

The Hubness Phenomenon: Fact or Artifact?

Thomas Low¹, Christian Borgelt², Sebastian Stober¹, and
Andreas Nürnberger¹

Abstract The hubness phenomenon, as it was recently described, consists in the observation that for increasing dimensionality of a data set the distribution of the number of times a data point occurs among the k nearest neighbors of other data points becomes increasingly skewed to the right. As a consequence, so-called *hubs* emerge, that is, data points that appear in the lists of the k nearest neighbors of other data points much more often than others. In this paper we challenge the hypothesis that the hubness phenomenon is an effect of the dimensionality of the data set and provide evidence that it is rather a boundary effect or, more generally, an effect of a density gradient. As such, it may be seen as an artifact that results from the process in which the data is generated that is used to demonstrate this phenomenon. We report experiments showing that the hubness phenomenon need not occur in high-dimensional data and can be made to occur in low-dimensional data.

1 Introduction

That working with high-dimensional data is difficult has been known for quite some time now, although not all of the effects of a large number of dimensions are completely understood yet. In 1961, R.E. Bellman was among the first who recognized the various problems that arise in high-dimensional spaces, for which he coined the term *curse of dimensionality* [1]. One property of this “curse” is the fact that with an increasing number of dimensions the volume of a unit hyperball grows considerably less quickly than the volume of a unit hypercube. As a consequence, most distance metrics, like the Euclidean

¹ Data and Knowledge Engineering Group, Otto-von-Guericke-University of Magdeburg, Universitätsplatz 2, D-39106 Magdeburg, Germany, firstname.lastname@ovgu.de

² European Centre for Soft Computing, c/ Gonzalo Gutiérrez Quirós s/n, E-33600 Mieres (Asturias), Spain, christian.borgelt@softcomputing.es

distance, suffer from a loss of relative contrast. This effect is also known as *distance concentration* [2] and causes the relative difference in the distance of a given query point to its nearest and its farthest neighbor to vanish.

Another (alleged) property of the curse of dimensionality is the *emergence of hubs*, which was first described as a general problem in [10]. Here *hubs* are defined as data points that appear unusually often among the k nearest neighbors of other data points. Described in statistical terms, the distribution of the number of times a data point occurs in the nearest neighbor lists of other data points becomes skewed to the right. This phenomenon has been examined and demonstrated to be present in many real-world data sets in [11], where it was also analyzed how it affects a broad spectrum of machine learning tasks and dimensionality reduction techniques. The core claim of both papers, [10] and [11], is that the emergence of hubs is an intrinsic effect of the dimensionality of the data—a view we dare to challenge here.

Our core claim in this paper is that the hubness phenomenon is an effect of a density gradient, not an effect of the dimensionality of the data space. Note, however, that if the data points are sampled from a region that is bounded, there is necessarily a density gradient at the boundary of the region. Since the ratio of the size of the (hyper-)surface (i.e. the boundary) of a region to its (hyper-)volume increases exponentially with the dimensionality of the data space, the density gradient at the boundary is emphasized (in the sense that it influences more data points) and thus makes the hubness phenomenon more notable. This explains the observations of [11]. However, high dimensionality alone is not sufficient to produce the hubness phenomenon as we demonstrate by sampling from a boundary-less high-dimensional space, for which the hubness phenomenon is essentially nonexistent. We also show that the strength of the hubness phenomenon depends on the relative size of the boundary of the sampling region. In addition, we show that introducing sufficiently many boundaries (and thus many places with a density gradient) in a low-dimensional sampling region creates the hubness phenomenon.

The remainder of this paper is organized as follows: in Section 2 we define the hubness phenomenon and several measures by which we try to capture its strength, thus obtaining proper means to quantify and compare this phenomenon over different data sets. In Section 3 we describe our data generator, that is, the procedures we employed to generate high-dimensional data sets as well as the structure of these data sets. In Section 4 we describe the experiments (on artificial data) we conducted and report and interpret our results. Finally, in Section 5 we draw conclusions from our discussion.

