Structural Intrinsic Dimensionality

Stephane Marchand-Maillet[1], Oscar Pedreira[2] and Edgar Chávez[3]

[1] Department of Computer Science, University of Geneva, Switzerland
stephane.marchand-maillet@unige.ch
[2] Universidade da Coruña, Coruña, Spain
oscar.pedreira@udc.es
[3] CICESE, Mexico
elchavez@cicese.mx

Abstract. The dimension of the space within which the data lives is a major driver for the performance of many processing operations. However, global dimensionality cannot be blindly trusted as the data may lie on structures of lower local dimensionality within the ambient space. Here, we address the problem of estimating the local dimensionality of the data space or to provide a consistent proxy for it.

The review of existing local dimensionality estimators shows the various types of geometric information they are based on. We propose the exploration of an alternative route using proximity constraints mapped into the structure of a spanner graph whose properties reflect the local geometry. We propose to adapt PageRank-like information propagation algorithms to infer the structural intrinsic dimensionality directly from the neighborhood structure of data points, taken as vertices. Further, the presence of the spanner over our dataset enables global operations to strengthen the coherence of our estimates and support similarity search.

Keywords: local intrinsic dimension· kissing number· geometric graph spanner

1 Introduction

The dimension of the space containing the data generally refers to the geometric dimension corresponding to the number of linearly independent vectors the space can accommodate. Global data dimension is not a proper characteristic of the data. If the data is uniformly distributed within its ambient space, there is no structural pattern to exploit to construct index structures. The assumption is therefore that the global data dimension may simply represent the dimension of an ambient space within which the data lies over finer structures as a subspace of lower dimension. It is further assumed that the dimension of this subspace may vary locally, therefore defining the notion of local dimensionality. The intrinsic nature of this dimensionality attaches it to the data rather than to its representation and therefore makes it more of an invariant of that data.

In this paper we investigate the nature of local dimensionality along with the proposed models for its estimation (section 2). We then uncover the issues related to its estimation in a practical setup. In particular, we address the issue of the stability of the estimate in relation to the various parameters (sections 3 and 4).

The main contribution of this paper is to make proposals to break the paradox that local dimensionality is a local notion but the statistical nature of its estimation requires to extend its support beyond mere locality. We present experimental measures that support our proposals.

2 The fundamental information behind local dimensionality

We define the *local dimensionality* at point $x \in \mathbb{R}^M$ as being a local indication of $\dim(x)$, the latent dimension at point x of a continuous information density distribution f immersed into the ambient space $\mathbb{R}^M$ ($f : \mathbb{R}^M \mapsto \mathbb{R}_+$; $\int_{\mathbb{R}} f = 1$). That is the lowest dimension of the subspace around x within which f could be embedded with no loss (isometrically).

f is thus a probability density function installed over the ambient space R^M from which we can sample discrete locations (the data). Call X a set of N points $X = \{x_i\}_{[\![N]\!]} \subset \mathbb{R}^M$ that is taken as a sample from the density distribution f. A metric (e.g. Euclidean distance) is used to define the neighborhood of every x_i. Then, the goal of *discrete local dimensionality estimation* is to infer the value of $\dim(x_i)$ at every x_i from the locations of points in the rest of X. In effect, the function $\dim(.)$ can take any positive scalar value ($\dim(x) \in \mathbb{R}^+$), i.e.”discrete” refers to the estimation being based on discrete point locations.

2.1 Motivation for an estimation

Formally, the Nearest Neighbor Indexing (NNI) theorem [21] and subsequent works state that for a workload of vanishing variance in high dimensions, the performance of the class of convex indexes will approach that of sequential search (i.e. $O(N)$). This is clearly supported in practice when working with data of dimensions approaching 20.

Underlying the proof of the NNI theorem is the idea that the indexing covers the dataset X with potentially overlapping convex tiles. As the dimension increases, the vanishing variance of the distribution (D) of distances makes the width of the indexing tiles of the same order than the distance to the nearest neighbor (the best answer to the query). As a result, all tiles need to be fetched during any search process. In this situation, mostly all N points of X are explored as candidates for the result of any search.

