Original Research Paper

Volume -10 | Issue - 5 | May - 2020 | PRINT ISSN No. 2249 - 555X | DOI : 10.36106/ijar

Mathematics

CONVERGENCE TO A DIFFUSION PROCESSES ON GRAPHS

B. Malathi — Department of Mathematics, The M.D.T Hindu College, Affiliated to Manonmaniam Sundaranar University, Tirunelveli-10

S. Chelliah* — Department of Mathematics, The M.D.T Hindu College, Affiliated to Manonmaniam Sundaranar University, Tirunelveli-10 *Corresponding Author

ABSTRACT Nowadays, graph representation learning has aroused a lot of research interest, which aims to learn the latent low dimensional representations of graph nodes, while preserving the graph structure. In many examples distances are equivalent to a snow flake of the natural distance on space. Based on the local smooth assumption, some existing methods have achieved significant success. Diffusion in narrow tubes processes with fast transmutations and convergence to a diffusion process on a graph. To alleviate this issue propose a graph Diffusion Network that can dynamically preserve local and global consistency of graph.

KEYWORDS : Graph representation learning, diffusion on graphs

INTRODUCTION

In graph theory the number of lines meeting at a vertex, i.e., incident to that vertex, is called the vertex degree; graphs whose vertex degrees are all equal are called regular graphs. A different approach was initiated by H. Wiener in 1947 [1-9] with topological indices (TI's). Theses TI's are numbers associated with chemical structures via their hydrogen-depleted graphs G. For hydrocarbons, the wiener index W is the sum of the number of bonds between all pairs of vertices in G. If one defines the (topological) distance between two vertices of a graph as the number of bonds between along the shortest path between these two vertices, then W is the sum of all distances in graph G. One can associate with any graph on n vertices several matrices: the adjacency matrix A(G) is a square, symmetrical, n×n matrix with entries $\alpha_{ij}=1$ for adjacent (directly bonded) vertices i and j and zero otherwise; the distance matrix D(G) is also a square n×n matrix with entries $d_{ii}=0$ on the main diagonal and $d_{ij}=1$ for adjacent vertices i and j as in A, but all other entries are integers bigger than 1 and represent the topological distance between vertices i and j. The sums over rows i or columns i for A(G) indicate the vertex degrees v_i ; the sums over rows i or columns i for D(G) indicate another graph invariant for each vertex (invariant from the arbitrary vertex labeling $i\epsilon\{0,1,\dots n\}$), called distance sum s_i . It is easy to see that $W=\sum_i {}^{s_i}/_2$ .Graph representation learning aims to represent each node in a graph as a low –dimensional vector that could facilitate tasks such as node . The structural analysis is an important problem motivating many studies of real – world networks. For instance, the efficient partitioning of a transport network is one of the most effective logistics optimization tools and the biggest opportunity to significantly reduce transportation costs. The structural analysis is also of importance when dealing with extremely large graphs, when we need to cluster the vertices into logical components for storage (to improve virtual memory performance) or for drawing purposes (to collapse dense subgraphs into single nodes in order to reduce cluttering). Finally, the structural analysis is important for accessing large databases. In the present chapter, we apply the methods related to random walks for analyzing of urban structures, evolution of languages, and musical compositions. An electrical network is considered as an interconnection of resistors. We demonstrate that random walks defined on connected undirected graphs have a profound connection to electric resistor networks (Doyle and Snell 1984; Tetali 1991; Chandra et al. 1996; Bollobas 1998)[11-15]. In the present chapter, we discuss the effective resistance of electrical networks, the relation between the shortest path (geodesic) distance and the effective resistance distance, Kirchhoff and Wiener indexes of a graph.

