JOURNAL OF COMBINATORIAL THEORY, Series B 44, 355–362 (1988)

The Johnson–Lindenstrauss Lemma and the Sphericity of Some Graphs

P. FRANKL* AND H. MAEHARA†

*CNRS, Paris, France, and †Ryukyu University, Okinawa, Japan

Communicated by the Managing Editors

Received May 2, 1986

A simple short proof of the Johnson–Lindenstrauss lemma (concerning nearly isometric embeddings of finite point sets in lower-dimensional spaces) is given. This result is applied to show that if G is a graph on n vertices and with smallest eigenvalue λ then its sphericity sph(G) is less than $c\lambda^2 \log n$. It is also proved that if G or its complement is a forest then sph$(G) \leqslant c \log n$ holds. © 1988 Academic Press, Inc.

1. Some Upper Bounds on the Sphericity of Graphs

We state a slightly improved version of the Johnson–Lindenstrauss lemma [4]. The proof—which is considerably shorter—will be given later.

LEMMA. *For an* ε $(0<\varepsilon<\frac{1}{2})$ *and an integer* n, *let* $k(n,\varepsilon)=\lceil 9(\varepsilon^2-2\varepsilon^3/3)^{-1}\log n\rceil+1$. *If* $n>k(n,\varepsilon)^2$, *then for any* n*-point set* S *in* R^n, *there exists a map* $f\colon S\to R^{k(n,\varepsilon)}$ *such that*

$$(1-\varepsilon)\,\|u-v\|^2<\|f(u)-f(v)\|^2 <(1+\varepsilon)\,\|u-v\|^2 \qquad \text{for all} \quad u, v \text{ in } S.$$

Remark. In [4] the constant 9 is not specified.

We are going to apply this lemma to the *sphericity problem* (see [2, 5–8], and also [10], where similar problems were considered). The *sphericity* of a graph $G=(V,E)$, sph(G), is the smallest integer n such that there is an embedding $f\colon V\to R^n$ such that $0<\|f(u)-f(v)\|<1$ if and only if $uv\in E$. An eigenvalue of a graph G is an eigenvalue of its adjacency matrix $A(G)$. $|G|$ denotes the number of vertices of G.

THEOREM 1. *Let* G *be a graph with minimum eigenvalue* $\lambda_{\min}\geqslant -c$ $(c\geqslant 2)$ *and suppose that* $|G|>[12(2c-1)^2\log|G|]^2$. *Then* sph$(G)<12(2c-1)^2\log|G|$.

0095-8956/88 $3.00

Proof. Let $n=|G|$ and $A(G)=(a_{ij})$ be the adjacency matrix of G. Then $A(G)+cI$ is positive semi-definite, where I is the identity matrix. Hence we can write $A(G)+cI=M\cdot M^t$. Let x_i be the ith row of M. Then $\|x_i-x_j\|^2=2c-2a_{ij}$. Let $\varepsilon=1/(2c-1)$ and let $S=\{x_i\in R^n\colon i=1, ..., n\}$. Since

$$k(n, \varepsilon)<11.58(2c-1)^2 \log |G|<12(2c-1)^2 \log |G| \text{ for } c\geqslant 2,$$

applying the lemma, we can conclude that there exists a map $f\colon S\to R^{k(n,\varepsilon)}$ such that

$$\|f(x_i)-f(x_j)\|^2<2c(1-\varepsilon) \qquad \text{iff} \quad a_{ij}=1.$$

Now setting $g(x_i)=(2c(1-\varepsilon))^{-1/2}f(x_i)$, we have

$$\|g(x_i)-g(x_j)\|^2<1 \qquad \text{iff} \quad a_{ij}=1.$$

Thus $\operatorname{sph}(G)<12(2c-1)^2\log|G|$. ∎

Reiterman, Rödl, and Šiňajová [12] showed by a different method that if G is a graph with maximum degree d, then

$$\operatorname{sph}(G)\leqslant 16(d+1)^3\log(8\,|G|\,(d+1)).$$

Our next result is an improvement of this bound.