The property of vanishing variance stating that

$$\exists \alpha \in \mathbb{R}^+_{\setminus\{0\}} \qquad \text{s.t.} \qquad \lim_{M \to \infty} \mathsf{Var}\left(\frac{D_M^\alpha}{\mathsf{E}[D_M^\alpha]}\right) = 0$$

is closely related to the so-called *concentration of distances* arising due to the Lipschitz structure of Minkowski distances (summing iid coordinates) and the Chebyshev inequality [3]. In other words, no convex indexing can provide exclusion, due to the concentration of distances. This makes this result rather universal in Data Analysis and motivates the quest for discrete local dimensionality estimation. In essence, by turning the argument upside down, we seek an estimate that correlates with the factor (called "dimensionality", $\dim(x)$) that influences the performance of any data nearest neighborhood indexing (and analysis).

Note further that most of the related literature focuses on estimating the dimensionality but a proper use of this characterization is yet to be proposed. This work is a step towards constructing a context where the analysis provides actionable tools to make effective similarity search in high dimensional spaces.

2.2 Expansion-based estimation

The class of expansion-based estimation techniques relies on the fact that the increase of volume of a M-dimensional hypersphere V_M is essentially related to the increase of its radius r by an exponential relationship.

$$V_M(r) = \frac{2\pi^{\frac{M}{2}}}{M\Gamma(\frac{M}{2})} r^M \qquad \Rightarrow \qquad \frac{\partial V_M}{\partial r} = \frac{2\pi^{\frac{M}{2}}}{\Gamma(\frac{M}{2})} r^{M-1}$$

This is exploited in the definition of the *Expansion Dimension* (ED) [18] and its generalization GED [12]. The strategy is to estimate a proxy for the volume of the hypersphere of radius r centered at x_i by counting the number n of data points in a r-range query from x_i. Hence, the dimension is estimated by a log of the relative increase of this number (from n_1 to n_2) versus the increase in radius (from r_1 to r_2): $\mathsf{GED}(x_i) = \frac{\log \frac{n_2}{n_1}}{\log \frac{r_2}{r_1}}$.

The above assumes (at least locally) a uniform distribution of data around x_i. It is further refined by considering (instead of the volume of the hypersphere) the cumulative function $F(r)$ of a distance distribution (whose 0 would be at -every- x_i). This allows to model a variable density within the space and to define the lID [13, 14] that matches the GED for a uniform distribution.

2.3 Concentration of correlates

Another route for exploring the local geometry of the space is to look at angles. Fixing one direction $\mathbf{u}_k$ from x_i (thru x_k say), one can study the distribution of the angles made between this direction and vectors $\mathbf{u}_j$ whose extremity x_j other than x_i is sampled over a hypersphere centered at x_i. Such an estimation amounts to compute the surface of spherical caps defined by the cones generated by $\mathbf{u}_k$ and angle $\theta \propto \angle(\mathbf{u}_k, \mathbf{u}_j)$. This distribution of correlates ($\cos(\theta) = \langle \mathbf{u}_k, \mathbf{u}_j \rangle$) is further known to concentrate with increasing dimensionality [5, 6, 9]:

$$\mathsf{P}(\theta) = \frac{\Gamma(\frac{M}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{M-1}{2})} \sin^{M-2}(\theta) \qquad \text{and} \qquad \mathsf{Var}(\cos(\theta)) = \frac{1}{M}$$

The latter second order information is then used to estimate the local dimensionality by local samples of angles [22].

2.4 Sphere packing

Another possible approach is to also use the notion of sphere packing [6] but in relation to the *kissing number*. Here, the observed estimate is the number of non-intersecting hyperspheres of diameter r able to be tangent to (to "kiss") a hypersphere of radius r centered at x_i. This is known as the kissing number $\mathsf{Kiss}(M)$ whose exact values are known only for a select number of dimension values $M \in \{1, 2, 3, 4, 8, 24\}$. For other values, upper and lower bounds which illustrate the dynamic of the kissing number with respect to the dimension have been proposed [16]:

$$(1+O(1))\sqrt{\frac{3\pi}{8}}\log\left(\frac{3}{2\sqrt{2}}\right)M^{\frac{3}{2}}\left(\frac{2}{\sqrt{3}}\right)^{M} \leq \mathsf{Kiss}(M) \leq (1+O(1))\sqrt{\frac{\pi}{8}}M^{\frac{3}{2}}2^{\frac{M}{2}}$$

The regular dependence of these bounds on dimension makes the kissing number another appropriate entry door to the estimation of the local dimensionality.