Fixed point theory is an important area of functional analysis. This chapter deals with the survey of literature, related to theory of fixed point theorems. Fixed point theory has fascinated thousands of researcher since1922 with the celebrated Banach's fixed point theorem. There exists vast literature on this topic and it is a very active field of research at present. A self map T of a metric space X is said have a fixed pointx if Tx=x. Theorems concerning the existence and properties of fixed points are known as fixed point theorems. Such theorems are very important tools for proving the existence and uniqueness of the solutions to various mathematical models representing phenomena arising in different fields, such as steady state temperature distribution, chemical equation, economic theories and flow of fluids. They are also used to study the problems of optimal control related to these systems.

A finite connected undirected graph G(V, E) can be seen as a *discrete time dynamical system* possessing a finite number of states (nodes) (Prisner 1995). The behavior of such a dynamical system can be studied by means of a transfer operator which describes the time evolution of distributions in phase space. The transfer operator can be represented by a stochastic matrix determining a discrete time random walk on the graph in which a walker picks at each node between the various available edges with equal probability[16-22]. An obvious benefit of the approach based on random walks to graph theory is that the relations between individual nodes and subgraphs acquire a precise quantitative probabilistic description that enables us to attack applied problems which could not even be started otherwise.

2. Related works

In the following, we will mainly provide a brief review on some related works.

2.1. Definition

Expected number of fixed points:

We have $C_n = \frac{1}{n!}\sum_{\sigma\epsilon S_n} x_i^{a_i(\sigma)}$

Then we defined $C(t) = \sum_{n=0}^{\infty} t^n C_n$

$= e^{\sum_{n=0}^{\infty} x_i \frac{t^i}{i}}$

This has a lot of information in it.

2.2. Example

Suppose $C(\sigma)$ is the number of cycle of σ.Then,

$C(\sigma)=\sum_{i=1}^{n} a_i(\sigma)$

Setting all $x_i = x$ we have

$$C_n(x) = \frac{1}{n!}\sum_{\sigma} x_0^{C(\sigma)}$$

Then $C(t) = \frac{1}{(1-t)^x}$

$= \sum_{j=0}^{\infty} \binom{-x}{j} (-t)^j$

$= \sum_j \frac{t^j}{j!} x(x+1)\dots(x+j-1)$

Since $\sum_i \frac{t^i}{i}$ is the power series expansion for $-\log(1-t)$

Therefore, $C_n = \frac{1}{n!}$ x(x+1)…(x+n-1)

= x($\frac{x+1}{2}$)($\frac{2+x}{3}$)….($\frac{n-1+x}{n}$)

=E(x^{S_n}) = $\pi E_x^{x_i}$

Here S_n denotes the sum, not the symmetric group. Here P(x_i=0)=$\frac{i-1}{i}$ and P(x_i=1)=$\frac{1}{i}$

E(x^{S_n}) = $\sum_{j=0}^{n} x^j \, P(x_n = j)$

E(f(s_n) = $\sum f(j)$P($S_n = j$)

We took f(j)=x^j

So, AV(S_n) = $1 + \frac{1}{2} + \cdots + \frac{1}{n} \sim \log n$

VAR(S_n)= $\sum_{n=1}^{n} \frac{1}{i}(1 - \frac{1}{i}) \sim logn$

P$\left\{\frac{C(\sigma) - \log n}{\sqrt{\log n}} \leq x\right\} \to \varphi(x)$

The coefficient of x_j is the number of permutations with j cycles. These happen to be called sterling numbers of the first kind.

2.3. Question

Who cares about all this stuff with fixed points?

There was a game played where someone took two decks of cards up to n. People play this game and you get a dollar if the same number comes up. The question is a question of the number of fixed points Monmort in 1708 proved the number of fixed points has a poission distribution as we proved last time. Note that we may as well call the cards on the first deck 1, 2…n. So the number of matches is just the number of fixed points in a random permutation.

We also have a metric,

D (π, σ)=♯{i:$\pi(i) \neq \sigma(i)$}

See, Diaconis, Gorolnick and Mulman on fixed points of permutations for a classification of possible fixed points of transitive primitive actions of the symmetric group.

