

Regularization Kernels and Softassign

Miguel Angel Lozano and Francisco Escolano

Robot Vision Group,
Departamento de Ciencia de la Computación e Inteligencia Artificial,
Universidad de Alicante, Spain
{malozano,sco}@dccia.ua.es
<http://rvg.ua.es>

Abstract. In this paper we analyze the use of regularization kernels on graphs to weight the quadratic cost function used in the Softassign graph-matching algorithm. In a previous work, we have showed that when using diffusion kernels on graphs such a weighting improves significantly the matching performance yielding a slow decay with increasing noise. Weights, relying on the entropies of the probability distributions associated to the vertices after diffusion kernel computation, transform the original unweighted matching problem into a weighted one. In this regard, as diffusion kernels are a particular case of regularization kernels it is interesting to study the utility of this family of kernels for matching purposes. We have designed an experimental set for discovering the optimal performance for each regularization kernel. Our results suggest that kernel combination could be a key point to address in the future.

1 Introduction

The Softassign algorithm [4] is a typical energy minimization approach for graph matching that relies on transforming the discrete search space into a continuous one and then optimizing a quadratic cost function through a polynomial computational complexity process in order to find a, typically approximate, solution. It has been reported [5] that the use of an alternative non-quadratic energy function complemented by a continuous editing process yields a slow decay of matching performance with increasing graph corruption (noise). However, in a recent paper [7] we report a similar decay with a simpler strategy consisting on computing structural attributes for the vertices of the original unweighted graph and then use these attributes to weight the quadratic function. Such good results are due to the fact that these attributes encode the structural similarities between each vertex and the rest of vertices of the graph, and such information is key to choose the proper attractor in contexts of high matching ambiguities where the classical Softassign fails, specially at higher levels of distortion.

Kernels on graphs provide a natural way of computing similarities between the vertices of the same graph. In the case of the diffusion kernel, the one we used in our initial experiments, Kondor and Lafferty [6] (see also [3] for a survey on kernels for structures like strings, trees and graphs) transferred to the domain

of graphs the well-known concept of *kernel* defined for the domain of vectors [2][10][8]. Diffusion kernels on graphs are positive semi-definite matrices (it is necessary to be a kernel or Gram matrix) having as many rows and columns as the number of vertices in the graph (like the adjacency matrix) where the value of each entry defines the similarity between two vertices and such a similarity decays exponentially with the *distance* between the vertices that we are comparing. The similarity can be interpreted as the probability of reaching one vertex from another by following a *lazy* random walk (a random walk with a given probability of remaining at each visited vertex). Given such a probabilistic interpretation, we retain as attribute for a given vertex the entropy of the probability distribution arising from considering the probability of reaching the rest of the nodes and the probability of resting. This approach is closely related to the use of distance matrices in matching and tests for isomorphism [9], and, more recently, to the use of powers of the adjacency matrix [13].

Although understanding the role of the diffusion kernel in graphs is intuitive because these kernels are the discrete version of continuous Gaussian kernels, this is not the case for the regularization kernels recently proposed by Smola and Kondor [12]. These latter kernels are derived from studying the usefulness of the Laplacian of a graph (and its normalized version) as a smoothing operator (section 2). Considering smoothing operators from a spectral point of view, it results that a family of kernels emerges from considering different penalization functions. It can be proved that the inverse of the so called *regularization matrix* for each element of the family yields a kernel (actually, the diffusion kernel belongs to this family). In (section 3) we review the weighted energy function and its minimization through the Softassign continuation process. In order to build the attributes of the vertices we consider that the similarities defined by each kernel induce a probability distribution and, in order to characterize such a distribution, we use its entropy. In (section 4) we evaluate the matching performance of each kernel in different conditions of noise and edge connectivity.