COROLLARY 1. *Let G be a graph with maximum degree d and suppose $|G|>[12(2d-1)^2\log|G|]^2$. Then* $\operatorname{sph}(G)<12(2d-1)^2\log|G|$.

Proof. If the maximum degree of a graph G is at most d, then the maximum eigenvalue $\lambda_{\max}$ of G is also at most d (see, e.g., [14]). Since $\lambda_{\min}\geqslant-\lambda_{\max}$ holds generally, we have $\lambda_{\min}\geqslant -d$. Hence the corollary follows from Theorem 1. ∎

Let $L(G)$ denote the line graph of G. Then it is well known that $\lambda_{\min}\geqslant -2$ (see, e.g., [14]). This implies the next result.

COROLLARY 2. *Let G be a graph with m edges. Then*

$$\operatorname{sph}(L(G))<108\log m \qquad \textit{for} \quad m>(108\log m)^2.$$

THEOREM 2. *Let T be a tree with sufficiently large order. Then* $\operatorname{sph}(T)<105\log|T|$.

Proof. Let v_i $(i=1, ..., n)$ be the vertices of T. For each v_i, there is a unique path P_i from v_1 to v_i. Let $x_i=(s_1, ..., s_n)$ in R^n, where $s_j=1$ if v_j appears in P_i and $s_j=0$ otherwise. Then the set $S=\{x_i\colon i=1, ..., n\}$

satisfies that $\|x_i - x_j\|^2 = 1$ if v_i and v_j are adjacent; $\geqslant 2$ otherwise. Hence letting $\varepsilon = \frac{1}{3}$ and applying the lemma, we have a map $f: S \to R^{k(n,\varepsilon)}$ such that $\|f(x_i) - f(x_j)\|^2 < \frac{4}{3}$ if v_i and v_j are adjacent, $> \frac{4}{3}$ otherwise. Now letting $g(x_i) = (\frac{4}{3})^{-1/2} f(x_i)$, we have $\|g(x_i) - g(x_j)\| < 1$ if and only if v_i and v_j are adjacent. Since $k(n, \varepsilon) = \lceil (729/7) \log n \rceil + 1 < 105 \log n$, we have the theorem. ▮

Remark. Using a different method we could improve the constant 105 to 7.3; see [3].

Concerning the sphericity of the complement of a tree or a forest, we have the following.

THEOREM 3. *For any forest F*, $\mathrm{sph}(\bar{F}) \leqslant 8 \lceil \log_2 |F| \rceil$.

Proof. The proof is based on the result of Poljak and Pultr [11]. They defined the "product" K_k^r of r copies of the complete graph K_k as the graph with vertices $\{x = (x_1, ..., x_r)\colon x_i \in V(K_k)\}$ and the edges $\{xy\colon x_i \neq y_i$ for every $i\}$. They proved then that each forest F can be embedded as an induced subgraph in K_3^r, where $r = 4 \lceil \log_2 |F| \rceil$. Now, let G be the complement of K_3^r. We show $\mathrm{sph}(G) \leqslant 2r$. Let u, v, w be the vertices of an equilateral triangle of sidelength $(1/r)^{1/2}$ in R^2 centered at the origin of R^2. Let $X = \{(s_1, ..., s_r)\colon s_i = u$ or v or $w\} \subset R^2 \times \cdots \times R^2 = R^{2r}$. If we connect each pair of points of X by a line segment whenever their distance is less than 1, then we have a geometric graph isomorphic to G. Hence $\mathrm{sph}(G) \leqslant 2r$, and hence $\mathrm{sph}(\bar{F}) \leqslant 8 \lceil \log_2 |F| \rceil$. ▮

Remark. Recently, Reiterman, Rödl, and Šiňajová [13] proved that $\mathrm{sph}(\bar{F}) \leqslant 6$ for every forest F. This result is further improved in [9] to $\mathrm{sph}(\bar{F}) \leqslant 3$.

2. A SIMPLE SHORT PROOF OF THE JOHNSON–LINDENSTRAUSS LEMMA

Let $\mathbf{v}$ be a unit vector in R^n and H a "random k-dimensional subspace" through the origin, and let us define the random variable X as the square length of the projection of $\mathbf{v}$ onto H.

