a Riemannian manifold $\mathcal{M}$.

Existence of Tangent Space

As a Riemannian manifold is differentiable, the derivatives at each $\vec{X}$ always exist, and such derivatives lie in the so called **tangent space** $T_{\vec{X}}$, which is a vector space in $\mathbb{R}^{d \times d}$.

Elements of the Tangent Space

- The tangent space at $\vec{T}_{\vec{X}}$ is endowed with an **inner product** $< .,. >_{\vec{X}}$ being

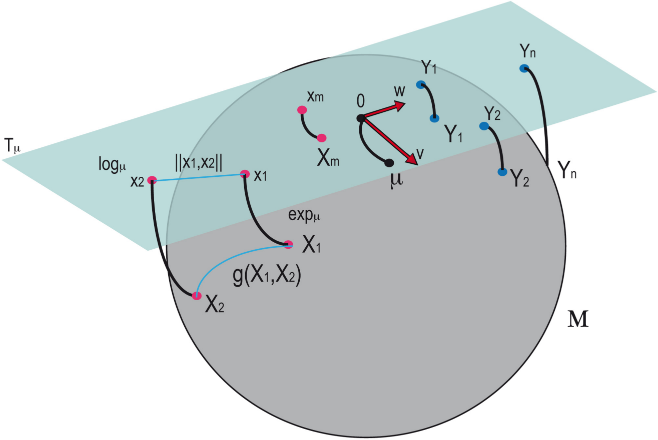
$$< \vec{u}, \vec{v} >_{\vec{X}} = trace(\vec{X}^{-\frac{1}{2}} \vec{u} \vec{X}^{-1} \vec{v} \vec{X}^{-\frac{1}{2}}).$$

- The tangent space is also endowed with an **exponential map** $\exp_{\vec{X}} : T_{\vec{X}} \to \mathcal{M}$ which maps a tangent vector $\vec{u}$ to a point

$$\vec{U} = \exp_{\vec{X}}(\vec{u}) \in \mathcal{M}$$

  that is, it maps $\vec{u}$ to the point reached by the unique **geodesic** from $\vec{X}$ to $\vec{U}$: $g(\vec{X}, \vec{U})$.

Information Geometry



Tangent Space (2)

Elements of the Tangent Space

The **inverse mapping** $\log_{\vec{X}} : \mathcal{M} \rightarrow T_{\vec{X}}$ is uniquely defined in a small neighborhood of $\vec{X}$. Therefore, we have the following mappings for going to the manifold and back (to the tangent space) respectively:

$$\begin{aligned} \exp_{\vec{X}}(\vec{u}) &= \vec{X}^{\frac{1}{2}} \exp(\vec{X}^{-\frac{1}{2}} \vec{u} \vec{X}^{-\frac{1}{2}}) \vec{X}^{\frac{1}{2}} \, . \\ \log_{\vec{X}}(\vec{U}) &= \vec{X}^{\frac{1}{2}} \log(\vec{X}^{-\frac{1}{2}} \vec{U} \vec{X}^{-\frac{1}{2}}) \vec{X}^{\frac{1}{2}} \, . \end{aligned} \tag{1}$$

and
$g^2(\vec{X}, \vec{U}) = < \log_{\vec{X}}(\vec{U}), \log_{\vec{X}}(\vec{U}) >_{\vec{X}} = trace\left(\log^2(\vec{X}^{-\frac{1}{2}} \vec{U} \vec{X}^{-\frac{1}{2}})\right) \, .$

Tangent Space (3)

Vectorization

The tangent space allows us to vectorize de result of the inverse mapping in order to **work in a vector space with Euclidean distances** which are approximations of the geodesics:

$$\begin{aligned} vec_{\vec{X}}(u) &= vec_{\vec{I}}(u)(\vec{X}^{-\frac{1}{2}} u \vec{X}^{-\frac{1}{2}}) \,. \\ vec_{\vec{I}}(u) &= (u_{11} \; \sqrt{2}u_{12} \; \ldots \; u_{22} \; \sqrt{2}u_{23} \; \ldots \; u_{dd})^T \,. \end{aligned} \tag{2}$$

Once we have a vector **we can use distributional dissimilarity measures for comparing both graphs!** However we must find in advance the **less biased location** for placing the tangent space.

