Implications of sparsity and high triangle density for graph representation learning

Hannah Sansford

September 7, 2022

Compass

EPSRC Centre for Doctoral Training in
Computational Statistics and Data Science

Talk structure

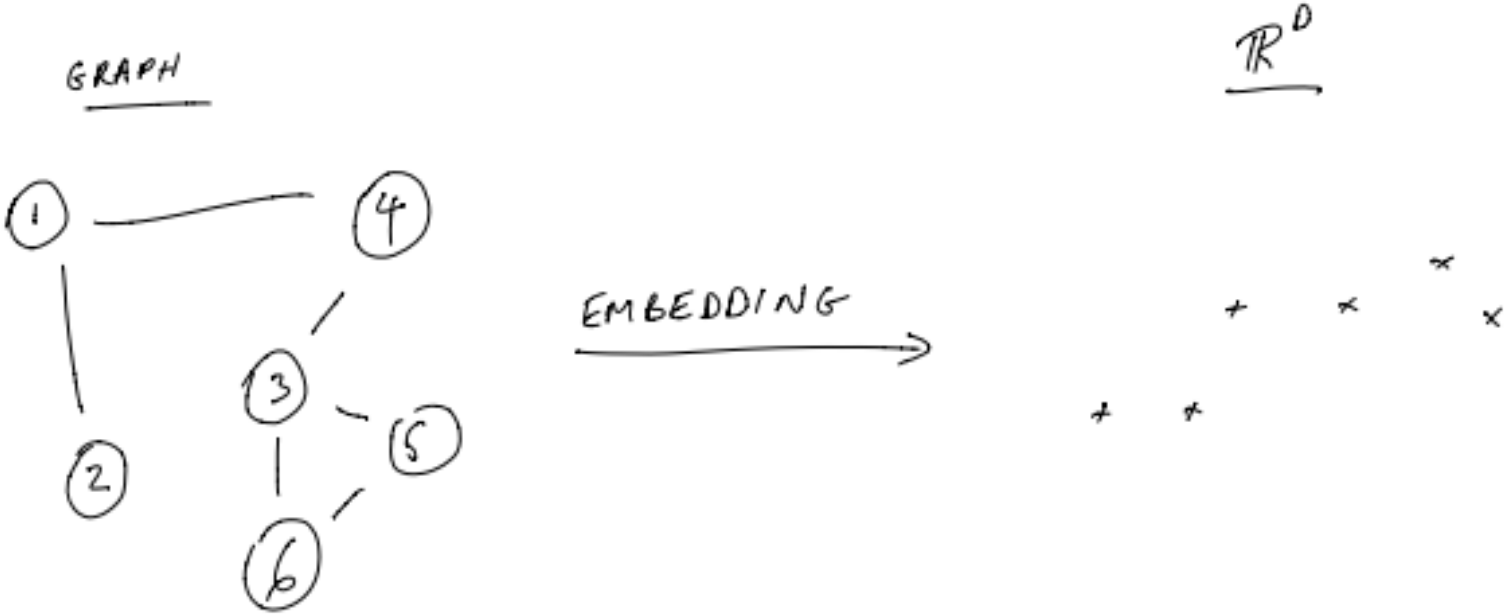
Mathematical prerequisites

Motivating work

Main result

Implications

Graph embedding



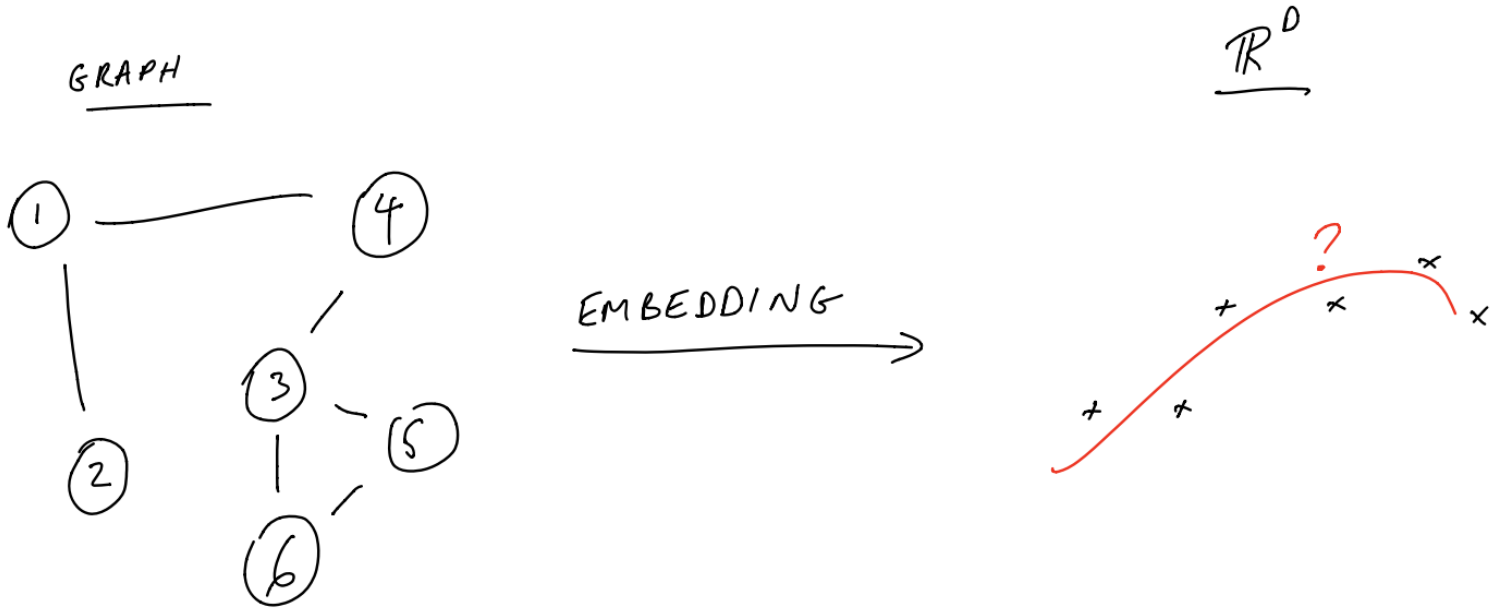
Adjacency spectral embedding

Input symmetric adjacency matrix $\mathbf{A}$ Compute $\mathbf{A} \approx \hat{\mathbf{X}}\hat{\mathbf{X}}^\top$ with

$\hat{\mathbf{X}} = [\hat{X}_1, \ldots, \hat{X}_n]^\top \in \mathbb{R}^{n \times D}$:

- compute truncated spectral decomposition $\mathbf{A} \approx \hat{\mathbf{U}}\hat{\mathbf{S}}\hat{\mathbf{V}}^\top$
- let $\hat{\mathbf{X}} = \hat{\mathbf{U}}\hat{\mathbf{S}}^{1/2}$

The manifold hypothesis



(Rubin-Delanchy, 2020)

Latent position model

Suppose the nodes have true latent positions $Z_1, \ldots, Z_n \in \mathcal{Z}_n$, where $\mathcal{Z}_n \subseteq \mathbb{R}^d$ is a compact topological manifold, and are i.i.d. from a probability distribution G_n.

Let $f : \mathcal{Z}_n \times \mathcal{Z}_n \to [0, 1]$ be a symmetric positive-definite function, called a kernel.

Next, suppose the graph's adjacency matrix satisfies

$$\mathbf{A}_{ij}^{(n)} \overset{ind}{\sim} \text{Bernoulli}\left\{f(Z_i, Z_j)\right\},$$

for $i < j$.

How is Z_i related to $\hat{X}_i$?