It can be noted that this information also relates to the above angle-based estimation in the sense that the kissing number counts the maximum number of non-intersecting spherical caps with total angle $\frac{\pi}{3}$ (generative angle $\frac{\pi}{6}$) one can segment the central hypersphere surface with. It is a particular instance of so-called "spherical codes" [4, 16] with $\theta = \frac{\pi}{3}$. Equivalently, the kissing number counts the number of points a given point can be nearest neighbor of (so-called reverse nearest neighbor). For example, a point of a 2D plane can only be the nearest neighbor of $\mathsf{Kiss}(2)$=6 points (arranged as a hexagon).

2.5 Discussion

The above three approaches are different in their computation but rely on essentially the same information.

Expansion-based estimations explicitly use the shape of the density distribution along the distance axis. That is, from the central point x_i where the local dimensionality is to be estimated the hyperspherical shell of radius r around that point is integrated into the point of coordinate r on the distance axis. It is the growth rate of this value that is explicitly modeled by GED and lID. In the discrete version [1, 2], the lID is a measure of the density of data within a thick spherical shell (from the 1-NN to the k-NN of x_i). The transition from the continuous model to the discrete estimation still imposes an assumption of local uniformity in the distribution of the k nearest neighbors.

The estimation of the ABID [22] is based on estimating the concentration of the cosine similarities between points on a hypersphere centered at x_i. Using fixed length vectors, the cosine similarities are known to correlate with squared distances (e.g. this is the basis for the MIPS problem [10]) [4]: $\langle \mathbf{u}_k, \mathbf{u}_j \rangle \propto d(x_k, x_j)^2$ In practice, the k nearest neighbors from x_i are used so that the ABID is also a reflection of the density of data within a thick spherical shell (from the 1-NN to the k-NN of x_i). The advantage of this estimation is that angles involve triplets of points and create a combinatoric volume of estimates, thus reducing the span of the neighborhood (value of k) required in practice to obtain a robust estimate.

Finally, using the kissing number to estimate the local dimensionality imposes complementary constraints: at fixed radius r from x_i (first constraint) the kissing number counts how many points can be organized so as to be at least r-distant from each other (second constraint). The first constraint may similarly be relaxed by exploring values of r along the distance axis. The second constraint may be imposed by selecting neighbors dispersed around x_i. This is handled via the generation of spanner graphs such as the reverse neighbor graph, the half-space proximal graph (HSP) [7] or the Yao and θ-graphs [20]. In earlier works [15, 8], we explored the correlation between indicators of some of these graphs with local dimensionality to partition the dataset in view of improving its *indexability*.

3 Information propagation on neighbor graphs

In their original presentations, the above measures essentially treat all points x_i individually and sequentially. They then operate some statistical analysis (e.g. mean or variance) on the distribution of the local dimensionality values throughout the dataset.

Considering the points independently creates the tension between the desire to compute a robust estimation over a large number of neighbors (large k in k-NN) and the intrinsic wish to stay local (small ε in ε-NN). The kissing number instructs us that for dimensions as limited as 20 the coverage of the hypersphere requires $O(10^4)$ neighbors already, which is beyond the density of any classical dataset[1]. One can therefore question the validity of the empirical estimates made using $k = O(10^2)$ neighbors. This is partly discussed in [22], for example.

In addition, the local dimensionality may vary from a point to another in X. Hence, computing global a posteriori statistics may not be so relevant for all datasets (e.g. a Saturn-shaped dataset). This can be related to the Yule-Simpson effect [11], which induces potentially contradictory interpretations, depending of the scale at which the data is studied[2].

In turn, a true local dimensionality estimate would enable operations like dimensionality-based clustering, and define *indexability* [15].

Here, we make steps in the direction pointed by the above remarks: using the global structural information provided by the full dataset X for estimating the local dimensionality at every $x_i \in X$. We relax the implicit above assumption of a constant local dimensionality by assuming that the local dimensionality bears

[1] We argue that estimating the $O(M)$ linearly independent vectors reflecting the geometric dimension (e.g. using rank-based methods such as local PCA) would not be reliable in that case due to quasi-orthogonality [17] and the issue of local neighborhood selection.