2 Measuring Hubness

Before we can define measures for the strength of the hubness phenomenon, we have to introduce the notions on which it is based. Let $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$

be an m -dimensional data set with n data points $\mathbf{x}_i = (x_{i1}, \dots, x_{im})$ for $i \in \{1, \dots, n\}$. Furthermore, let $N_k(\mathbf{x}) \subseteq \mathbf{X} - \{\mathbf{x}\}$ be the set of the k nearest neighbors ($k < n$) of \mathbf{x} in \mathbf{X} . That is, $\forall \mathbf{y} \in N_k(\mathbf{x}) : \forall \mathbf{z} \in \mathbf{X} - N_k(\mathbf{x}) - \{\mathbf{x}\} : d(\mathbf{x}, \mathbf{y}) \leq d(\mathbf{x}, \mathbf{z})$ and $|N_k(\mathbf{x})| = k$ (assuming that ties are broken arbitrarily). We consider mainly the Euclidean distance $d(\mathbf{x}, \mathbf{y}) = (\sum_{j=1}^m (x_j - y_j)^2)^{\frac{1}{2}}$, but in principle other distance measures may also be studied (cf. [11]).

The quantity $o_k(\mathbf{x}) = \sum_{\mathbf{y} \in \mathbf{X}} \mathbb{1}_{N_k(\mathbf{y})}(\mathbf{x})$, where $\mathbb{1}_{N_k(\mathbf{y})}$ is the indicator function of $N_k(\mathbf{y})$ w.r.t. \mathbf{X} (that is, $\mathbb{1}_{N_k(\mathbf{y})}(\mathbf{x}) = 1$ if $\mathbf{x} \in N_k(\mathbf{y})$ and 0 otherwise), counts the number of times the data point \mathbf{x} occurs in the sets of nearest neighbors of other data points.¹ We call $o_k(\mathbf{x})$ the k -occurrence of the data point $\mathbf{x} \in \mathbf{X}$. The hubness phenomenon can now be described as the observation that the distribution of the values $o_k(\mathbf{x})$ for $\mathbf{x} \in \mathbf{X}$ is (considerably) skewed to the right or that some data points have unusually high k -occurrence values (i.e., considerably larger than the mean value, which is obviously k).

In order to obtain an objective evaluation of the strength of the hubness phenomenon, we rely on a few very simple measures. The most straightforward approach is obviously to compute the *skewness* (or simply *skew*) of the distribution of the $o_k(\mathbf{x}_i)$, $i \in \{1, \dots, n\}$, which is defined as

$$\gamma = \frac{\frac{1}{n} \sum_{i=1}^n (o_k(\mathbf{x}_i) - \bar{o}_k)^3}{\left(\frac{1}{n} \sum_{i=1}^n (o_k(\mathbf{x}_i) - \bar{o}_k)^2\right)^{3/2}},$$

where $\bar{o}_k = k$ is the mean of the k -occurrence values (see [4] for a general definition of skewness). If $\gamma > 0$, the distribution is skewed to the right. However, usually only (very) few of the points in a data set are hubs, that is, exhibit high k -occurrence. Thus the skewness may not always be sensitive enough to measure the phenomenon properly. On the other hand, the skewness is computed from all data points and thus large values may result even if there are no sizable hubs, namely if there is an asymmetry close to the mean.

An alternative approach that immediately suggests itself is to use the largest k -occurrence in the data set. However, this measure has the disadvantage that it is strongly affected by the randomness of the sampling process. Thus it is only sufficiently expressive if averaged over a certain number of runs. To obtain a better measure we average the k -occurrence of the fraction q of data points with the highest k -occurrences, where q should be small. The averaging makes the measure more robust, yet allows us to properly capture the value of the highest k -occurrences.² In our experiments we tried $q = 0.1\%$, $q = 0.5\%$ and $q = 1\%$. Since this measure depends directly on the number k of nearest neighbors that are considered, we finally divide by k :

¹ Note that the sum need not exclude the data point \mathbf{x} , because $\mathbf{x} \notin N_k(\mathbf{x})$ by definition.

² We refrained from using the $(1 - q)$ -quantile (which would be an even more robust choice) because of the integer nature of the k -occurrences, which limits the number of possible values, especially for small k (that is, for few nearest neighbors). In addition, the $(1 - q)$ -quantile does not capture the distribution of values at and beyond it.

$$h_1(q) = \frac{1}{k|O_k(q)|} \sum_{\mathbf{x} \in O_k(q)} o_k(\mathbf{x}),$$

where $O_k(q)$ contains the $\lfloor qn \rfloor$ data points with the highest k -occurrences. Note that this measure captures the maximum k -occurrence for $q = 1/n$.

