Distance and Reconstruction in RGG: Breaking the $\Theta(r)$ error

Josep Díaz

Joint work with:

{V.Dani, T.Hayes, C.McDiarmid, D.Mitsche, C.Moore, G.Perarnau, X.Pérez}

Oct. 2021

Random Geometric Graphs

Given $n \in \mathbb{N}$, a set $V = \{v_i\}_{i=1}^n$, together with an embedding $\Psi : V \to \mathbb{R}^2$ into a convex subset of $\mathbb{R}^2$, for $\mathcal{S}_n = \left[0, \sqrt{n}\right]^2$ (a realization), and given a threshold distance $r > 0$, define a random geometric graph $G = G(\Psi, r)$, where $v_i, v_j \in V$ are adjacent iff $d_E(\Psi(v_i), \Psi(v_j)) \leq r$.

Two main models of distribution of the n vertices in $\mathcal{S}_n$:

- The *uniform distribution*, where the number of vertices in a subset of $\mathcal{S}_n$ of area A follows a Binomial distribution, and
- *Poisson distribution with intensity* $\lambda = 1$.

Asymptotically both models have the same properties.

If the realization Ψ is deterministic and r is rescaled to 1 then $G = G(\Psi, r)$ is said to be a Unit Disk Graph.

Alternative input for a RGG: adjacency matrix

The RGG G is given by its adjacency matrix A_G.

- *We don't have neither the realization Ψ, or the value of r.*
- But from A_G we do know the sets $V(G)$, $E(G)$, and for $u \in V(G)$ we know its degree $\delta(u)$.

$$A_G = \begin{array}{c} \\ v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \end{array} \begin{array}{c} \begin{array}{cccccc} v_1 & v_2 & v_3 & v_4 & v_5 & v_6 \end{array} \\ \begin{pmatrix} 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix} \end{array}$$

$E = \{(v_1, v_3), (v_1, v_5), \ldots\}$

$|E| = 7$

$\delta v_4 = 4; \mathcal{N}(v_2) = \{v_2, v_3, v_5, v_6\}$

Using transitive closure we can evaluate any graph distance between two vertices.

$d_E(v_i, v_j)$?

Estimating $\hat{r}$

Notice that i $X_V = \delta(v)$, $\mathbf{E}[X_v] = \frac{\pi r^2}{n}(n-1)$ and we exactly know the value $\delta(v)$, so we can get a sharp estimator $\hat{r}$ for r.

Formally Díaz, McDiarmid, Mitsche-2019.

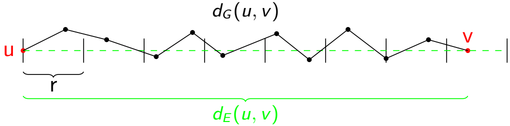
Thm. *Let $r = r(n) > 0$ be s.t. $1/\sqrt{n} < r < \sqrt{n}$ as $n \to \infty$. and $\rho = \sqrt{n}/r$. Let $\omega(n)$ a function tending to infinity with n arbitrarily slowly Then there is an $O(n^2)$ time algorithm to compute an estimator $\hat{r}$ s.t.*

$$|\hat{r} - r| < \omega \cdot (n^{-1/2} + \rho^{-3/2}) \ \ \textit{w.h.p.},$$

so that $\hat{r}/r \to 1$ in probability as $n \to \infty$.

Distances in RGG

Relate $d_E(u, v)$ and $d_G(u, v)$.



- Muthukrishnan, Panduragan (2005)
- Ellis, Martin, Yan (2007)
- Friedrich, Sauerwald, Martin, Yan (2007)
- Brandonjic, Elsässer, Sauerwald, Stauffer (2010)
- Merhabian, Wormald (2013)
- Díaz, Mitsche, Pérez, Perarnau (2016)
- Arias-Castro, Channarond, Pelletier, Verzelen (2017)
- Araya-Valdivia, De Castro (2019)
- Dani, Díaz, Hayes, Moore (2021)

Bounding $d_G(u,v)$ with $d_E(u,v)$

Thm. (D,M,P,P-16) Given $G \in \mathcal{G}(\Psi, r)$, $\exists c < 6 \cdot 10^6$ s.t. if $r \geq 224(\log n)^{3/4}$, w.h.p. for any $u, v \in V(G)$:

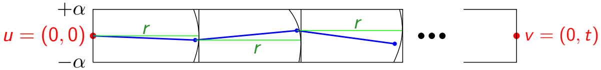
$$\left\lfloor \frac{d_E(u,v)}{r} \right\rfloor \leq d_G(u,v) \leq \left\lceil \frac{d_E(u,v)}{r} + 1 + c \cdot \max\left\{ \frac{n^{1/2}}{r^{7/3}}, \frac{n^{1/6}(\log n)^{2/3}}{r^{5/3}} \right\} \right\rceil .$$

If $r > n^{3/14}$, $\exists \epsilon(n) = o(1)$: $d_G(u,v) \leq \frac{d_E(u,v)}{r} + 1 + \epsilon(n)$,

If we want to bound $d_E(u,v)$:

If $r > n^{3/14}$, $d_G(u,v)r - \underbrace{(1+o(1))r}_{\text{error } \Theta(r)} \leq d_E(u,v) \leq d_G(u,v)r$.

Random greedy construction of path $u \to v$ in strip $d_E(u,v) \times 2\alpha$.