2 Regularization Kernels on Graphs

Given a undirected and unweighted graph $G = (V, E)$ with vertex-set V of size m , and edge-set $E = \{(i, j) | (i, j) \in V \times V, i \neq j\}$, its adjacency matrix and degree matrix are respectively defined as

$$A_{ij} = \begin{cases} 1 & \text{if } (i, j) \in E \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad D_{ij} = \begin{cases} \sum_{j=1}^m A_{ij} & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}.$$

Then, the Laplacian of G , $L = D - A$, and its degree-normalized version, $\tilde{L} = D^{-\frac{1}{2}} L D^{-\frac{1}{2}} = I - D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$ or Normalized Laplacian [1], are defined as

$$L_{ij} = \begin{cases} -1 & \text{if } (i, j) \in E \\ D_{ii} & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad \tilde{L}_{ij} = \begin{cases} -D_{ii}^{-\frac{1}{2}} D_{jj}^{-\frac{1}{2}} & \text{if } (i, j) \in E \\ 1 & \text{if } i = j \text{ and } D_{ii} \neq 0 \\ 0 & \text{otherwise} \end{cases}.$$

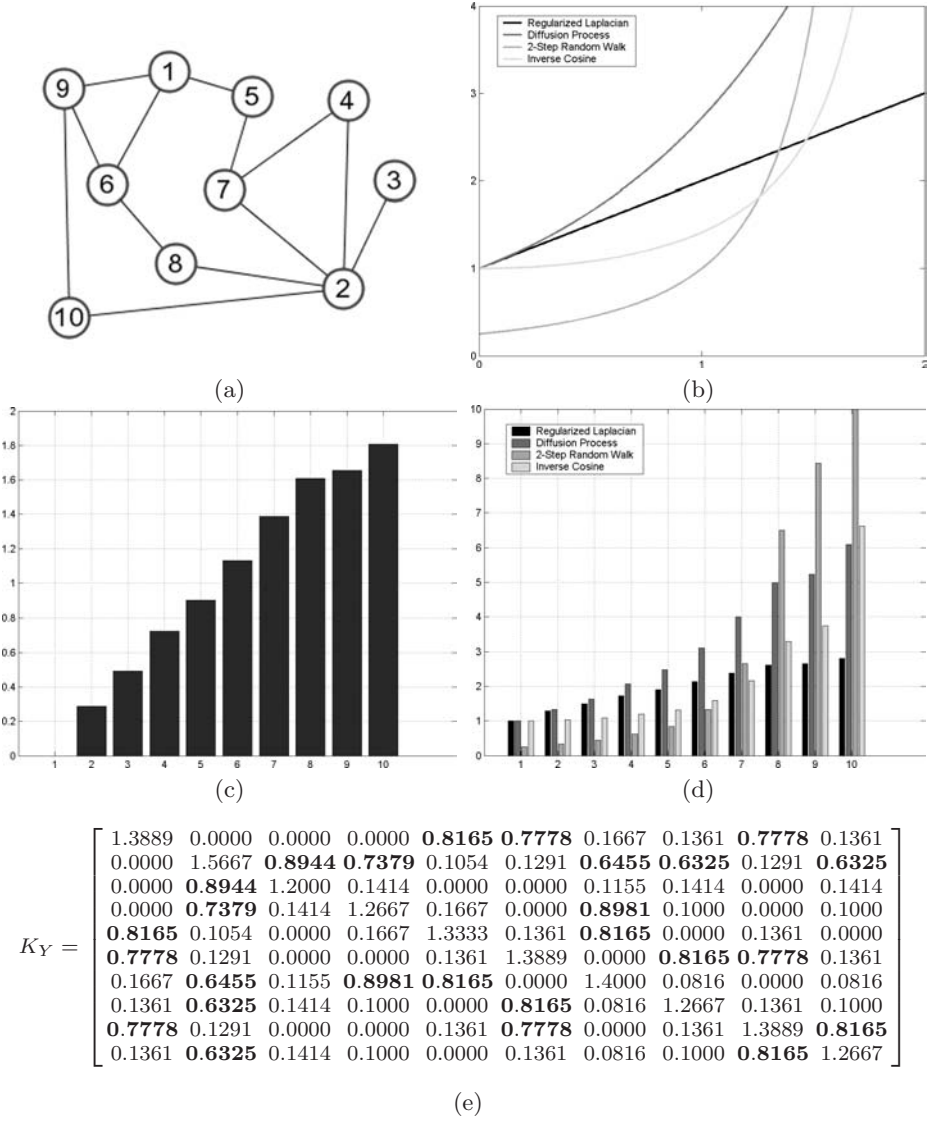


Fig. 1. Illustrating regularization kernel computation. (a) Graph with $m = 10$ nodes. (b) Penalizing functions corresponding to Table 1. (c) Eigenvalues of the Normalized Laplacian sorted in ascending order and contained in $[0, 2]$. (d) Result of applying the penalizing functions on the graph spectrum. (e) Resulting p -step kernel with $a = 2$ and $p = 2$.