As an alternative, we consider what percentage of the data points have a k -occurrence value at least β times the mean value k , that is, the percentage of data points $\mathbf{x} \in \mathbf{X}$ with $o_k(\mathbf{x}) \geq \beta k$. Formally we have

$$h_2(\beta) = \frac{|\{\mathbf{x} \in \mathbf{X} \mid o_k(\mathbf{x}) \geq \beta k\}|}{|\mathbf{X}|} \cdot 100\%.$$

In particular, we experimented with $\beta = 2$, $\beta = 3$ and $\beta = 4$. That is, if we consider, for example, the 10 nearest neighbors, we compute what percentage of the data points occurs in at least 20, 30, and 40 nearest neighbor lists. Note that h_2 highlights the number of hubs, while h_1 focuses on their size.

3 Data Generation

The design of a data generator starts with the choice of a (pseudo-)random number generator (RNG), usually for a uniform distribution. Here we rely on a simple and very fast RNG producing 32 bit unsigned integer numbers, which was suggested by G. Marsaglia [6]. This RNG computes the next (pseudo-)random number from the previous five numbers, has a period of about 2^{160} , and seems to pass all standard quality tests for RNGs. We prefer this RNG over the more fashionable Mersenne Twister [8] due to its simplicity and much higher speed. We use this RNG to generate uniformly distributed (pseudo-)random floating point numbers in the interval $[0, 1)$ by generating two 32 bit unsigned integers i_1 and i_2 and then computing $r = i_1 \cdot 2^{-32} + i_2 \cdot 2^{-64}$, thus filling all bits of the mantissa of a double precision floating point number. We ensure that $r \in [0, 1)$ by rejecting r and generating a new random number should the (highly unlikely) event occur that (due to rounding) $r = 1$.

In order to obtain normally distributed (pseudo-)random numbers (which we also need for sampling from (hyper-)spheres and (hyper-)balls, see below), we employ the so-called polar method [7, 5], which consists in generating two (pseudo-)random numbers x and y that are uniformly distributed in $[-1, 1)$ until $s = x^2 + y^2 < 1$, that is, until the point (x, y) lies inside a unit circle. Then the transformed numbers $x' = \xi x$ and $y' = \xi y$, where $\xi = \sqrt{-2 \ln(s)/s}$, are normally distributed with mean 0 and variance 1.

Our data generator can sample from a multivariate standard normal distribution as well as uniformly from a (hyper-)cube, a (hyper-)ball, and a (hyper-)sphere (i.e., the surface of a (hyper-)ball). Sampling uniformly from an m -dimensional (hyper-)cube $[-1, 1)^m$ (also called an m -cube) is trivial:

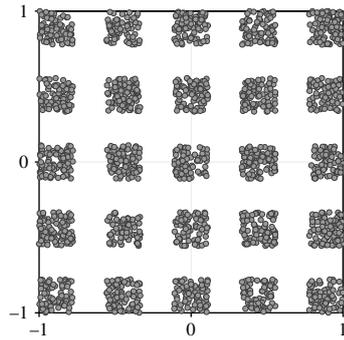


Fig. 1 A data set with 2000 points that were sampled uniformly from a grid of small squares described by the parameters $\eta = 5$ and $\alpha = 0.5$ (see the explanations in the main text).

simply generate m random coordinates x_1, \dots, x_m that are uniformly distributed in $[0, 1)$ and transform them with $x'_i = 2x_i - 1$ for $i = 1, \dots, m$. Sampling from an m -dimensional standard normal distribution is equally simple: use the polar method (see above) to generate m normally distributed coordinates. The coordinates are combined in a vector $\mathbf{x} = (x_1, \dots, x_m)$.

For sampling uniformly from an m -dimensional (hyper-)sphere (also called an m -sphere) we exploit the insight that a multivariate standard normal distribution is spherically symmetrical. Therefore, if x_1, \dots, x_m are sampled independently from a standard normal distribution, the vector $\mathbf{x}' = \mathbf{x}/\|\mathbf{x}\|$, where $\mathbf{x} = (x_1, \dots, x_m)$ and $\|\mathbf{x}\| = (\sum_{i=1}^m x_i^2)^{\frac{1}{2}}$, is uniformly distributed on an m -sphere [12]. In order to obtain points that are uniformly distributed over an m -dimensional (hyper-)ball (also called an m -ball), we start by sampling a vector \mathbf{x}' uniformly from an m -sphere (see above) and in addition generate a (pseudo-)random number u that is uniformly distributed in $[0, 1)$. Since the radius r of a random vector that is uniformly distributed over an m -ball satisfies $P(r \leq z) = z^m$, we can write $r = u^{1/m}$. Therefore the vector $\mathbf{x}'' = r\mathbf{x}'$ is uniformly distributed over an m -ball [12].