2.4. Definition

The Cayley distance between two permutations

$d_c(\sigma, \pi) =$
$minimum\ number\ of\ transpostions\ needed\ to\ express\ \pi\sigma^{-1}$

ie., this is the distance in the cayley graph where the vertices are permutations and the edges join two elements differing by a permutations.

2.5 Remarks

The above two distance measures are the only two biinvariant distances that persi knows of.

2.6. Definition

Graph diffusion, which is equivalent to linear weighting for nodes by large-scale random walk on graph. However, Graph Diffusion focuses on node-level transformations rather than content-level transformations. To reveal the relationship between node features, we consider looking for a function $f(A) = \sigma(AC + x) to\ non\ linearly\ map\ A$ from the input space to the representation space, where A=A_0 is original node features and C∈ $R^{d \times h}$ is the transformation matrix. Combined with Graph diffusion, single diffusion can be expressed as follows:

$f(B) = \sigma(BC + x)$
$=\sigma((I - \delta S)^{-1}AC + x).$

2.7. Diffusion

A particular important issue in harmonic analysis is to connect the smoothness of a function with the speed of convergence of its diffused version to itself, in the limit as time goes to zero. In order to consider the smoothness of diffusing functions in more general settings, a distance defined in terms of the diffusion itself seems particularly appropriate.

Defining diffusion distances is of interest in applications as well. As discussed in [3, 7, 12, 13, 25], dimensionality reduction of data and the concomitant issue of finding structures in data are highly important objectives in the fields of information theory, statistics, machine learning, sampling theory, etc. It is often useful to organize the given data as nodes in a weighted graph, where the weights reflect local interaction between data points. Random walks, or diffusion, on graphs may then help understand the interactions among the data points at increasing distance scales. To even consider different distance scales, it is necessary to define an appropriate diffusion distance on the constructed data graph.

2.8. Properties

The symmetric diffusion operator T_t has the following properties of a symmetric diffusion semi group:

(i)T_0 is the identity
(ii)$T_{t+s} = T_t \bullet T_s$, for all s, t≥ 0
(iii)$\|T_t(f)\|_{L_p} \leq \|f\|_{L_p}$, for 1≤ $p \leq \infty$
(iv)T_t is a self ad joint operator on $L_2(X)$
(v) $T_t(f) \to f\ in\ L_2, as\ t \to 0^+$
(vi) $T_t(f) \geq 0$ if $f \geq 0$
(vii)$T_t(1) = 1$

See Stein's book, in which the author derives various harmonic analysis results for symmetric diffusion semi groups without explicitly using kernels.

2.9. Definition

We consider a general symmetric diffusion semigroup $\{T_t f\}_{t \geq 0}$ on a topological space X with a positive σ-finite measure (i.e., X is a countable union of measurable sets with finite measure), given, for t> 0, by an integral kernel operator:$T_f f(x) \triangleq \int_x \rho_t(x, y) f(y) dy$.

Coifman and Leeb introduce a family of multiscale diffusion distances and establish quantitative results about the equivalence of a bounded function *f* being Lipschitz, and the rate of convergence of $T_t f$ to f, as t→ 0^+ (we are discussing some of their results using a continuous time for t convenience; most of Coifman's and Leeb's derivations are done for dyadically discretized times. Moreover, most of the authors' results are in fact established without the assumption of symmetry and under the weaker condition than positivity of the kernel, namely, an appropriate L_1integrability statement. To prove the implication that Lipschitz implies an appropriate estimate on the rate of convergence, Coifman and Leeb make a quantitative assumption about the decay of