Random dot product graph (RDPG)

$\hat{X}_i$ estimates X_i, where

$$\mathbf{A}_{ij}^{(n)} \overset{ind}{\sim} \text{Bernoulli}\left\{\langle X_i, X_j \rangle\right\}, \quad \text{for } i < j,$$

and $\langle \cdot, \cdot \rangle$ denotes the (possibly indefinite) inner product, a model known as the (generalised) random dot product graph (Rubin-Delanchy et al., 2017).

Claim: The X_i are a high-dimensional but somehow faithful transformation of Z_i,

$$X_i = \phi(Z_i).$$

Infinite-dimensional RDPG

The inner product can be used to represent any positive definite kernel f (although typically in infinite dimension) using the associated Mercer feature map $\phi : \mathcal{Z}_n \to \mathbb{R}^{\mathbb{N}}$,

$$\langle \phi(x), \phi(y) \rangle = f(x, y), \quad \text{for all } x, y \in \mathcal{Z}_n,$$

defining $X_i := \phi(Z_i)$. Therefore, the latent position model also defines an infinite-dimensional RDPG.

Theorem

The map ϕ is a bi-Lipschitz homeomorphism. As a result, the X_i are supported on a topological manifold $\mathcal{M}_n := \phi(\mathcal{Z}_n)$ of the same dimension as $\mathcal{Z}_n$.

(Whiteley et al., 2021)

Motivating work

PNAS

RESEARCH ARTICLE | COMPUTER SCIENCES |

The impossibility of low-rank representations for triangle-rich complex networks

C. Seshadhri, Aneesh Sharma, Andrew Stolman, and Ashish Goel Authors Info & Affiliations

Edited by Mark E. J. Newman, University of Michigan, Ann Arbor, MI, and accepted by Editorial Board Member Peter J. Bickel February 1, 2020 (received for review June 26, 2019)

March 2, 2020 | 117 (11) 5631-5637 | https://doi.org/10.1073/pnas.1911030117

Intuition

Under a random dot product graph,

$$\mathbf{A}_{ij}^{(n)} \overset{ind}{\sim} \text{Bernoulli}\left\{\langle X_i, X_j \rangle\right\},$$

- the graph is sparse if most inner products are small
- inner products are small if either the X_i are small (low norm) or they are 'almost orthogonal'
- but we can't have arbitrarily many 'almost orthogonal' vectors in $\mathbb{R}^D$
- so for low D and large n, the X_i must be small to achieve sparsity, making triangles almost impossible

Sparsity and triangle density

Under a random dot product graph with i.i.d. latent positions X_i, define the graph sparsity factor, ρ_n, and triangle density, Δ_n, via

$$n\rho_n = n \iint \langle x, y \rangle F_n(x) F_n(y), \quad \text{(expected degree)}$$

$$n\Delta_n = \binom{n}{3} \iiint \langle x, y \rangle \langle y, z \rangle \langle z, x \rangle F_n(x) F_n(y) F_n(z),$$

(expected number of triangles)

where F_n is the distribution of $X_1, \ldots, X_n$.

Question

Can a random dot product graph be sparse, i.e., $\rho_n = o(1)$, with high triangle density, i.e., $\Delta_n = \Omega(1)$?

1. The answer is *no* if X_i are finite-dimensional (Seshadhri et al.), but:
2. The answer is *yes* if X_i are infinite-dimensional, yet supported on a low-dimensional manifold.

Proof approach

- Start with a compact topological manifold $\mathcal{Z}_n$, e.g. unit cube/sphere,
- build manifold $\mathcal{M}_n$ using a homeomorphism $\mathcal{M}_n := \phi(\mathcal{Z}_n)$ which is non-distortive,
- this results in $\mathcal{M}_n$ being a manifold of the same dimension as $\mathcal{Z}_n$,
- then can push forward a probability distribution G_n on $\mathcal{Z}_n$ to $\mathcal{M}_n$ using ϕ to get an identical distribution F_n on $\mathcal{M}_n$

We do this for settings which give a random graph with calculable sparsity and triangle density.