[2] This further relates the question of discrete local dimension estimation to local scale estimation, an important topic, addressed in [2], left for further investigation.

some smoothness of the form: $d(x_i, x_j) < r \Rightarrow |\dim(x_i) - \dim(x_j)| < \alpha$ for small values of r and α.

3.1 Structural regression

Following the above discussion we propose to enforce that the dimensionality at x_i is the weighted average of the dimensionality of its neighbors x_j:

$$\dim(x_i) = \frac{1}{Z_i} \sum_{x_j \in \mathcal{N}(x_i)} w_{ij} \dim(x_j) \tag{1}$$

for some neighborhood $\mathcal{N}(x_i)$ and some influence weighting w_{ij} with proper normalization $Z_i = \sum_j w_{ij}$ (note that $x_i \notin \mathcal{N}(x_i)$ and $w_{ii} = 0$). This smoothness condition alone makes the dimensionality estimates prone to translation, as the above stays valid if $\dim'(x_i) = \dim(x_i) + K$ with any constant K. Hence, we apply this strategy starting from an estimate ϵ of the dimensionality (e.g. lID or ABID).

Given $\epsilon = [\epsilon_i]^\mathsf{T}$ as estimate for local dimensionality $\mathbf{d} = [d_i]^\mathsf{T}$, we resolve the classical regression:

$$\mathbf{d}^* = \min_{\mathbf{d} \in \mathbb{R}^N} \mathcal{L}(\mathbf{d}, \epsilon) \quad \text{where} \quad \mathcal{L}(\mathbf{d}, \epsilon) = \frac{1}{2}\|\mathbf{d} - \epsilon\|_2^2 + \frac{\lambda}{2}(d_i - \sum_{x_j \in \mathcal{N}(x_i)} d_j)^2$$

with $\lambda > 0$ controlling the smoothness when maintaining the volume $\sum_i d_i$ constant. The above is a classical convex minimization solved iteratively using:

$$d_i^{(t+1)} \leftarrow d_i^{(t)} - \eta \left[(d_i^{(t)} - \epsilon_i) + \lambda(d_i^{(t)} - \sum_{x_j \in \mathcal{N}(x_i)} w_{ij} d_j^{(t)}) \right] \tag{2}$$

for learning rate $0 < \eta < 1$. It is easy to see that this guarantees $\mathsf{Var}[\mathbf{d}^*] \leq \mathsf{Var}[\epsilon]$.

3.2 Experiments

We wish to validate empirically our analysis on the regularity of local dimensionality estimates. We use the lID [13] and its MLE estimate [2], and the HSP degree [15] over uniform datasets of known dimensionality to calibrate our study.

We generate datasets of various dimensionalities ($M \in \{3, 6, 10, 15, 20, 30, 50\}$) and various densities ($N \in \{10000, 20000, 30000, 40000, 50000, 100000, 150000\}$) to study the influence of dimension with respect to data density.

The datasets are composed of N samples of a distribution (Uniform or Gaussian) restricted to a M-dimensional hypersphere of radius 1. Spherical datasets are chosen to eliminate "corner" effects.

Figure 1(top) shows the estimates provided by the HSP built over an hypersphere filled with uniform (left) and Gaussian (right) sampling, using its degree as a function of the data density and for the dimensionalities listed above. One