$$\sup_x \int_X |\rho_t(x,y)| d(x,y) dy \, , as \; t \to 0^+ \quad (1)$$

for their distances *d*, namely, that

$$\sup_x \int_X |\rho_t(x,y)| d(x,y) dy \; \leq Ct^{\alpha} \quad (2)$$

for some $\alpha > 0$. Coifman and Leeb also establish that (2) above, in the case of positive diffusion kernels, is in fact equivalent to their conclusion about the rate of convergence of $T_t f \; to \; f$, as $t\to 0^+$, for a Lipschitz function *f*. Additionally, Coifman and Leeb show that, in some of the settings they consider (with decay and continuity assumptions on the diffusion kernels relative to an intrinsic metric), their multi scale diffusion distance is equivalent to (localized)$D(x,y)^{\alpha}$ where *D(x,y)* is the intrinsic metric of the underlying space and α is a positive number strictly less than 1. The authors emphasize that α cannot be taken to equal 1.

We introduce a new family of diffusion distances generated by the diffusion semigroup $\{T_t f\}_{t\geq 0}$. We provide several reasons as to why we think our definition is natural; in particular, we show that, for a convolution diffusion kernel on $\mathbb{R}^n$, we achieve $\alpha = 1$in the discussion just above; i.e., we can recover (local) Euclidean distance to the "full" power 1.

The implication established in [27, 28] that smoothness of *f* implies control of the speed of convergence of $T_t f$ to *f* seems to us to be a more notable result than the converse (which the authors establish without assuming the decay of (1)). However, if *f* is Lipschitz for the multi scale diffusion distance introduced in [29, 36, 37, 63, 74], as the authors themselves point out their assumed estimate (2) almost tautologically leads to the desired estimate for the speed of convergence of $T_t f$ to *f*.

The main reason is that we wish to avoid making any assumptions about the decay of (1) and still establish a correspondence between some version of smoothness of a function *f* and convergence of $T_t f$ to f, as $t\to 0^+$. Our main contribution is to establish, under almost no assumptions, that local equi continuity (in t) is equivalent to local convergence; i.e., local control of the differences $T_t f(x)$-$T_t f(y)$ for all *t* small is equivalent to local control of the differences $T_t f(x)$-*f(x)* for all small t. Here "local" is defined relative to a representative of our family of proposed diffusion distances.

2.10. Theorem

For A>0, Ax=$\delta_{max} x$ are strictly positive.

Proof

The key idea is to look at all number t such that $Ax \geq tx$ for some non negative vector x (other than x=0). We are allowing inequality in Ax≥tx in order to have many positive candidates t. For the largest value t_{max} (Which attained), we will show that equality holds Ax=t_{max}x

Otherwise if Ax≥ $t_{max} x$ is not a equality, multiply by A. Because A is positive that produces a strict inequality $A^2 x > t_{max} Ax$ .Therefore the positive vector $y = Ax$ satisfies Ay>$t_{max} y$ and t_{max} could be increased. This contradiction forces the equality Ax = t_{max}x and we have an eigen value. Its eigen vector x is positive because on the left side of that quality, Ax is sure to be positive.

To see that no eigen value can be larger than t_{max}, suppose Az=δ z . Since δ and z may involve negative or complex number, we take absolute values $|\delta||z| = |Az| \leq A|z|$ by the triangle inequality. This $|z|$ is a non negative vector, so $|\delta|$ is one of the possible candidate t. Therefore $|\delta|$cannot exceed t_{max} which must be δ_{max}.

2.11. Definition

The mathematics of oscillation deals with the quantification of the amount that a sequence or functions trends to move between extremes. There are several related notions. Oscillation of a sequence of real numbers, oscillation of a real valued function at a point and oscillation of a function on an interval(or an open set).

2.12. Theorem

Let $\sum_A$ be topologically mixing $\varphi \in \gamma_A \cap C(\sum_A^+)$ and μ=μ_φ as above. There are $\delta > 0$, $h \in C(\sum_A^+)$ with h>0 and v∈M($\sum_A^+$) for which$\mu h = \delta h$, $\mu^* v = \delta v$, vh=1 and

$\lim_{m\to\infty} \|\delta^{-m}\mu^m g - v(g)h\| = 0$ for all $g \in C(\sum_A^+)$.