In addition to these basic data generation modes, our implementation supports sampling uniformly from a regular grid of small (hyper-)cubes with a user-specified size. With this method we try to obtain a low-dimensional space with a large boundary in order to show that the hubness phenomenon can be produced in this way as well. The procedure is essentially the same as for sampling uniformly from a hypercube, only that the hypercube is cut into the requested number of small cubes and gaps are introduced by an appropriate transformation of the coordinates. To be precise: given a number η of (hyper-)cubes per dimension and a fraction α , which specifies how much (per dimension) of a grid cell is covered by the small (hyper-)cubes, we sample m coordinates x_1, \dots, x_m uniformly from $[0, 1)$ and transform them according to $x'_i = 2(\lfloor \eta x_i \rfloor + \alpha(\eta x_i - \lfloor \eta x_i \rfloor)) / (\eta - 1 + \alpha) - 1$ for $i = 1, \dots, m$. An example of such a 2-dimensional grid-structured data set with 2000 points, which was generated with $\eta = 5$ and $\alpha = 0.5$, is shown in Figure 1.

Finally, our implementation supports jolting points that were sampled uniformly from a (hyper-)cube into a (hyper-)ball. Intuitively, this can be seen

as “pushing in” the corners of the hypercube. Technically, this is achieved as follows: let $\mathbf{x} = (x_1, \dots, x_m)$ be a data point in a (hyper-)cube. We determine $z = \max_{i=1, \dots, m} |x_i|$ as well as $\|\mathbf{x}\| = (\sum_{i=1}^m x_i^2)^{\frac{1}{2}}$. Then $\mathbf{x}' = z\mathbf{x}/\|\mathbf{x}\|$ lies inside a (hyper-)sphere with radius 1. Of course, the distribution of the resulting points is not uniform in the (hyper-)sphere, but denser near the axes from the center towards the corners of the original (hyper-)cube. However, it is very interesting to see how this transformation affects the hubness phenomenon, as the result is not quite what one might expect.

4 Experimental Results

With the four experiments we describe in the following we try to clarify the inherent properties of the hubness phenomenon. In the first experiment we show that hubness need not occur in high-dimensional spaces by sampling from a finite, but boundary-less space. We demonstrate that it is rather directly related to a density gradient. In the second experiment we show that hubs also occur in low-dimensional spaces and reveal the true cause of the hubness phenomenon. The third experiment demonstrates the dependence of the hubness phenomenon on artificially introduced density gradients, especially the size of the surface of the sampling region. Finally, the fourth experiment examines the effect of jolting a (hyper-)cube into a (hyper-)ball.

Experiment 1: As already reported in [11], the distribution of the k -occurrences becomes skewed to the right if a data set is sampled uniformly from a high-dimensional hypercube, and even more so if the data set is sampled from a high-dimensional normal distribution. Our experiments confirm this observation, as can be seen from the curves labeled “cube” and “normal” in Figure 2.³ For $m \geq 15$, and certainly for $m \geq 20$, all three hubness measures (skewness γ , $h_1(1\%)$, i.e. the average k -occurrence, divided by k , of the data points with the 1% highest k -occurrences, and $h_2(3)$, i.e. the percentage of data points occurring in a least $3k = 30$ nearest neighbor lists) clearly indicate a strongly skewed distribution and the existence of sizable hubs.

It is remarkable that data sampled from a normal distribution exhibit a much stronger hubness phenomenon than data sampled uniformly from a hypercube. This finding already provides a hint that the hubness phenomenon is caused by a density gradient, because a normal distribution possesses a strong density gradient everywhere, while data sampled from a hypercube possess such a gradient only at its hyperfaces. As these faces cover a considerable hyperarea for high-dimensional data, many data points are influenced by the gradient they cause, but still fewer than in a normal distribution.

³ All such diagrams in this paper have been obtained by averaging over 200 runs in order to reduce the effects of randomness and to achieve representative results.

