The Shape of Edge Differential Privacy

Siddharth Vishwanath and **Jonathan Hehir**

The Pennsylvania State University

1A Random Dot-Product Graphs

RDPGs encompass a large class of commonly used models

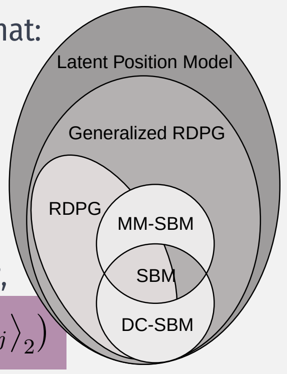
(Definition) Given $\mathbb{P}$ on $\mathbb{R}^d$ and $p+q=d$ such that:

- For $\mathbb{I}_{p,q} = \text{Diag}\left(\mathbf{1}_p^\top, -\mathbf{1}_q^\top\right)$, and
- For all $\boldsymbol{X}, \boldsymbol{Y} \sim \mathbb{P}$ such that $\boldsymbol{X} \perp \boldsymbol{Y}$

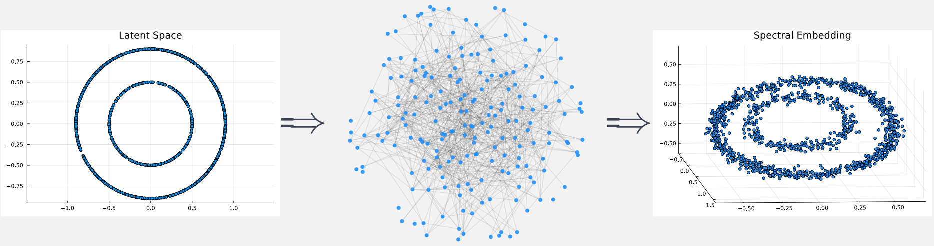
$$\langle \boldsymbol{X}, \mathbb{I}_{p,q}\boldsymbol{Y}\rangle_2 \in [0,1] \text{ a.s.}$$

Then $G \sim$ RDPG($\mathbb{P}$) if, for all $\{\boldsymbol{X}_1, \boldsymbol{X}_2 \dots \boldsymbol{X}_n\} \sim \mathbb{P}$,

$$\text{edge}(\boldsymbol{X}_i, \boldsymbol{X}_j) \,\big|\, \boldsymbol{X}_1 \dots \boldsymbol{X}_n \sim \text{Bernoulli}\left(\langle \boldsymbol{X}_i, \mathbb{I}_{p,q}\boldsymbol{X}_j\rangle_2\right)$$



Spectral embeddings of RDPGs recover topological information



2A Theoretical Results

For $\epsilon > 0$ and $G \sim$ RDPG($\mathbb{P}$), where supp($\mathbb{P}$) $= \mathcal{M} \subset \mathbb{R}^d$

(i) $\mathcal{A}_\epsilon(G) \sim$ RDPG($\mathbb{P}_\epsilon$) with supp($\mathbb{P}_\epsilon$)$=\mathcal{M}_\epsilon \subset \mathbb{R}^{d+1}$ s.t. $\mathcal{M}_\epsilon = \xi(\mathcal{M})$ and $\mathbb{P}_\epsilon = \xi_\sharp \mathbb{P}$ is the pushforward of $\mathbb{P}$ via

$$\xi : \boldsymbol{x} \mapsto \left(\sqrt{1-2\pi(\epsilon)}\right)\boldsymbol{x} \oplus \sqrt{\pi(\epsilon)}$$

(ii) $\mathcal{M}_\epsilon$ is diffeomorphic to $\mathcal{M}$, and diam$\mathcal{M}_\epsilon \downarrow$ as $\epsilon \downarrow$.