2.13. Definition

Given a reported type profile u and a node v with non-nil reported type, define $R_v(u) = \{\cap \mathfrak{D}\}_{\mathfrak{D}\in\mathfrak{D}_i(u)}$ is the set of all feasible trading paths from the seller s to node v.

2.14. Result

Given any type profile x and any node $v \in \mathfrak{D}_i^n(x) \backslash \mathfrak{D}_j^n(x)$ cannot increase her utility by misreporting.

2.15. Result

The weighted diffusion mechanism is individually rational and incentive compatible.

3. Application

Within pure mathematics, graph theory is studied in the pioneering book on topology by veblen. A simplicial complex (or briefly a complex) is defined to consist of a collection v of points together with a prescribed collection S of nonempty subsets of v called simplexes satisfying the following two conditions.

- Every point is a simplex
- Every nonempty subset of a simplex is also simplex.

Simplex is also simplex. The dimension of a simplex is one less than the number of points in it; that of a complex is the maximum dimension of any simplex in it. In these terms, a graph may be defined as a complex of dimension 1 or 0. We call a 1-dimensional simplex a line and note that a complex is 0-dimensional if and only if consists of points, but no lines or other higher dimensional simplexes. A side from these totally disconnected graphs, every graph is a 1-dimensional complex. It is precisely because of the traditional use of the words points and line as undefined terms in axiom systems for geometric structures that we have chosen to use this terminology. Whenever we are speaking of geometric simplicial complexes as subsets of a Euclidean space, as opposed to the abstract complexes, we shall then use the words vertex and edge.

3.1. Markov Process

A number of asymptotic problems for classical stochastic processes leads to diffusion processes on graphs. We study several such examples and develop a general technique for these problems. Diffusion in narrow tubes [10] , processes with fast transmutations and small random permutations of Hamiltonian systems are studied.

Let $X^{\varepsilon}(t)$, $\varepsilon > 0$, be a family of Markov Processes on a space M.It is possible that as $\varepsilon \to 0$ the process $X^{\varepsilon}(t)$ moves faster and faster in some directions, whereas the motion in other directions does not accelerate. This is the situation where one can expect that the so-called averaging principle works: we can identify the points of the space M in the “fast” directions, obtaining a new space Y(M) (Y is the mapping effecting the identification).The “fast” motion “across” Y(M) is not a Markov process in general, but in its “fast” time it is nearly one because the characteristics of the “fast” motion depend on the “slow” variables and vary slowly compared to the “fast” motion itself. The slow process Y($X^{\varepsilon}(t)$) also is not a Markov one ,but the averaging principle means that it converges in some sense to a Markov process Y(t) on Y(M) as $\varepsilon \to 0$, and the characteristics of this limiting process are obtained by averaging the characteristics of the process Y($X^{\varepsilon}(t)$)over the “fast” directions with respect to the stationary distribution of the “fast” Markov process.

3.2. Complex analysis

For a given graph, since the diffusion matrix $(I - \alpha S)^{-1}$is dense , the computational complexity of the diffusion layer is $\mathbb{O}(N^2 d)$.In general, we can approximate the diffusion matix with k th-order ($2 \leq k \leq 4$) expansion , which is sufficient sparse. Although it will lose some performance, the efficiency can be greatly improved to $\mathbb{O}(kd|\varepsilon|)$.IN theory, the efficiency of Graph diffusion network can be at the same level as Graph convolutional Networks.

3.3. Graph Diffusion Network

In General, given a set of iid data points X$\epsilon R^{n\times d}$ in d-dimentional space, we focus on finding an intrinsic representation P=H(X)(P$\epsilon R^{n\times m}$).where H(.) is the function mapping X from the input space to the low dimensional vector space. Thus the key of graph representation learning is to exploit the structure information information more effectively. Due to powerful ability of nonlinear representation, the deep learning architecture is increasingly used for graph representation learning. Experimental results on node classification tasks demonstrate the effectiveness of the proposed Graph Diffusion Network model.