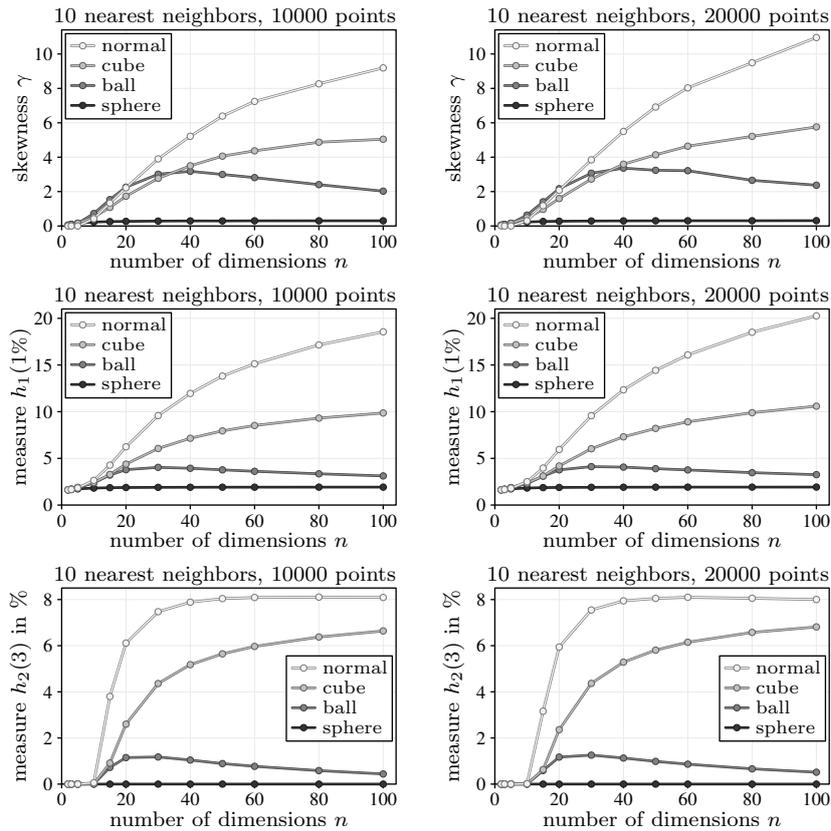


Fig. 2 The hubness phenomenon for different sampling distributions (multivariate normal and uniformly from a (hyper-)cube, (hyper-)ball and (hyper-)sphere) and two data set sizes, assessed by different hubness measures (as defined in Section 2).

A further hint is provided by the fact that data sampled uniformly from a hyperball (see curves labeled “ball” in Figure 2) exhibit a much lesser hubness phenomenon, which even is reduced again beyond $m \approx 30$ –40. Since the sample is still drawn uniformly, the different strength of the hubness phenomenon must be due to the different shape of the sampling region. We believe that the absence of “corners” (at which the density gradient is particularly high) and the much smaller hypersurface relative to the enclosed hypervolume are the reason for this effect. Since the hypersurface of a hyperball is much smaller compared to that of a hypercube, fewer data points are affected.

However, the strongest argument that high dimensionality alone does not cause a hubness effect is provided by the following consideration: an $(m + 1)$ -dimensional hypersphere is essentially an m -dimensional space, but finite and boundary-less. As a consequence, there is no density gradient anywhere, at least if we confine ourselves to the topology of the hypersphere (which is

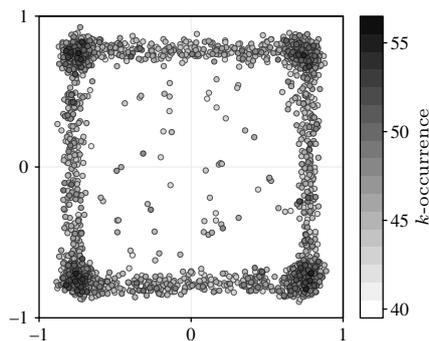


Fig. 3 The 12 largest hubs collected from 250 sets of 1000 points each, sampled uniformly from a square $[-1, 1]^2$. For each point the 30 nearest neighbors were considered. The darkness of a circle encodes a hub's k -occurrence.

equivalent to using an elliptic geometry). Although this does not pose any problems (the shortest path between two points follows the meridian through them and thus, for a unit hypersphere, its length is equal to the angle between the points, measured in radians), we rely on the Euclidean distance in the $(m + 1)$ -dimensional space, which yields essentially the same result. As can be seen from the curves labeled “sphere” in Figure 2, none of the measures detects a hubness phenomenon, regardless of the number of dimensions.