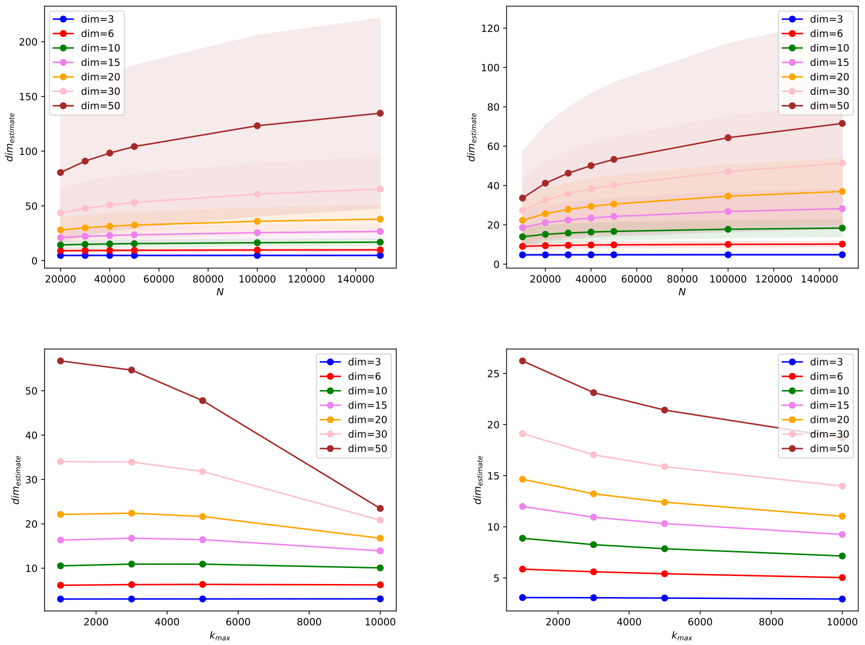


Fig. 1. Variation of dimensionality estimates versus parameters (N or $k_{\max}$) for [left: spherical data] [right: gaussian data] [up: HSP degree] [down: lID with $N = 50'000$]

clearly sees that although only correlated with the true dimensionality, the estimates stabilize with an increase of the density but that the variance augments with density.

Turning to the lID MLE estimate, we use the same datasets with fixing a high density with $N = 50'000$ and varying the size of the neighborhood over which the estimate is computed: $k_{\max} \in \{1000, 3000, 5000, 10000\}$. Clearly, the estimates get corrupted using a too large $k_{\max}$. Further, as dimensionality increases this phenomenon is more drastic (the variance here is too large to be properly displayed).

The above clearly illustrates the contradiction in extending the support of estimation of a local estimator. It shows that even though the estimates may be considered as overall reliable (e.g. when averaged), the sensitivity to their parameters and location of estimation is so that they cannot be blindly applied without some knowledge of an appropriate scale and the presence of a reasonable local data density.

The main issue lies in the variance of these estimates, as we seek a factor that correlates with what could be referred to as local dimensionality. We therefore look at the behavior of these measures over datasets of varying dimensionality to demonstrate the capability of our smoothing (Eq 1) to reduce the variance of the estimates. We generate a dataset with 3 non-overlapping spherical uniform clusters containing 20'000, 15'000 and 15'000 points respectively and of dimensionality 20, 10 and 5 respectively. We initially estimate the degree of the HSP and smooth it using our iterative convolution (Eq 2) where we fix $\lambda = 5$ and $\eta = 0.1$ everywhere.

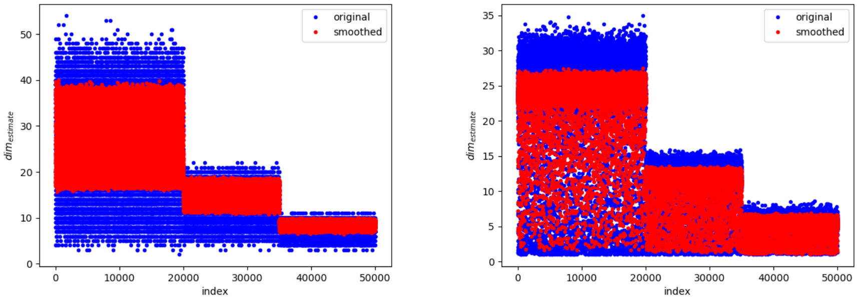


Fig. 2. Estimates before and after iterative convolution. The initial values are [left: HSP degree], [right: lID using k=1000]. Both diffusions happen over the HSP

The result is reported in figure 2(left). We first clearly see that the estimate is correct with respect to our calibration in fig 1. The three clusters are clearly identified. However, the estimate vary significantly within clusters. As result of the regression, dimension estimates are corrected (from blue to red dots) and the variance diminish appropriately. Here, inspired by the gravitational physical model, we use $w_{ij} = \frac{1}{d^2(x_i,x_j)}$ as influence weight.

We use the same dataset to perform an estimate with the lID (fixing $k_{\max} = 1000$) (fig 2(right)). The regression clearly alleviates the problem of high variance in the estimates. However, due to the initial distribution, the estimates remain rather spread, indicating a need for exploring stronger constraints in the regression or using a stronger value for λ.

Table 1 gives the variation of mean and standard deviation (between brackets) of the estimates in both cases and per cluster. Note that the result of the convolution is not the mere mean and variance of the original, showing that the weighting structure does play a role.