(iii) When $\epsilon = 0$, $\mathcal{M}_\epsilon = \{\boldsymbol{x}_0\}$ with $\|\boldsymbol{x}_0\| = \frac{1}{2}$ and $\boldsymbol{x}_0 \perp \mathcal{M}$

$$\text{RDPG}\left(\delta_{\boldsymbol{x}_0}\right) \sim \text{Erdős-Rényi}\left(\tfrac{1}{2}\right)$$

(iv)* When $\mathbb{X}_n = \Phi(G)$ and $\mathbb{Y}_n = \Phi(\mathcal{A}_\epsilon(G))$ denote the spectral embeddings of G and $\mathcal{A}_\epsilon(G)$, then as $n \to \infty$

$$W_\infty^{\text{SI}}(\mathbb{X}_n, \mathbb{Y}_n) \xrightarrow{p} 0$$

* Under some mild regularity assumptions

TL; DR

Given a graph $G \sim \mathcal{G}$, the ϵ-edge DP graph $\mathcal{A}_\epsilon(G)$ preserves topological structure for a large class of random graphs $\mathcal{G}$.

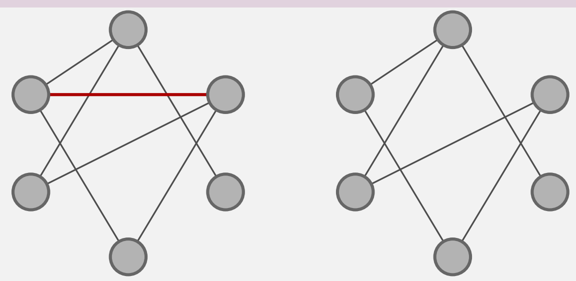
$$\begin{tikzcd}
\mathcal{M} \arrow[r, "\text{RDPG}(\mathbb{P})"] \arrow[d, leftrightarrow, "\simeq"'] & G \arrow[r, "\text{spectral embedding}"] \arrow[d, mapsto, "\text{edgeFlip}"] & \mathbb{X}_n \arrow[d, dash, "W_\infty^{\text{SI}}(\mathbb{X}_n, \mathbb{Y}_n) \xrightarrow{p} 0"] \\
\mathcal{M}_\epsilon & \mathcal{A}_\epsilon(G) \arrow[l, "\text{RDPG}(\mathbb{P}_\epsilon)"'] \arrow[r, "\text{spectral embedding}"'] & \mathbb{Y}_n
\end{tikzcd}$$

1B Differential Privacy via edgeFlip

Let $\mathcal{G}^n = \{(V,E) : |V| = n\}$ = Class of graphs with n vertices

(Definition) $\mathscr{M} : \mathcal{G}^n \to \mathcal{G}^n$ satisfies ϵ–edge DP if for all graphs $G_1 \overset{e}{\sim} G_2$ differing in a single edge, i.e. $E_1 \Delta E_2 = \{e\}$

$$\mathbb{P}\left(\mathscr{M}(G_1) \in S\right) \le e^\epsilon\, \mathbb{P}\left(\mathscr{M}(G_2) \in S\right) \quad \forall S \subseteq \mathcal{G}^n$$



(edgeFlip) For graph G, $\epsilon > 0$ and $\pi(\epsilon) := (1+e^\epsilon)^{-1} \in (0,1)$, edgeFlip is the mechanism $\mathcal{A}_\epsilon(G) : \mathcal{G}^n \to \mathcal{G}^n$ such that

$$\mathcal{A}_\epsilon\left(\mathsf{e}(i,j)\right) \big| \mathsf{e}(i,j) = \begin{cases} \mathsf{e}(i,j) & \text{w.p. } 1-\pi(\epsilon) \\ 1-\mathsf{e}(i,j) & \text{w.p. } \pi(\epsilon) \end{cases}$$

References

[1] F. Chazal and B. Michel. *An introduction to topological data analysis.* arXiv:1710.04019, 2017

[2] V. Karwa, et al. *Sharing social network data.*Journal of the Royal Statistical Society, 2017

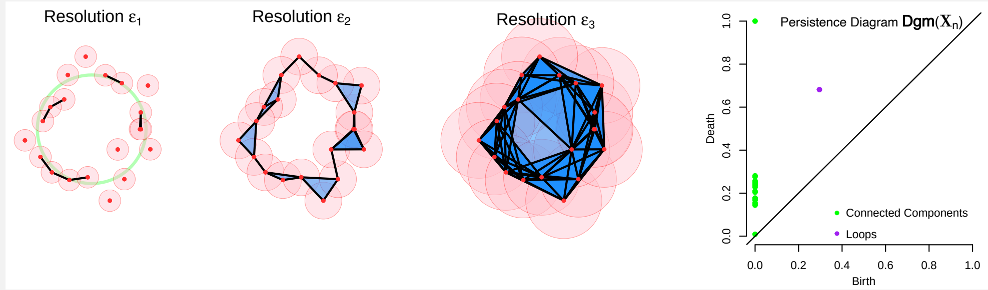
[3] A. Athreya, et al. *Statistical inference on random dot product graphs.*Journal of Machine Learning Research, 2017