3.4. Neural network on graph

Most recently the Graph convolutional Networks proposed by Kife and Welling [1] for graph representation receives much attention. Compared with Graph Convolutional Network , the features of the proposed Graph Diffusion Network in this work can be summarized as:

1) A novel diffusion induced graph representation learning approach by applying high-order diffusion model.
2) Through high –order diffusion in each layer of graph network, both the local and global structures of data can be well preserved; Meanwhile , the nonlinear information propagation across layers facilitates the compact clustering of the data;
3) During training the graph network ,a recurrent refinement on the adjacency relationship across the input layer and the hidden layer is implemented to progressively discover the intrinsic structure of data ;
4) Different from Graph Convolutional Network, embedded node representations can be learned in a fully unsupervised way.

3.5. Convergence to a diffusion process on graph

The tool we will use to establish weak convergence of distributions in a functional space is martingale problems; Let, for any ε>0, M^{ε} be a metric space; and let Y^{ε} be a continuous mapping of M^{ε} into some graph Y(M).

Let g^{ε} be a closed set in M^{ε},its image Y^{ε} (g^{ε}) being, for small ε,closer to $\{O_1, O_2, \dots, O_m\}$then some $l(\varepsilon), l(\varepsilon) \to 0$ as $\varepsilon \to 0$ .Denote by g_k^{ε} the part of the set g^{ε} such that $Y^{\varepsilon}(g_k^{\varepsilon})$ is near the vertexO_k.

For a small positive $\delta > l(\varepsilon)$, let G^{δ} be the set of all points x$\in M^{\varepsilon}$ such that Y^{ε}(x) lies closer than δ to the set $\{O_1, O_2, \dots, O_m\}$; r^{δ}, the set of points at exactly the distance δ from this set (of course,

G^{δ} *and* Γ^{δ} *depend also on* ε). The set Γ^{δ} is the union of mutually disjoint sets Γ^{δ}_{ki} of points $x \in M^{\varepsilon}$ such that $Y^{\varepsilon}(x)$ lies at distance δ from O_k,on the segment I_i;similarly, $G^{\delta}=\cup_{k-1}^{m} G^{\delta}_{k}$,where $G^{\delta}_{k}$the part of G^{δ}*near* O_k.

Suppose that $(X^{\varepsilon}(t),P^{\varepsilon}_{x})$ is a strong Markov Process on M^{ε}.Let us denote by τ^{ε} the time when this process reachesg^{ε}; byσ^{δ}, the time it reachesΓ^{s}. We do not suppose $X^{\varepsilon}(t)$ to be continuous, but we will suppose that $Y^{\varepsilon}(X^{\varepsilon}(t))$ is continuous.

If x (ε),>0,is a family of points of M^{ε} such that $Y^{\varepsilon}\big(x(\varepsilon)\big) \to y$ as $\varepsilon \to 0$, We obtain that the probability distribution of $Y^{\varepsilon}(X^{\varepsilon}(\cdot))$ in $\mathbb{C}(Y(M))$ corresponding to the probability $P^{\varepsilon}_{x(\varepsilon)}$ converges weakly to the solution of the solution of the martingale problem corresponding to the operator A, starting from the point y; that is , to the probability P_y corresponding to the diffusion process on the graph.