Note that this finding explains why a (hyper-)ball exhibits a much less pronounced hubness effect: for increasing dimensionality the mean Euclidean norm of points sampled uniformly from a unit (hyper-)ball converges to 1 (see Section 3: $P(r \leq z) = z^m$). Thus we may say that in high-dimensional spaces almost all points in a (hyper-)ball lie close to its surface and thus almost on a (hyper-)sphere. With this regard it is no longer surprising that the hubness phenomenon reduces again for very high-dimensional (hyper-)balls.

Experiment 2: In our second experiment we show that the hubness phenomenon also occurs in low dimensional spaces, although not as pronounced, and reveal its true cause. In the diagram in Figure 3 the twelve largest hubs have been collected from 250 data sets that were sampled uniformly from a square. The darkness of the hubs encodes their k -occurrence, which shows that there are not only *more hubs* close to the corners, but that these hubs also tend to be *larger* (that is, they tend to have higher k -occurrences).

This effect is clearly due to the boundaries of the square and can be seen as a kind of mirroring effect. Points close to the sides and corners do not have as many options to choose their nearest neighbors compared to points in the interior of the square. Therefore points that lie near points that are close to the sides and corners are more likely to be chosen as nearest neighbors and thus become hubs. Hubs are more frequent close to the corners, because here up to $3/4$ of the space (for a point exactly at the corner of square) are void of points, while in the middle of a side only up to $1/2$ of the space is void of points. If one extrapolates this finding to more dimensions, it becomes clear why hypercubes exhibit such a strong hubness phenomenon.

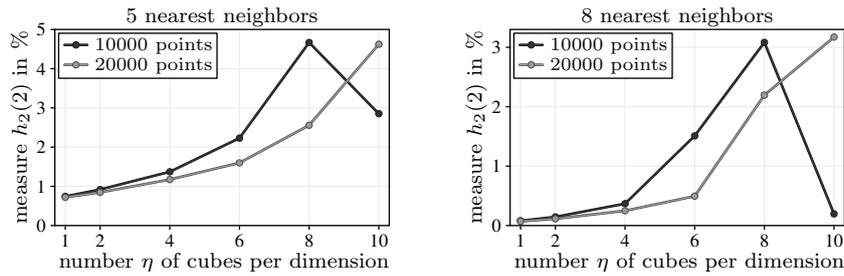


Fig. 4 Hubness in 3-dimensional data sets sampled uniformly from a grid of cubes: dependence on the number of cubes/grid cells per dimension (gap size $\alpha = 50\%$).

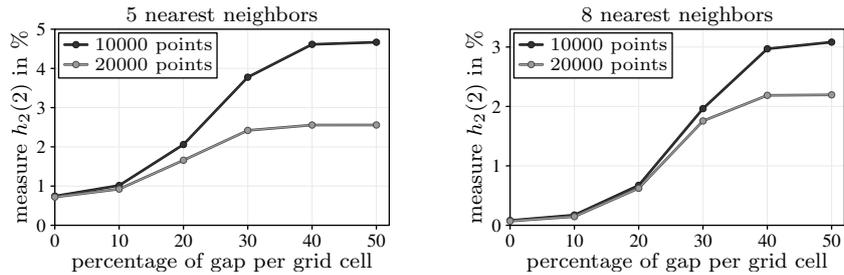


Fig. 5 Hubness in 3-dimensional data sets sampled uniformly from a grid of cubes ($\eta = 8$ cubes per dimension): dependence on the size of the gaps (measured per dimension).

Experiment 3: Our third experiment expands on our view that the size (and shape) of the (hyper-)surface (and the density gradient it causes) produces the hubness phenomenon. If this view is correct, it should be possible to create a hubness phenomenon in a low-dimensional space by sampling from a region with a large boundary. Our core idea is to sample data from a grid of squares or cubes. If the gaps between these squares or cubes are big enough, so that a nearest neighbor is almost surely found in the same cube, there should also be a certain, though weaker hubness phenomenon.