4 Structural intrinsic dimensionality

We push the idea of information propagation over a neighborhood structure a step further. The geometry described in section 2.4 suggests that dimensionality may be captured by the connectivity structure of the neighborhood graph itself.

	Cluster 1	Cluster 2	Cluster 3
HSP initial	28.02 (10.97)	13.89 (3.37)	7.54 (1.26)
HSP smoothed	26.08 (6.71)	15.29 (2.06)	8.71 (0.64)
lID initial	22.70 (6.51)	10.95 (2.83)	5.23 (1.06)
lID smoothed	22.74 (3.65)	10.93 (1.95)	5.21 (0.83)

Table 1. Mean and standard deviations for HSP degree and lID before and after iterative convolution ($\lambda = 5$ and $\eta = 0.1$)

We already demonstrated that the degree of the HSP whose construction relates to the kissing number correlates with dimensionality. It is known (notably from the development of the PageRank algorithm) that information propagation over directed graphs can provide essential information about the underlying connectivity structure. We therefore hypothesize that the local dimensionality may be inferred via information diffusion, provided the graph encodes this information.

Given a directed graph $G = (X, E)$ with edge set E defined from geometric constraints, i.e. $(x_i, x_j) \in E$ iff $x_j \in \mathcal{N}(x_i)$, we define information propagation of value $d(x)$ as the convergence of the (directed) iterative process:

$$d_i^{(t+1)} \leftarrow \sum_{x_j \text{ s.t. } x_i \in \mathcal{N}(x_j)} w_{ji} d_j^{(t)} \tag{3}$$

Classically, the diffusion is done so as to preserve the value $\sum_i d(x_i)$ constant. The directed setup thus imposes $\sum_j w_{ij} = 1 \quad \forall i \in [\![N]\!]$, making matrix $W = [w_{ij}]_{ij \in [\![N]\!]}$ a row-stochastic matrix ($w_{ij} = 0$ if $(x_i, x_j) \notin E$). It is known that under proper conditions, this process converges to the principal eigenvector (with eigenvalue 1) of W, the weighted adjacency matrix of G. In PageRank-like diffusion algorithms, edge weights w_{ij} are tuned so as to distribute the value at node x_i to forward neighbors x_j based on the degree (e.g. $w_{ij} = \frac{1}{\deg^+(x_i)}$).

Adapting to our geometrical context we read $\deg^+(x_i) = \sum_{x_j \in \mathcal{N}(x_i)} 1$. That is, every outgoing edge from x_i counts 1, so that $w_{ij} = \frac{1}{\sum_{x_k \in \mathcal{N}(x_i)} 1}$. We transform this to influence by inserting an inverse dependence $\phi(.)$ to distance as edge weight, while preserving the row-stochasticity constraint:

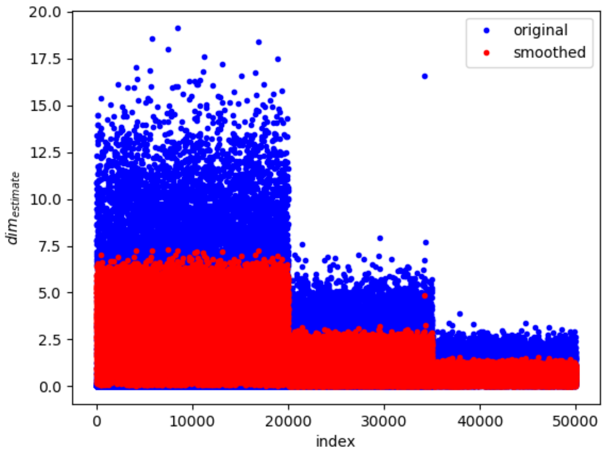
$$w_{ij} = \frac{\phi(d(x_i, x_j))}{\sum_{x_k \in \mathcal{N}(x_i)} \phi(d(x_i, x_k))} \qquad \text{where, for example} \qquad \phi(x) = \frac{1}{x} \tag{4}$$

Using an inverse dependence as base edge weight (Eq 4) therefore induces a softmax-like filter on edges, thus favoring the shortest edge emanating from every x_i. Combining this with our diffusion strategy (Eq 3), every vertex x_i receives mostly influence from the other vertices x_j of which it is the closest neighbor. This corresponds exactly to the definition of the kissing number at x_i. We therefore expect diffusion over such graph structures to exhibit an information that correlates with local dimensionality and that we can refer to as *structural intrinsic dimensionality.*

The relationship with eigencentrality in graphs is also clear as it corresponds to the case where $\phi(x) = 1$. This nicely connects with and continues our earlier proposals [15, 8], where we proposed graph centrality measures as indicators that correlate with local dimensionality. In this context, the degree of the HSP is seen as its degree centrality indicator, itself an approximation of the eigencentrality.