The connection of the latter Laplacian matrices with regularization theory stands from the fact that given a real-valued function f defined over the vertices of G , that is, $f : V \rightarrow \mathbb{R}$ both, L and \tilde{L} can be seen as discrete differential operators which tend to penalize changes of f between adjacent edges. Considering

now f a column vector, that is $f \in \mathbb{R}^m$, the following inner product is a measure of the smoothness of f over the graph G :

$$\langle f, Lf \rangle = f^T Lf = \sum_{i=1}^m f_i^2 D_{ii} - \sum_{(i,j) \in E} 2f_i f_j = \sum_{(i,j) \in E} (f_i - f_j)^2,$$

and the result is called the Dirichlet sum of G . Furthermore, given $g = D^{\frac{1}{2}}f$ the Dirichlet sums associated to L and \tilde{L} are related by:

$$\langle f, Lf \rangle = (D^{-\frac{1}{2}}g)^T L(D^{-\frac{1}{2}}g) = g^T \underbrace{D^{-\frac{1}{2}}LD^{-\frac{1}{2}}}_{\tilde{L}}g = \langle g, D^{-\frac{1}{2}}LD^{-\frac{1}{2}}g \rangle = \langle g, \tilde{L}g \rangle$$

An alternative way of formulating regularization is through spectral analysis. In [11], Smola et al established the connection between regularization, Fourier analysis and kernels in continuous spaces. A smoothness operator in Fourier space can be built by multiplying the Fourier transform by a penalizing function increasing in frequency. As such a multiplication in Fourier space becomes the application of the latter function on the continuous Laplacian operator, an spectral-based regularization operator in graphs comes from

$$\langle f, r(\mathcal{L})f \rangle = f^T \underbrace{\left[\sum_{i=1}^m r(\lambda_i) \phi_i \phi_i^T \right]}_{r(\mathcal{L})} f = \sum_{i=1}^m \langle f, \phi_i \rangle r(\lambda_i) \langle \phi_i, f \rangle,$$

where \mathcal{L} denotes both L and \tilde{L} , $\{\lambda_i, \phi_i\}$ are the eigenvalues and eigenvectors of \mathcal{L} , and $r(\lambda_i)$ is a monotone increasing function. Actually, $r^{-1}(\lambda)$, the inverse of such a function is the Fourier transform of the associated kernel in the continuous case, and the discrete regularization kernel K is the inverse (or the pseudo-inverse if necessary) of the so called *regularization matrix* $r(\mathcal{L})$. Then we have that

$$K = r^{-1}(\mathcal{L}) \text{ where } r^{-1}(\mathcal{L}) = \sum_{i=1}^m r^{-1}(\lambda_i) \phi_i \phi_i^T, \quad (1)$$

and $0^{-1} = 0$. For instance, in the particular case of the diffusion kernel, which relies on *matrix exponentiation* but not on componentwise exponentiation, we have that

$$K = e^{-\beta \mathcal{L}} = \left(\sum_{i=1}^m e^{\beta \lambda_i} \phi_i \phi_i^T \right)^{-1} = \sum_{i=1}^m e^{-\beta \lambda_i} \phi_i \phi_i^T.$$

In the general case, the relation $K = r^{-1}(\mathcal{L})$ is derived from the fact that given a regularization operator, for instance $M = r(\mathcal{L})$, the matrix K must satisfy the *self-consistency condition* $KMK = K$ to be a kernel, and therefore $K = M^{-1}$ or equal to the pseudo-inverse if M is not invertible. Furthermore, it can be proved [12] that such regularization operator defines a reproducing kernel Hilbert space whose kernel is $K = M^{-1}$. In table 1 we show several penalization functions and their associated regularization kernels. The process of obtaining a regularization kernel is summarized in Fig. 1. As it can be seen, choosing \tilde{L} yields and spectrum contained in $[0, 2]$.