PROPOSITION. *Suppose* $\frac{1}{2} > \varepsilon > 0$, $n > k^2$, $k > 24 \log n + 1$. *Then* $P_\varepsilon = \mathrm{Prob}(|X - k/n| > \varepsilon k/n) < 2\sqrt{k} \exp(-(k-1)(\varepsilon^2/4 - \varepsilon^3/6))$.

Proof. First we state a few easy consequences from the above restriction on k, n for later use:

$$k/n < 1/20, \qquad kn > (5\pi)^2, \qquad \sqrt{(k-1)/32} > 1.4.$$

Note that to compute the above probability we can reverse the roles and take a fixed k-space H and then a random unit vector $\mathbf{v}$ (uniformly distributed on the surface of the unit sphere in R^n). Let Θ be the angle between $\mathbf{v}$ and H. Then $X = \cos^2 \Theta$. Thus the event we are interested in is

$$\Theta \notin [\arccos \sqrt{(1+\varepsilon)k/n}, \arccos \sqrt{(1-\varepsilon)k/n}].$$

Let V_i denote the surface area of the unit sphere in R^i. We use the following formula (a proof of which will be given later):

$$V_n = \int_0^{\pi/2} V_k (\cos\theta)^{k-1} V_{n-k} (\sin\theta)^{n-k-1}\, d\theta \qquad \text{(valid for all } 1 \leqslant k < n). \tag{1}$$

Let $A(t) = \arccos \sqrt{(1+t)k/n}$. Then

$$P_\varepsilon = \left(\int_0^{A(\varepsilon)} V_k V_{n-k} f(\theta)\, d\theta + \int_{A(-\varepsilon)}^{\pi/2} V_k V_{n-k} f(\theta)\, d\theta \right) \Big/ V_n,$$

where $f(\theta) = (\cos\theta)^{k-1} (\sin\theta)^{n-k-1}$. Let us estimate the value of $f(\theta)$ for $\theta = A(t) = \arccos \sqrt{(1+t)k/n}$.

$$\begin{aligned} f(\theta) &= ((1+t)k/n)^{(k-1)/2} (1-(1+t)k/n)^{(n-k-1)/2} \\ &= \underbrace{(k/n)^{(k-1)/2} ((n-k)/n)^{(n-k-1)/2}}_{C} \\ &\quad \times \underbrace{(1+t)^{(k-1)/2} (1-tk/(n-k))^{(n-k-1)/2}}_{B}. \end{aligned}$$

Using the inequalities

$$1+t < \exp(t - t^2/2 + t^3/3), \qquad (1-tk/(n-k))^{(n-k)/(tk)} < 1/e$$

we obtain

$$B < \exp(t - t^2/2 + t^3/3)(k-1)/2) \exp(-(tk/2)(n-k-1)/(n-k)).$$

Since $(n-k-1)/(n-k) > (k-1)/k$ for $n > 2k$, we infer

$$B < \exp(-(k-1)(t^2/4 - t^3/6)).$$

Since

$$A'(t) = dA(t)/dt = -(4(1+t)(1-(1+t)k/n))^{-1/2} (k/n)^{1/2},$$

$$A''(t) = d^2A(t)/dt^2 > 0,$$

and since $k/n < 1/20$, using the mean value theorem, we have

$$|A(-\tfrac{1}{2}) - A(-\varepsilon)| = |(\tfrac{1}{2} - \varepsilon)\, A'(\xi)| < \tfrac{1}{2}\, |A'(-\tfrac{1}{2})|$$
$$= (k/n)^{1/2}\,(8 - 4k/n)^{-1/2} < 0.36(k/n)^{1/2}.$$

Similarly

$$|A(\varepsilon) - A(1)| < |A'(0)| = (n/k)^{1/2}\,(4 - 4k/n)^{-1/2}$$
$$< 0.52(k/n)^{1/2}.$$

Hence

$$\int_{A(1)}^{A(\varepsilon)} B\, d\theta + \int_{A(-\varepsilon)}^{A(-1/2)} B\, d\theta$$
$$< 0.9(k/n)^{1/2} \exp(-(k-1)(\varepsilon^2/4 - \varepsilon^3/6)). \qquad (2)$$