REFERENCES:

[1] T.N.Kipf, M.Welling , Semi-supervised classification with graph convolutional networks, ICLR,2017.https://openreview.net/forum?id=SJU4ayYgl.
[2] D.K. Hammond, P.Vandergheynst, R.Gribonval,Wavelets on graphs via spectral graph theory, Appl. Comput. Harmonic Anal. 30 (2) (2011) 129-150.
[3] Cayley,A.Philos. Mag. 1857, 13(1),172; 1859,18,374; 1874,47,444; 1877,3,34; Rep.Br.Assoc.Adv.Sci 1875,45,257; Ber. Dtsch. Chem. Ges. 1875,18,1056.
[4] Brown, R. (1828). A brief account of microscopical observations made on the particles contained in the pollen of plants. Philosophical Magazine, 4, 161-173.
[5] Fick, A. (1855). Ueber Diffusion [On diffusion]. Annalen der Physik und Chemie von J. C. Pogendorff, 94, 59-86.
[6] Graham, T. (1829). A short account of experimental researches on the diffusion of gases through each other, and their separation by mechanical means. Quarterly Journal of Science, Literature, and Art, 2, 74–83.
[7] STROOCK, D. W. and VARADHAN, S. R. S. (1979). Multidimensional Diffusion Processes. Springer, Berlin.
[8] Wiener, H.J. Am. Chem. Soc. 1947,69,17,2636;J. Chem. Phys.1947,15, 766;J.Chem. Phys. 1948,52,425,1082.
[9] Butzer, P.L.Berens, H. Semigroups of operators and Approximation. Springer,Berlin(1967).
[10] Malathi selvam Subramanian, Dr.S.Chelliah. On diffusion in Graphs, In:International conference on sustainable Development, 245,1571,(2019).
[11] Triebel , H. Theory of function spaces, Birkhauser , Basel (1983).
[12] Stein, E.M. Topics in Harmonic Analysis, Related to Littlewood-Paley Theory.Princeton University Press, Princeton (1970).
[13] E. M. Stein, Topics in Harmonic Analysis Related to the Littlewood-Paley Theory, vol. 63 of Annals of Mathematics Studies, Princeton University Press, Princeton, NJ, USA, 3rd edition, 1985. View at MathSciNet
[14] K. T. Sturm, "Diffusion processes and heat kernels on metric spaces," Annals of Probability, vol. 26, no. 1, pp. 1–55, 1998. View at Publisher • View at Google Scholar • View at MathSciNet
[15] H.-T. Wu, "Embedding Riemannian manifolds by the heat kernel of the connection Laplacian," Advances in Mathematics, vol. 304, pp. 1055–1079, 2017. View at Publisher • View at Google Scholar • View at MathSciNet • View at Scopus
[16] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," Neural Computation, vol. 15, no. 6, pp. 1373–1396, 2003. View at Publisher • View at Google Scholar • View at Scopus
[17] R. R. Coifman and S. Lafon, "Diffusion maps," Applied and Computational Harmonic Analysis , vol. 21, no. 1, pp. 5–30, 2006. View at Publisher • View at Google Scholar • View at Scopus
[18] R. R. Coifman, S. Lafon, A. B. Lee et al., "Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps," Proceedings of the National Academy of Sciences of the United States of America, vol. 102, no. 21, pp. 7426–7431, 2005. View at Google Scholar
[19] R. R. Coifman and W. E. Leeb, "Earth Mover's distance and equivalent metrics for spaces with semigroups," Tech. Rep. YALEU/DCS/TR-1481, 2013. View at Google Scholar
[20] R. R. Coifman and M. Maggioni, "Diffusion wavelets," Applied and Computational Harmonic Analysis , vol. 21, no. 1, pp. 53–94, 2006. View at Publisher • View at Google Scholar • View at MathSciNet
[21] M. J. Goldberg and S. Kim, "Some Remarks on Diffusion Distances," Journal of Applied Mathematics, vol. 2010, Article ID 464815, 17 pages, 2010. View at Publisher • View at Google Scholar • View at MathSciNet
[22] M. J. Goldberg and S. Kim, "An efficient tree-based computation of a metric comparable to a natural diffusion distance," Applied and Computational Harmonic Analysis , vol. 33, no. 2, pp. 261–281, 2012. View at Publisher • View at Google Scholar • View at MathSciNet