3 Kernelizing Softassign

A feasible solution to the graph matching problem between G_X and G_Y , with adjacency matrices X_{ab} , and Y_{ij} , is encoded by a matrix M of size $m \times n$, being $m = |V_X|$ and $n = |V_Y|$. Following the Gold and Rangarajan formulation we are interested in finding the feasible solution M that minimizes the following cost function,

$$F(M) = -\frac{1}{2} \sum_{a=1}^m \sum_{i=1}^n \sum_{b=1}^m \sum_{j=1}^n M_{ai} M_{bj} C_{aibj}, \quad (2)$$

where typically $C_{aibj} = X_{ab}Y_{ij}$. Furthermore, considering the entropies defined in the previous section a simple way of *kernelizing* the latter energy function is to redefine C_{aibj} as

$$C_{aibj}^K = X_{ab}Y_{ij} \exp -[(H_a^{K_X} - H_i^{K_Y})^2 + (H_b^{K_X} - H_j^{K_Y})^2], \quad (3)$$

where H^{K_X} and H^{K_Y} are the entropies of the probability distributions associated to the vertices of the graph and induced respectively by K_X and K_Y . Given a pair of vertices, for instance a, b of graph G_X the kernel K^X induces the following probability distribution

$$p_{ab}^X = K_{ab}^X \left(\sum_{c=1}^m K_{ac}^X \right)^{-1} \quad \text{and} \quad H_a^{K_X} = \sum_{b=1}^m p_{ab}^X \log p_{ab}^X,$$

and the same holds for p_{ij}^Y and $H_i^{K_Y}$ in case of vertices i, j of G_Y . We use entropy because building attributes in the properties of distributions yields more robustness than building them in the crude values of the kernels.

The latter definition of C_{aibj}^K ensures that $C_{aibj}^K \leq C_{aibj}$, and the equality is only verified when nodes a and i have similar entropies, and the same for nodes b and j . In practice, this weights the rectangles in such a way that rectangles with compatible entropies in their opposite vertices are preferred, and otherwise they are underweighted and do not attract the continuation process.

To see intuitively the difference between two kernels in Fig. 2 we show the results obtained both for the cosine and the p-step kernel. We show the matching strength prior to performing clean-up. Cosine kernel is very ambiguous (similar to the classical Softassign) whereas the p-step one finds the most coherent subgraph in terms of structural similarity.

Table 1. Penalization Functions and Regularization Kernels.

$r(\lambda)$	$K = r^{-1}(\mathcal{L})$	Name
$1 + \beta\lambda$	$(I + \beta\mathcal{L})^{-1}$	Regularized Laplacian
$e^{\beta\lambda}$	$e^{-\beta\mathcal{L}}$	Diffusion Process
$(a - \lambda)^{-p}$	$(aI - \mathcal{L})^p$	p -step Random Walk
$(\cos \lambda\pi/4)^{-1}$	$\cos \mathcal{L}\pi/4$	Inverse Cosine

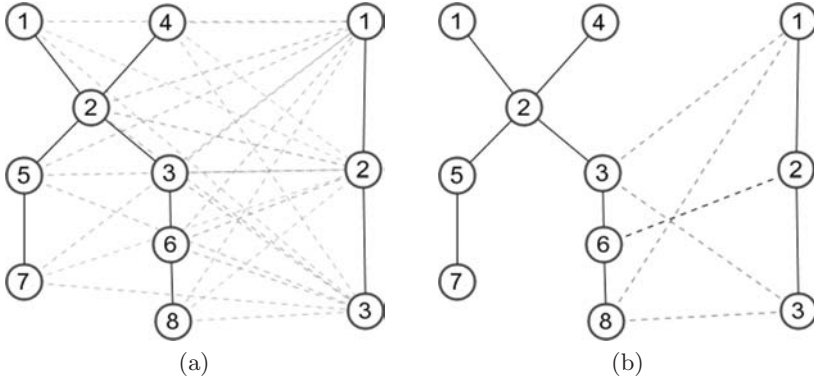


Fig. 2. Illustrating matching differences between the cosine kernel (a) and the p-step one (b).