Calculation of sparsity and triangle density

Let $f : \mathcal{Z}_n \times \mathcal{Z}_n \to [0,1]$ be a positive-definite kernel which is Lipschitz continuous in each variable, with eigenvalues λ_k with respect to G_n, and Mercer feature map ϕ.

Lemma

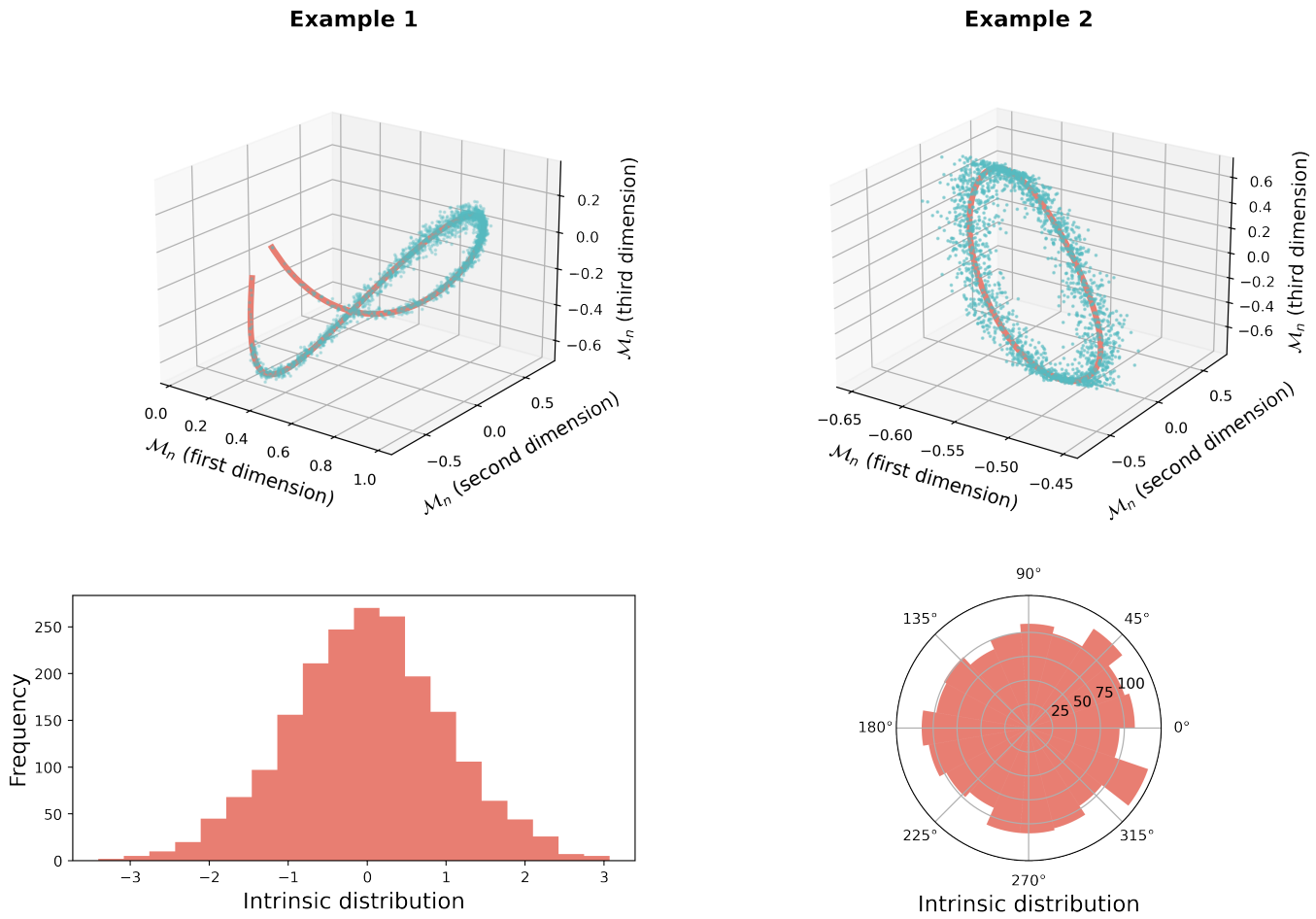
$$\iiint f(x,y)f(y,z)f(z,x)G_n(x)G_n(y)G_n(z) = \sum_{k=1}^{\infty} \lambda_k^3.$$