This effect is demonstrated in Figure 4, which shows how a grid of cubes (gap size $\alpha = 50\%$) leads to a hubness effect with an increasing number of cubes. The effect is weak, though, but can be detected with the skewness γ or with the measure $h_2(2)$ used in the diagrams. Note that the effect is generally bounded due to the topology of a 3-dimensional space, as can be seen from the relation of the hubness phenomenon to the kissing number problem [3, 9]. A kissing number is the number of non-overlapping unit spheres that touch another given unit sphere. As was already noticed in [11], this problem is relevant for the hubness phenomenon, because a data point cannot be the nearest neighbor of more data points than the kissing number of the space it resides in. Since the kissing number for three dimensions is 12, sizable hubs are extremely unlikely as they require highly symmetric point arrangements.

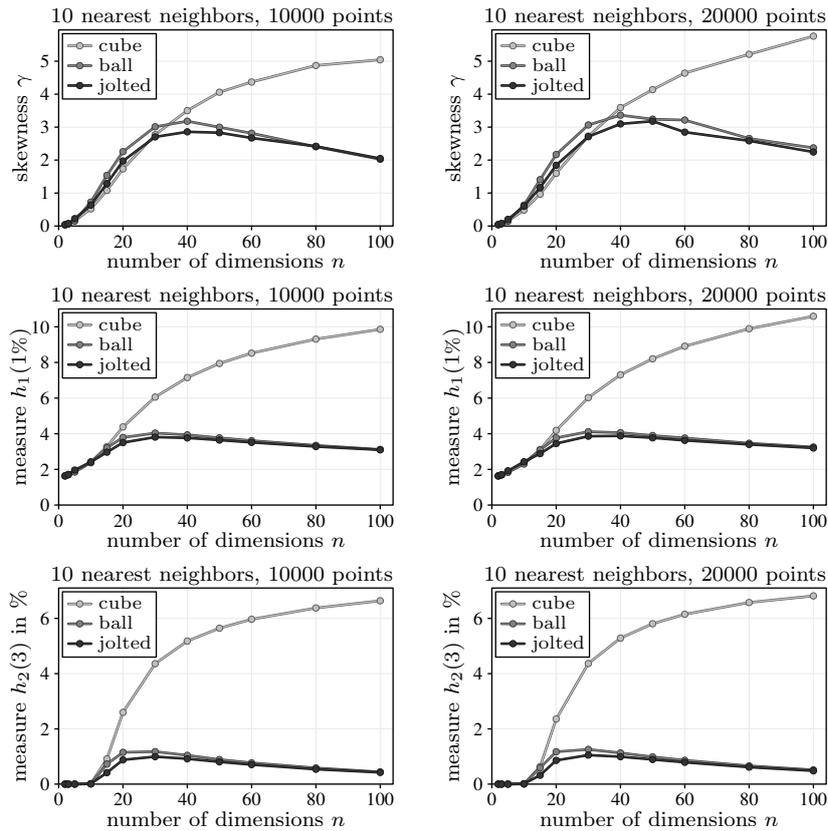


Fig. 6 The hubness phenomenon for data that was sampled uniformly from a (hyper-)cube and then jolted into a (hyper-)ball, as compared to sampling uniformly from a (hyper-)cube or (hyper-)ball, assessed with the same measures as in Figure 2.

Note that for 10000 points the hubness phenomenon disappears again for 10 cubes per dimension. The reason for this effect is simply that under these circumstances the average number of points per cube is 10 (as there are $10 \times 10 \times 10 = 1000$ cubes). If 8 nearest neighbors are considered, almost all points in a cube are the nearest neighbors of all other points in the same cube. As a consequence, the location of the points relative to the boundary, which is responsible for the hubness effect (see Experiment 2), becomes irrelevant.

Note also that the gaps between the cubes have to be large enough, as can be seen in Figure 5: for small gaps there is basically no hubness effect, because nearest neighbors may still be found in neighboring cubes, thus reducing or even eliminating the effect of the cube surfaces/boundaries.

Experiment 4: Our last experiment reveals a somewhat unexpected behavior that we discovered during our analysis. As we have seen in Experiment 1, the skewness of the distribution of k -occurrences for data sampled uniformly

from a (hyper-)cube is much stronger than for data sampled uniformly from a (hyper-)ball. This led to the idea to “jolt” a (hyper-)cube into a (hyper-)ball in order to check whether this operation reduces the hubness phenomenon. Given our view of the causes of the hubness phenomenon, we certainly expected it to be reduced (because this operation significantly reduces the surface of the sampling region), but that it was even reduced slightly below the level of data that was sampled uniformly from a (hyper-)ball was somewhat surprising (see Figure 6, curve labeled “jolted”). We rather expected it to lie between a (hyper-)cube and a (hyper-)ball based on the argument that the jolting introduces density gradients inside the (hyper-)ball.