Tangent Space (4)

Karcher Mean

Let us denote by $\vec{Z}_k$ with $k = 1, \ldots, N$ (where $N = n_X + n_Y$) each covariance matrix coming from graphs X or from Y with $n_X = |V_X|$ and $n_Y = |V_Y|$.

- A fair selection of the tangent space origin is the **Karcher mean** defined as

$$\mu = \arg\min_{\vec{Z} \in \mathcal{M}} g^2(\vec{Z}_k, \vec{Z}).$$

- The Karcher mean can be obtained after few iterations of $\mu^{t+1} = \exp_{\mu^t}(\vec{X}^t)$ where $\vec{X}^t = \frac{1}{N} \sum_{k=1}^{N} \log_{\mu^t}(\vec{Z}_k)$.
- Then we deproject and compute the vectorization before comparing distributions representing both X and Y.

Henze-Penrose Divergence

Definition [Henze & Penrose., Annals of Stat.'99]

$$D_{HP}(f||g) = \int \frac{p^2 f^2(z) + q^2 g^2(z)}{pf(z) + qg(z)} dz \ , \tag{3}$$

where $p \in [0,1]$ and $q = 1 - p$. This divergence is the limit of the Friedman-Rafsky run length statistic, that in turn is a generalization based on MST of the Wald-Wolfowitz test for $d > 1$:

1. Build the MST over the samples from both X and O.
2. Remove the edges that do not connect a sample from X with a sample from O.
3. The proportion of non-removed edges converges to 1 minus the Henze Penrose divergence between f_X and g_O.

Henze-Penrose Divergence (2)

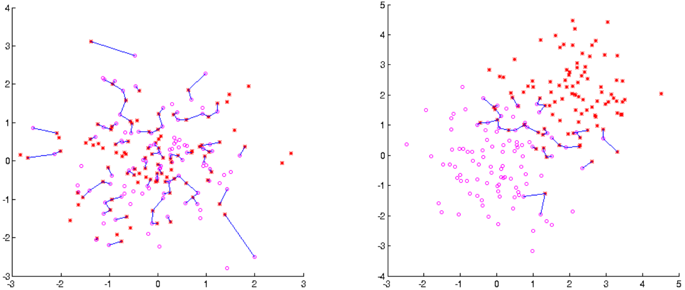


Figure: Two examples of Friedman-Rafsky estimation of the Henze and Penrose divergence applied to samples drawn from two Gaussian densities. Left: the two densities have the same mean and covariance ($D_{HP}(f||g) = 0.5427$). Right: different means ($D_{HP}(f||g) = 0.8191$).

Total Variation k-dP Divergence

Definition [Escolano et al., SPR'10]

Let $f(x)$ and $g(x)$, from which we draw a set X of n_x samples and a set O of n_o samples, respectively. If we apply the partition scheme of the k-d partition algorithm [Stowell and Plumbey., IEEE-SP'09] to the set of samples $X \bigcup O$, the result is a partition A of $X \bigcup O$, being $A = \{A_j | j = 1, \ldots, p\}$. For $f(x)$ and $g(x)$ we have

$$f(A_j) = \frac{n_{x,j}}{n_x} = f_j, \;\; g(A_j) = \frac{n_{o,j}}{n_o} = g_j \; .$$

The k-dP total variation divergence is then given by

$$D_{kdP}(f||g) = \frac{1}{2} \sum_{j=1}^{p} |f_j - g_j|$$

Total Variation k-dP Divergence (2)

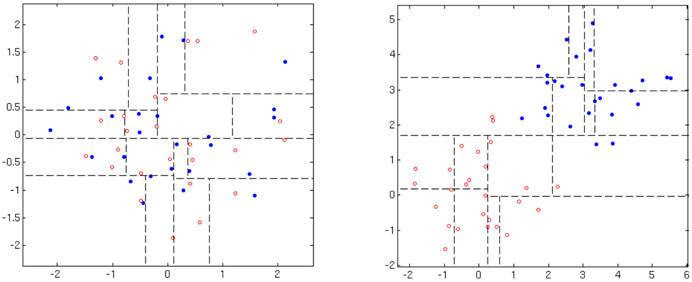


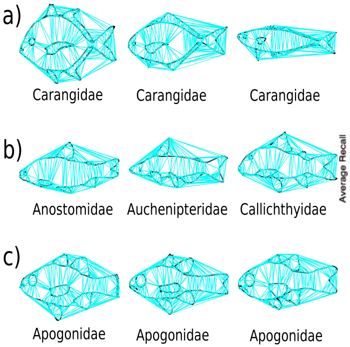
Figure: Two examples of divergence estimation applied to samples drawn from two Gaussian densities. Left: both densities have the same mean and covariance matrix ($D(f||g) = 0.24$). Right: different means. Almost all the cells contain samples obtained from only one distribution ($D(f||g) = 0.92$).