For the random dot product graph with $F_n = \phi(G_n)$,

$$\rho_n = \iint f(x,y)G_n(x)G_n(y), \qquad \text{and} \qquad \Delta_n = \Theta\left(n^2\right) \sum_{k=1}^{\infty} \lambda_k^3.$$

Results

In explicit examples, we can achieve $\rho_n = O(1/n)$ (constant expected degree) and $\Delta_n = \Omega(1)$ with e.g. normal or uniform on-manifold distributions



Triangles and community structure

It is known that certain forms of community structure do not result in triangles, i.e.

$$\text{community structure} \not\Longrightarrow \text{high triangle density.}$$

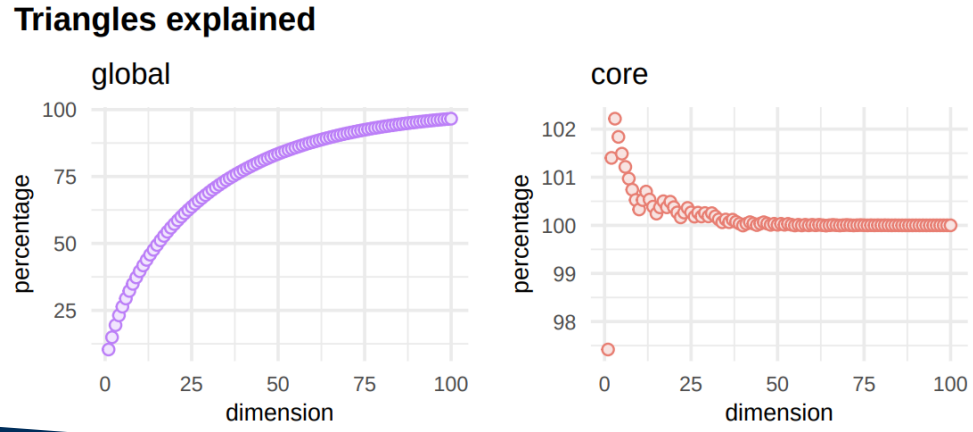
There is a perception that the reverse implication is true. Our constructions achieve high triangle density when no sensible notion of community structure is present, i.e.

$$\text{high triangle density} \not\Longrightarrow \text{community structure.}$$

Global vs local embedding

Representing a sparse graph, $n\rho_n = o(\log n)$, on a low-dimensional manifold leaves large parts of the manifold uncovered. Therefore, fully identifying the manifold is, in general, impossible.

By 'zooming in' on a small neighbourhood of the graph, we can form an embedding that represents a flat approximation of the manifold over that region.



References I

Rubin-Delanchy, P. (2020). Manifold structure in graph embeddings. In Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., and Lin, H., editors, *Advances in Neural Information Processing Systems*, volume 33, pages 11687–11699. Curran Associates, Inc.

Rubin-Delanchy, P., Cape, J., Tang, M., and Priebe, C. E. (2017). A statistical interpretation of spectral embedding: the generalised random dot product graph.

Seshadhri, C., Sharma, A., Stolman, A., and Goel, A. (2020). The impossibility of low-rank representations for triangle-rich complex networks. *Proceedings of the National Academy of Sciences*, 117(11):5631–5637.

References II

Whiteley, N., Gray, A., and Rubin-Delanchy, P. (2021). Matrix factorisation and the interpretation of geodesic distance. In Beygelzimer, A., Dauphin, Y., Liang, P., and Vaughan, J. W., editors, *Advances in Neural Information Processing Systems*.