[4] V. Solanki, et al. *Persistent homology of graph embeddings.* arXiv:1912.10238, 2019

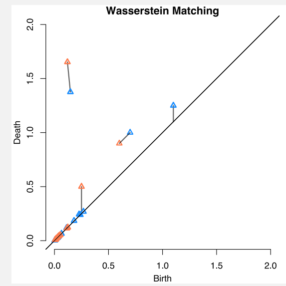
1C Measuring Shape using Topology

Topological Data Analysis has emerged as a propitious tool for uncovering low-dimensional structures underlying data

(Persistence Diagram) Given $\mathbb{X}_n = \{\boldsymbol{X}_1, \dots, \boldsymbol{X}_n\}$, the **multiscale** evolution of topological features is summarized in Dgm $(\mathbb{X}_n)$



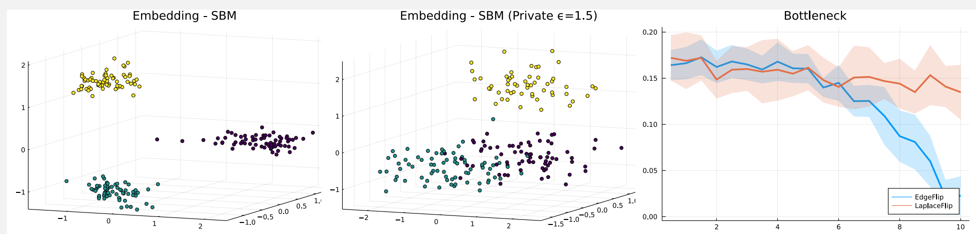
- Dgm $(\mathbb{X}_n)$ lives in a metric space $(\mathfrak{D}, W_\infty)$
- $W_\infty(\cdot,\cdot)$ is the Wasserstein metric for matchings
- The "shape distortion" between points $\mathbb{X}_n$ and $\mathbb{Y}_n$ can be quantified by $W_\infty(\text{Dgm}(\mathbb{X}_n), \text{Dgm}(\mathbb{Y}_n))$
- However, W_∞ is sensitive to the "units" of the underlying metric, e.g., distances in inches vs. cm



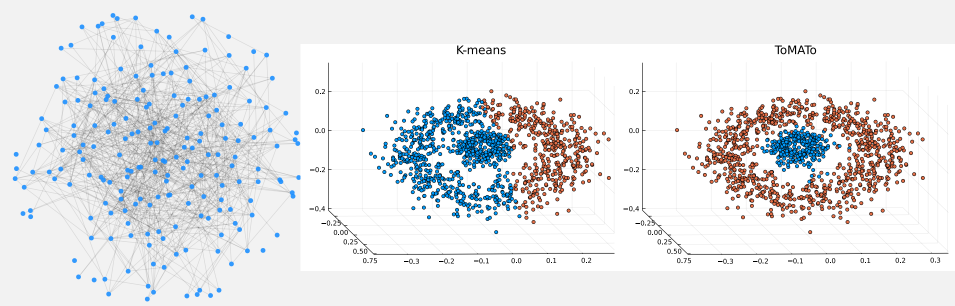
This can be overcome by considering the **shift-invariant distance**–$W_\infty^{\text{SI}}(\cdot,\cdot)$

$$W_\infty^{\text{SI}}(D_1, D_2) = \inf_{s \in \mathbb{R}} W_\infty(D_1 \oplus s, D_2)$$

2B Simulations & Experiments



1. The effect of edgeFlip brings the clusters closer together as per result 2A (ii)
2. edgeFlip outperforms LaplaceFlip, which is another ϵ-edge DP mechanism



3. Topology aware spectral clustering algorithms, which are more appropriate for the data and the privacy mechanism, lead to noticeably better results