On the other hand, since $f(\theta)$ is "unimodal" with maximum value at $\theta = \arccos\sqrt{(k-1)/(n-2)}$, it follows that for $t < -\frac{1}{2}$ or $t > 1$, B is less than $\exp(-(k-1)/12)$. Hence

$$\int_{0}^{A(1)} B\, d\theta + \int_{A(-1/2)}^{\pi/2} B\, d\theta < (\pi/2)\exp(-(k-1)/12). \qquad (3)$$

Since $\varepsilon \leqslant \frac{1}{2}$ and $k - 1 > 24 \log n$,

$$(\pi/2)\exp(-(k-1)/12) < (\pi/2)\exp(-(k-1)(\varepsilon^2/4 - \varepsilon^3/6 + 1/24))$$
$$< (\pi/2)n^{-1} \exp(-(k-1)(\varepsilon^2/4 - \varepsilon^3/6))$$
$$< 0.1(k/n)^{1/2} \exp(-(k-1)(\varepsilon^2/4 - \varepsilon^3/6)).$$

Combining (2) and (3) we get that the numerator of P_ε is less than

$$V_k V_{n-k} C(k/n)^{1/2} \exp(-(k-1)(\varepsilon^2/4 - \varepsilon^3/6)).$$

Now we estimate V_n. Using the inequalities

$$\exp(t - t^2/2) < 1 + t \qquad (t > 0),$$
$$1/e < (1 - tk/(n-k))^{((n-k)/(tk)) - 1}$$

we have that for $t > 0$,

$$B > \exp(t(k-1)/2 - t^2(k-1)/4)\exp(-tk(n-k-1)/(2(n-k-tk)))$$
$$= \exp(-(k-1)t^2/4 - (t/2)(1 + (tk^2 - k)/(n-k-tk))).$$

Thus $B > e^{-1/4} \exp(-(k-1)t^2/4)$ for $0 < t < \frac{1}{4}$. Since

$$|d\theta/dt| = |A'(t)| > (k/(5n))^{1/2} \qquad \text{for} \quad 0 < t < \tfrac{1}{4}, n > k^2,$$

we have

$$\int_{A(1/4)}^{A(0)} B\, d\theta > \int_0^{1/4} (k/(5n))^{1/2} e^{-1/4} \exp(-(k-1)t^2/4)\, dt$$

(letting $\sigma = (2/(k-1)^{1/2})$

$$= e^{-1/4}(k/(5n)^{1/2} (2\pi)^{1/2} \sigma \int_0^{1/4} (2\pi)^{-1/2} \sigma^{-1} \exp(-t^2/(2\sigma^2))\, dt.$$

(Using the standard normal distribution function $\Phi(x)$, the last integral is represented as $\Phi(1/(4\sigma)) - \Phi(0)$.) Since $1/(4\sigma) = ((k-1)/32)^{1/2} > 1.4$, and $\Phi(1.4) - \Phi(0) = 0.4192$, this is greater than

$$2e^{-1/4}(\pi/5)^{1/2} n^{-1/2}(0.4192) > (4n)^{-1/2}.$$

Thus $V_n > V_k V_{n-k} C(4n)^{-1/2}$. Therefore

$$P_\varepsilon < 2\sqrt{k} \exp(-(k-1)(\varepsilon^2/4 - \varepsilon^3/6)). \quad \blacksquare$$

Proof of Formula (1). We have

$$\int_0^{\pi/2} \sin^{a-1}\theta \cos^{b-1}\theta\, d\theta = \tfrac{1}{2}\Gamma(a/2)\,\Gamma(b/2)/\Gamma((a+b)/2)$$

(cf. [1, Sect. 534, Exercise 4a]). On the other hand, the surface of an n-dimensional sphere of radius 1 is

$$V_n = 2\pi^{n/2}/\Gamma(n/2)$$

(cf. [1, Sect. 676, Exercise 3]). Thus

$$V_{n-k} V_k = 4\pi^{n/2}/(\Gamma(k/2)\,\Gamma((n-k)/2))$$

and hence

$$V_n/(V_k V_{n-k}) = \tfrac{1}{2}\Gamma(k/2)\,\Gamma((n-k)/2))/\Gamma(n/2)$$

$$= \int_0^{\pi/2} \sin^{k-1}\theta \cos^{n-k-1}\theta\, d\theta. \quad \blacksquare$$