4.1 Experiments

We now propose results for our structural intrinsic dimensionality estimation using again the cluster dataset presented above. Our initial experiment confirms that a careful design of the edge weighting scheme is important. We found that using $\phi(x) = \frac{1}{\sqrt{x}}$ in Eq 4 produces interesting results. As before, we smoothed these results via iterative convolution. The results are presented in figure 3.



	Cluster 1	Cluster 2	Cluster 3
Initial	3.14 (3.60)	1.61 (1.48)	0.86 (0.58)
Smoothed	3.44 (1.54)	1.40 (0.75)	0.65 (0.32)

Fig. 3. Dimensionality indicator from information propagation over the HSP ($\phi(x) = \frac{1}{\sqrt{x}}$). Estimates before and after iterative convolution

Of course, the value of the estimate does not match the dimensionality as understood as space dimension. However, information propagation does produce a proper indicator of this dimensionality. More investigation is required to understand the most favorable structure of underlying spanner (HSP, k-NN, reverse-k-NN, ...) to use for propagation and the best weighting scheme. It can even be envisaged to join both, starting from the complete graph and decimating it with respect to the defined edge weight (i.e. removing edges with negligible transfer in order to reduce the computational cost).

Finally, it seems natural to target the integration of the iterative convolution process (Eq 2) with the propagation (Eq 3). We therefore believe that this graph-based strategy for the estimation of the structural intrinsic dimensionality opens many interesting questions.

5 Conclusion

Local dimensionality is a major driver for the performance of data processing techniques. Its effects are deeply rooted into statistics, as demonstrated by the concentration of distances that is one aspect of the curse of dimensionality. Obtaining indicators for local dimensionality in the discrete space is therefore of interest and most existing local dimensionality indicators are based on the estimation of the variation of local density.

Here, we consider any indicator that show a monotonic relationship with local dimensionality. We propose to exploit the definition of the kissing number to obtain such an indicator. Using a graph structure over the dataset, we show that information propagation can not only help strengthening classical indicators but also being used as an estimator itself. This work therefore gives a formal grounding for our earlier proposals [15, 8].

The results open the question of the underlying graph structure that would be best suited for such an exploration. We suggest that this question is equivalent to defining a proper edge weight, capturing the geometry of the dataset in the graph structure. As this weighting naturally makes use of the underlying metric, this clearly relates to the construction of appropriate geometric t-spanners that will be one direction we wish to explore. One can note that graph-based computations also provide a computational solution to the problem of combining local and global structures. Computation can further be distributed using the tight equivalence between information propagation algorithms and random walk processes.

Finally, determining local dimensionality does not directly provide a solution to counter its adverse effects. We have proposed to use it to partition the dataset based on its *indexability* [15]. Following that route, we believe that the graphs arising from the estimation of the structural intrinsic dimensionality will be useful for constructing efficient indexing strategies in the line of recent graph-based indexing techniques using navigable structures [19].

Another option for using local dimensionality estimates in an operational setting may be their use to drive local embedding for adapting the indexing locally into a lower dimensional context.

Acknowledgments

O. Pedreira is partially funded by: MICIU/FEDER-UE BIZDEVOPSGLOBAL: RTI-2018-098309-B-C32 and Xunta de Galicia/FEDER-UE GRC: ED431C 2017/58.