4 Experiments

We have performed several matching experiments with graphs of 30 nodes, considering three levels of edge density (10%, 30% and 50%), and six different noise levels (0% or isomorphism, 10%, 20%, 30%, 40%, and 50%). We have also considered both the Laplacian L and the Normalized Laplacian \tilde{L} for kernel computation. In all cases $\beta = 1$ unless we specify that β is normalized by the number of nodes of the graph. In all experiments we compare kernels with the classical Softassign driven by degree similarity. Furthermore, in the plots (see Fig. 3) each point corresponds to the averaged result for 100 graphs randomly generated. We have registered the fraction of complete graphs successfully matched.

Analyzing the obtained results, at low edge densities (10%) we observe that cardinality outperforms all kernels at zero noise (isomorphism). However, kernels, specially p-step with \tilde{L} , yield slower decays as the noise increases. P-step has a similar behavior at 30% edge density, although the diffusion kernel with L and normalized β is the best choice even at isomorphism. This latter kernel yields the slower decay for a density of 50% although p-step with \tilde{L} is only acceptable for a corruption percentage greater than 30%. We conclude that p-step is a good choice for low densities, whereas the diffusion one with L and normalized β is preferable for higher densities.

5 Conclusions and Future Work

In this paper we have analyzed the use of regularization kernels in the context of graph matching, particularly in the Softassign algorithm. We have studied the performance of several kernels in different noise and edge-density conditions and we conclude that p-step is good for low densities and diffusion is good for higher densities. This is why in the future we will develop an strategy for combining them. Another question in which we are currently working is how to formulate a pure kernelized energy function and its associated matching algorithm.