GatorBait100 Database

http://www.cise.ufl.edu/anand/publications.html

- 100 shapes with fishes from 30 different classes (genus) quantized to form Delaunay graphs.
- High intraclass variability in many cases and many similar species from different genus and few homogeneous classes.
- There are 10 classes with one species (not included in the analysis and curves), 11 with $1-3$ individuals, 5 with $4-6$ ones and only 4 classes with more than 6 species.
- Hence, it is hard to devise a mesure which produces an average recall curve far above the diagonal. HP improves KDP.
- $d=5$, where the 5D setting is selected experimentally since intrinsic dimensions are overestimated in this case.

Performance Analysis



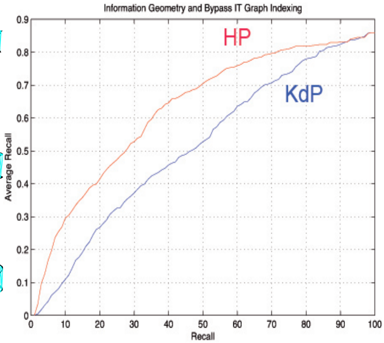


Figure: Left: Gator Samples. Right: Average recall curves

Performance Analysis (2)

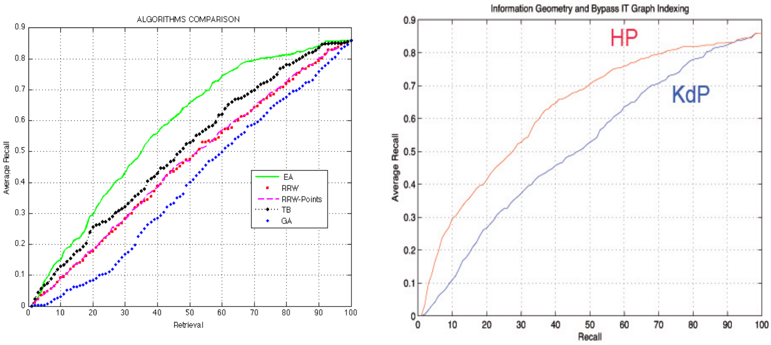


Figure: Left: [Escolano, Hancoch & Lozano, CVPR'11] . Right: Average recall curves

Selecting Optimal Scale

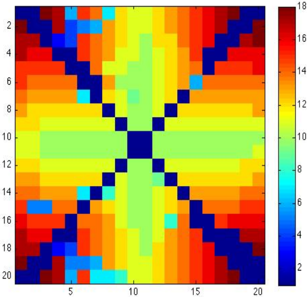
Scale Space for Graphs

- The history of a node is the basis of a scale space as we increase the degree of the coverage.
- Not all nodes of the graph need a maximum degree of coverage.
- What is the optimal degree k_i^* for each node $i \in V$?

Node Saliency on Graphs

- As we have covariance matrices we can compute $det(k\vec{X}_i^k) - \mu(trace(k\vec{X}_i^k)^2$.
- If we detect a maximum in the scale space between k and $k+1$ then $k^* = k$. Otherwise $k^* = 1$.

Selecting the Optimal Scale



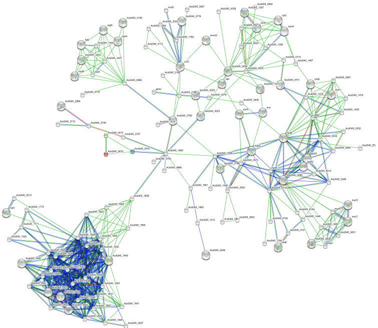


Figure: Left: Scale Space Analysis for a Grid graph with 8−neighborhood topology. Right: Histidine-kinase PPI.

Discussion and Future Work

Discussion

- We have defined an information geometric approach for graph indexing.
- The more competitive IT measure is the Henze-Penrose divergence improving entropic manifold alignment for Gator.

Ideas for the Future

- Introduced optimal scale selection on structure. Only practical for small graphs or for analyzing specific nodes in PPIs.
- The purpose of analyzing these specific nodes is to track the evolution of functional specialization.