References

1. Amsaleg, L., Chelly, O., Furon, T., Girard, S., Houle, M.E., Kawarabayashi, K., Nett, M.: Extreme-value-theoretic estimation of local intrinsic dimensionality. Data Mining and Knowledge Discovery **32**(6), 1768–1805 (2018)
2. Amsaleg, L., Chelly, O., Houle, M.E., Kawarabayashi, K., Radovanović, M., Treeratanajaru, W.: Intrinsic dimensionality estimation within tight localities. In: Proceedings of the SIAM International Conference on Data Mining (SDM). SIAM (2019)
3. Boucheron, S., Lugosi, G., Massart, P.: Concentration Inequalities: A Nonasymptotic Theory of Independence. OUP Oxford (2013)
4. Boyvalenkov, P., Dodunekov, S., Musin, O.R.: A survey on the kissing numbers. CoRR **abs/1507.03631** (2015)
5. Cai, T., Fan, J., Jiang, T.: Distributions of angles in random packing on spheres. Journal of Machine Learning Research **14**(21), 1837–1864 (2013), http://jmlr.org/papers/v14/cai13a.html
6. Cai, T., Jiang, T.: Phase transition in limiting distributions of coherence of high-dimensional random matrices. Journal of Multivariate Analysis **107**, 24–39 (2012). https://doi.org/https://doi.org/10.1016/j.jmva.2011.11.008
7. Chavez, E., Dobrev, S., Kranakis, E., Opatrny, J., Stacho, L., Tejeda, H., Urrutia, J.: Half-space proximal: A new local test for extracting a bounded dilation spanner of a unit disk graph. In: International Conference On Principles Of Distributed Systems. pp. 235–245. Springer (2005)
8. Chavez, E., Pedreira, O., Marchand-Maillet, S.: Reverse k-nearest neighbors centrality measures and local intrinsic dimension. In: Proceedings of the 13th International Conference on Similarity Search and Applications (SISAP 2020) (2020)
9. Connor, R., Dearle, A.: Sampled angles in high-dimensional spaces. In: Proc. of Int. Conf. on Similarity Search and Applications (SISAP'20). Springer (2020)
10. Ding, Q., Yu, H.F., Hsieh, C.J.: A fast sampling algorithm for maximum inner product search. In: Chaudhuri, K., Sugiyama, M. (eds.) Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics. Proceedings of Machine Learning Research, vol. 89, pp. 3004–3012. PMLR (16–18 Apr 2019)
11. Good, I.J., Mittal, Y.: The Amalgamation and Geometry of Two-by-Two Contingency Tables. The Annals of Statistics **15**(2), 694 – 711 (1987)
12. Houle, M.E., Kashima, H., Nett, M.: Generalized expansion dimension. In: 2012 IEEE 12th International Conference on Data Mining Workshops. pp. 587–594 (2012)
13. Houle, M.E.: Local intrinsic dimensionality I: an extreme-value-theoretic foundation for similarity applications. In: Proc. of Int. Conf. on Similarity Search and Applications (SISAP'17). pp. 64–79. Springer (2017)
14. Houle, M.E.: Local intrinsic dimensionality II: multivariate analysis and distributional support. In: Proc. of Int. Conf. Similarity Search and Applications (SISAP'17). pp. 80–95. Springer (2017)

15. Hoyos, A., Ruiz, U., Marchand-Maillet, S., Chávez, E.: Indexability-based dataset partitioning. In: Proc. of Int. Conf. on Similarity Search and Applications (SISAP'19). pp. 143–150. Springer (2019)
16. Jenssen, M., Joos, F., Perkins, W.: On kissing numbers and spherical codes in high dimensions. CoRR **abs/1803.02702** (2018)
17. Kainen, P., Kůrková, V.: Quasiorthogonal dimension of Euclidean spaces. Applied Mathematics Letters **6**(3) (1993)
18. Karger, D.R., Ruhl, M.: Finding nearest neighbors in growth-restricted metrics. In: 34th ACM Symposium on Theory of Computing. ACM (2002)
19. Malkov, Y.A., Yashunin, D.A.: Efficient and robust approximate nearest neighbor search using hierarchical navigable small world graphs. IEEE Transactions on Pattern Analysis and Machine Intelligence **42**(04), 824–836 (apr 2020). https://doi.org/10.1109/TPAMI.2018.2889473
20. Narasimhan, G., Smid, M.: Geometric Spanner Networks. Cambridge University Press (2007). https://doi.org/10.1017/CBO9780511546884
21. Shaft, U., Ramakrishnan, R.: Theory of nearest neighbors indexability. ACM Trans. Database Syst. **31**(3), 814–838 (Sep 2006)
22. Thordsen, E., Schubert, E.: ABID: Angle based intrinsic dimensionality. In: Proc. of Int. Conf. on Similarity Search and Applications (SISAP'20). Springer (2020)