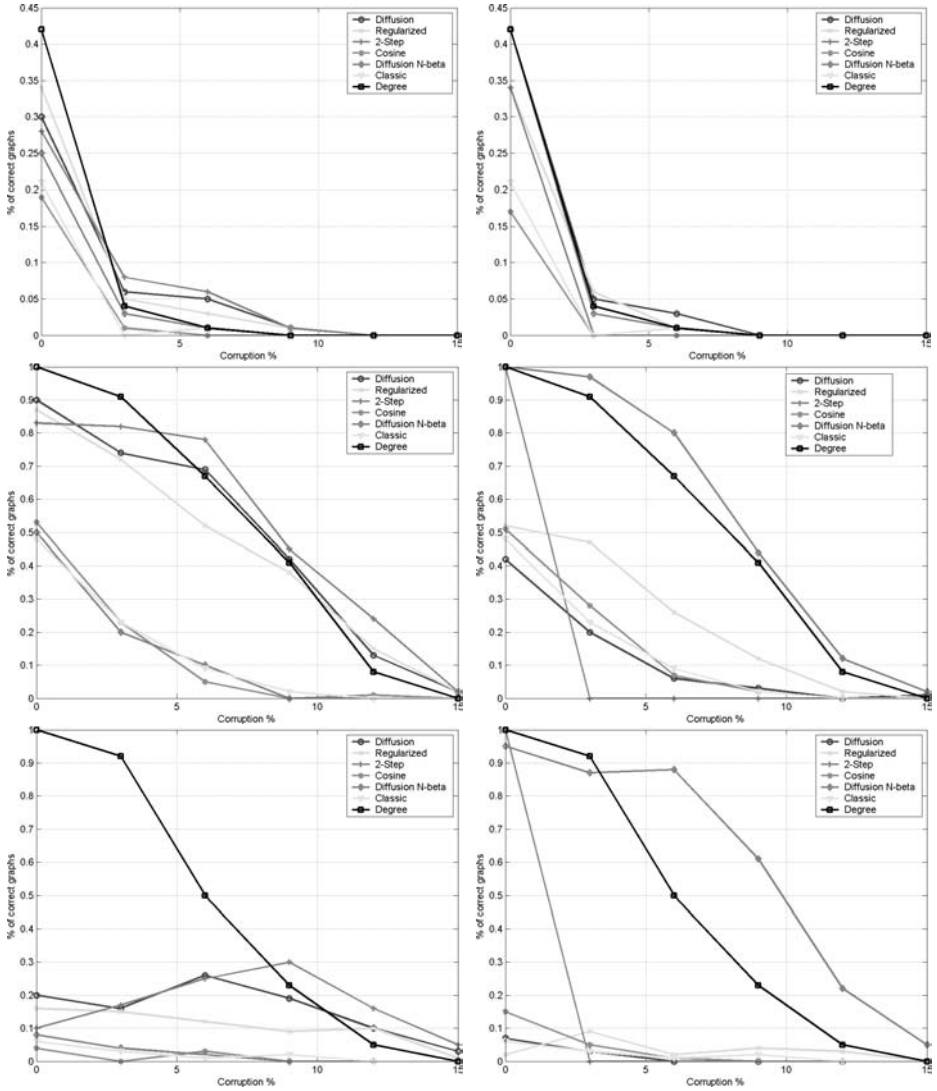


Fig. 3. Matching results. Graphs successfully matched. Left column: results with the Normalized Laplacian \tilde{L} . Right column: results with the classical Laplacian L . Each row corresponds to a different edge density: 10%, 30% and 50%.

Acknowledgements

This work was partially supported by grant *TIC2002 – 02792* funded by *Ministerio de Ciencia y Tecnología* of the Spanish Government and by *FEDER*.

References

1. Chung, F.R.K.: Spectral Graph Theory. Conference Board of the Mathematical Sciences (CBMS) **92**. American Mathematical Society (1997)
2. Cristianini, N., Shawe-Taylor, J.: An Introduction to Support Vector Machines Cambridge University Press (2000)
3. Gärtner: A Survey of Kernels for Structured Data. ACM SIGKDD Explorations Newsletter **5**(1) (2003) 49–58
4. Gold, S., Rangarajan, A.: A Graduated Assignment Algorithm for Graph Matching. IEEE Transactions on Pattern Analysis and Machine Intelligence **18** (4) (1996) 377–388
5. Finch, A.M., Wilson, R.C., Hancock, E.: An Energy Function and Continuous Edit Process for Graph Matching. Neural Computation, **10** (7) (1998) 1873–1894
6. Kondor, R.I., Lafferty, J.: Diffusion Kernels on Graphs and other Discrete Input Spaces. In: Sammut, C., and Hoffmann, A. G. (eds) Machine Learning, Proceedings of the Nineteenth International Conference (ICML 2002). Morgan Kaufmann (2002) 315–322
7. Lozano, M.A., Escolano, F.: A Significant Improvement of Softassign with Diffusion Kernels. In Proceedings of the IAPR International Workshop on Syntactical and Structural Pattern Recognition SSPR 2004. Lecture Notes in Computer Science (2004)(accepted for publication)
8. Müller, K.-R., Mika, S., Räsch, Tsuda, K., Schölkopf, B.: An Introduction to Kernel-based Learning Algorithms. IEEE Transactions on Neural Networks, **12**(2) (2001) 181–201.
9. Schmidt, D.C., Druffel, L.E. : A Fast Backtracking Algorithm to Test Direct Graphs for Isomorphism Using Distance Matrices. Journal of the ACM **23** (3) (1976) 433–445
10. Schölkopf, B., Smola, A.: Learning with Kernels. MIT Press (2002).
11. Smola, A., Schölkopf, B., Müller, K.-R.: The Connection between Regularization Operators and Support Vector Kernels. Neural Networks **11** (1998) 637–649
12. Smola, A., Kondor, R.I.: Kernels and Regularization on Graphs. In: Schölkopf, B., and Warmuth, M. K. (eds) Computational Learning Theory and Kernel Machines, 16th Annual Conference on Computational Learning Theory and 7th Kernel Workshop, COLT/Kernel 2003. Lecture Notes in Computer Science. Springer. Vol. **2777** (2003) 144–158
13. DePiero, F.W., Trivedi, M., Serbin, S.: Graph Matching Using a Direct Classification of Node Attendance. Pattern Recognition, Vol. **29**(6) (1996) 1031–1048
14. Ozer, B., Wolf, W., Akansu, A.N.: A Graph Based Object Description for Information Retrieval in Digital Image and Video Libraries. In: Proceedings of the IEEE Workshop on Content-Based Access of Image and Video Libraries (1999) 79–83