However, on second thought, the effect becomes understandable. The jolting operation, even though it causes density gradients inside the (hyper-)ball, also reduces the effect of the (remaining) (hyper-)surface, because it pushes more data points into the interior of the (hyper-)ball, thus leaving less at the surface that cause the hubness effect (cf. Experiment 2).

5 Conclusions

In this paper we demonstrated that the hubness phenomenon is not an effect of the (high) dimensionality of a data set, but an effect of a density gradient. However, a density gradient may be intrinsic to the data set (if the data is not uniformly distributed) or it may be a boundary effect. Since a boundary necessarily introduces a density gradient and the ratio of the size of the boundary to the size of the enclosed space grows exponentially with the dimensionality of the data space, high-dimensional bounded data is prone to exhibit the hubness phenomenon. However, it is important to note that this phenomenon can also be produced, though much weaker, in a low-dimensional space by artificially increasing the size of the boundary. Another factor is the shape of the boundary: “corners” intensify the effect, as can be seen from the pronounced hubness phenomenon exhibited by (hyper-)cubes.

Software

Our implementation of the data generator and evaluation routines, which we used for the experiments in this paper as well as the corresponding Python scripts automating the experiments will soon be available for download at: <http://www.borgelt.net/hubness.html>

Acknowledgements The work presented in this paper was supported by Short-Term Scientific Mission (STSM) grant 7684 (Thomas Low) of COST Action IC0702.

References

1. Bellman RE (1961) *Adaptive Control Processes: A Guided Tour*. Princeton University Press, Princeton, NJ, USA 1961
2. Beyer K, Goldstein J, Ramakrishnan R, Shaft U (1999) When is “Nearest Neighbor” Meaningful? *Proc. 7th Int. Conf. on Database Theory (ICDT 1999, Jerusalem, Israel)*, LNCS 1540:217–235. Springer-Verlag, Berlin, Germany
3. Conway JH, Sloane NJA (1999) *Sphere Packings, Lattices and Groups (3rd edition)*. Springer-Verlag, New York, NY, USA
4. Groeneveld RA, Meeden G (1984) Measuring Skewness and Kurtosis. *J. of the Royal Statistical Society, Series D (The Statistician)* 33(4):391399. Blackwell Publishing, Oxford, United Kingdom
5. Knuth DE (1998) *The Art of Computer Programming, Vol. 2: Seminumerical Algorithms*. Addison-Wesley, Reading, MA, USA
6. Marsaglia G (2003) Re: good C random number generator. Post on newsgroup `comp.lang.c`, date: 2003-05-13 08:55:05 PST. http://groups.google.com/group/comp.lang.c/browse_thread/thread/a9915080a4424068/
7. Marsaglia G, Bray TA (1964) A Convenient Method for Generating Normal Variables. *SIAM Review* 6:260-264. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA
8. Matsumoto M, Nishimura T (1998) Mersenne Twister: A 623-dimensionally Equidistributed Uniform Pseudorandom Number Generator. *ACM Trans. on Modeling and Computer Simulation* 8:3-30. ACM Press, New York, NY, USA
9. Nebe G, Sloane NJA (2012) *Table of the Highest Kissing Numbers Presently Known*. <http://www.math.rwth-aachen.de/~Gabriele.Nebe/LATTICES/kiss.html> (retrieved 2012.01.16)
10. Radovanović M, Nanopoulos A, Ivanović M (2009) Nearest Neighbors in High-Dimensional Data: The Emergence and Influence of Hubs. *Proc. 26th Int. Conf. on Machine Learning (ICML 2009, Montreal, Canada)*, 865–872. ACM Press, New York, NY, USA
11. Radovanović M, Nanopoulos A, Ivanović M (2010) Hubs in Space: Popular Nearest Neighbors in High-Dimensional Data. *J. Machine Learning Research* 11(Sep):2487–2531. MIT Press, Cambridge, MA, USA
12. Rubinstein RY, Kroese DP (2007) *Simulation and the Monte Carlo Method (2nd ed.)*. J. Wiley & Sons, Chichester, United Kingdom