Proof of the Lemma. Let $S = \{v_1, ..., v_n\} \subset R^n$ and let H be a random k-space in R^n, where $k = k(n, \varepsilon)$. Let w_i be the projection of v_i on H, $i = 1, ..., n$. We denote the event

$$| \, \|w_i - w_j\|^2 / \|v_i - v_j\|^2 - k/n| > \varepsilon k/n$$

by E_{ij}. Then by the above proposition,

$$\text{Prob}(E_{ij}) < 2\sqrt{k} \exp(-(k-1)(\varepsilon^2/4 - \varepsilon^3/6)) \qquad \text{for} \quad i \neq j.$$

Hence the probability that E_{ij} occurs for some $i \neq j$ is less than

$$\binom{n}{2} 2\sqrt{k} \exp(-(k-1)(\varepsilon^2/4 - \varepsilon^3/6)) < n^{9/4} \exp(-\log n^{9/4}) = 1.$$

Therefore there exists a k-space H in R^n for which

$$(1-\varepsilon)k/n < \|w_i - w_j\|^2 / \|v_i - v_j\|^2 < (1+\varepsilon)k/n \quad (i \neq j)$$

i.e.,

$$(1-\varepsilon) \|v_i - v_j\|^2 < (n/k) \|w_i - w_j\|^2 < (1+\varepsilon) \|v_i - v_j\|^2 \quad (i \neq j).$$

Hence, letting $f(v_i) = \sqrt{n/k}\, w_i$ $(i = 1, ..., n)$, we obtain a desired embedding of S in k-dimension. ▌

Acknowledgments

We thank the referees for many helpful suggestions and V. Rödl for the present proof of (1).

References

1. G. M. Fichtenholz, "Introduction to Differential and Integral Calculus," Moscow, 1966.
2. P. Frankl and H. Maehara, Embedding the n-cube in lower dimensions, *Eur. J. Combin.* **7** (1986), 221–225.
3. P. Frankl and H. Maehara, On the contact dimension of graphs, *Discrete Comput. Geom.*, in press.
4. W. B. Johnson and J. Lindenstrauss, Extensions of Lipshitz mapping into Hilbert space, *Contemp. Math.* **26** (1984), 189–206.
5. H. Maehara, Space graphs and sphericity, *Discrete Appl. Math.* **7** (1984), 55–64.
6. H. Maehara, On the sphericity for the join of many graphs, *Discrete Math.* **49** (1984), 311–313.
7. H. Maehara, On the sphericity of the graphs of semiregular polyhedra, *Discrete Math.* **58** (1986), 311–315.

8. H. MAEHARA, Sphericity exceeds cubicity for almost all complete bipartite graphs, *J. Combin. Theory Ser. B* **40** (1986), 231–235.
9. H. MAEHARA, J. REITERMAN, V. RÖDL, AND E. ŠIŇAJOVÁ, Embedding of trees in Euclidean spaces, *Graphs Combin.* **4** (1988), 43–47.
10. J. PACH, Decomposition of multiple packing and covering, 2, *Kolloq. Diskr. Geom.* (1983), 69–78.
11. S. POLJAK AND A. PULTR, On the dimension of trees, *Discrete Math.* **34** (1981), 165–171.
12. J. REITERMAN, V. RÖDL, AND E. ŠIŇAJOVÁ, Geometrical embeddings of graphs, *Discrete Math.*, in press.
13. J. REITERMAN, V. RÖDL, AND E. ŠIŇAJOVÁ, Embeddings of graphs in Euclidean spaces, submitted for publication.
14. A. J. SCHWENK AND R. J. WILSON, On the eigenvalue of a graph, *in* "Selected Topics in Graph Theory" (L. W. Beineke and R. J. Wilson, Eds.), Academic Press, New York, 1978.