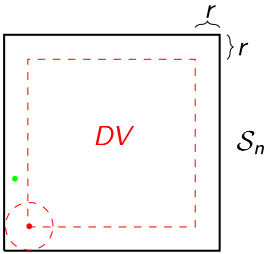


Breaking the $\Theta(r)$ error barrier: Deep vertices

Dani, Díaz, Hayes, Moore-21

The setting: Given an RGG G by A_G for u, v in $V(G)$ we want to get bounds for $d_E(u, v)$ conditioned on $d_G(u, v)$.

Given an RGG G in $\mathcal{S}_n$, define $v \in V(G)$ to be deep if there are $\geq 12r^2$ vertices at $d_G \leq 2$.



For $r > r_c$, w.h.p $v \in V(G)$ is a deep vertex iff $v \in DV$.

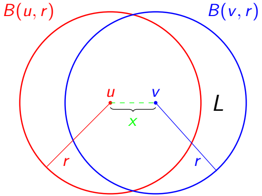
Breaking the $\Theta(r)$ error barrier: Short distances

Let G be given by A_G and $u, v \in V(G)$, s.t. v is deep and $d_G(u, v) \leq 2$, so $d_E(u, v) = x \leq 2r$.

Thm. Given A_G, if $d_G(u, v) \leq 2$ and v is deep, then w.h.p.

$$\left| d_E(u, v) - \tilde{d}(u, v) \right| \leq c\sqrt{\log n}.$$

For $0 < x \leq 2r$, for the lune (lense) $L = B(v, r) \setminus B(u, r)$, define $F(x) =$ the area $A(L)$ of L.
We want to approximate $A(L)$ by the number of points in L, and compute $F^{-1}(A(L))$ to approximate $x = d_E(u, v)$.



Breaking the $\Theta(r)$ error barrier: Long distances

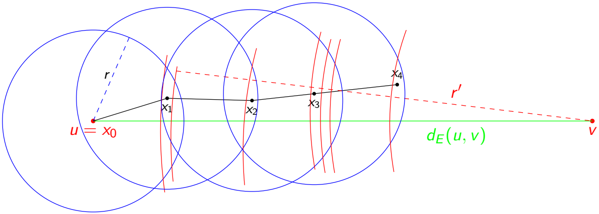
Thm. Given an RGG G, with $r > r_c$, for all $u, v \in V(G)$, w.h.p.

$$\lceil d_E(u,v)/r \rceil \leq d_G(u,v) \leq \lceil (d_E(u,v) + \kappa)/r \rceil,$$

where $\kappa/r = \Theta(d_E(u,v) \cdot r^{-7/3} + \log(n) \cdot r^{-4/3})$.

If $r = n^{\alpha}$ for $(0 < \alpha < 1/2)$ then $\kappa = O(n^{\beta})$ for $\beta = \frac{1}{2} - \frac{4}{3}\alpha$
$\alpha > 3/14$ *then* $\kappa = o(r)$.

Randomized greedy path $u \to v$



Breaking the $\Theta(r)$ error barrier: Hybrid distances

Cor. If $r = \Omega(n^{3/14})$ then $\kappa = o(r)$, and w.h.p.

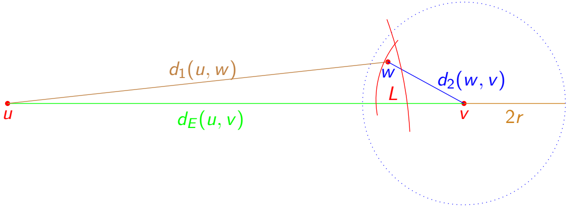
$$d_G(u,v) \cdot r - (r + \kappa) \le d_E(u,v) \le d_G(u,v) \cdot r.$$

If we can find a w s.t. $d_E(u,w)$ is near a multiple of r, say tr, the error could be diminished: for $r > n^{3/14}$, $rd_G(u,v)$ is a good estimator for $d_E(u,v)$:

Let $u, w \in G$, for $r = n^\alpha$, if $\exists t \in \mathbb{N}$, and a $\delta > 0$ s.t. $tr - (\kappa + \delta) < d_E(u,w) < tr - \kappa$, then

$$d_G(u,w)r - \underbrace{(\kappa + \delta)}_{\text{error}} \le d_E(u,w) \le d_G(u,w)r.$$

Breaking the $\Theta(r)$ error barrier: Hybrid distances



Thm. Given A_G, let $r = n^{\alpha}$ for $0 < \alpha < 1/2$. For all pairs $u, v \in V(G)$, with v deep define $\hat{d} = \min_{w \mid d_G(w,v) \leq 2}(d_1(u,w) + d_2(w,v))$. Then w.h.p.

$$\hat{d}(u,v) - \underbrace{\hat{\epsilon}(u,v)}_{\text{error}} \leq d_E(u,v) \leq \hat{d}(u,v),$$

where

$$\hat{\epsilon}(u,v) \leq \begin{cases} n^{\frac{1}{2} - \frac{4}{3}\alpha} & \alpha < 3/8, \\ \sqrt{\log n} & 3/8 \leq \alpha < 1/2. \end{cases}$$

Therefore, for $r = n^{\alpha}$, $3/14 < \alpha < 1/2$, the error is $o(r)$.

The reconstruction problem on 2D: $\mathcal{S}_n$

Given as input the adjacency matrix A_G of RGG G on $\mathcal{S}_n$, the goal is finding the realization $G(\Psi, r)$.
(i.e. Finding an embedding $\Phi : V \to \mathcal{S}_n$, which recovers G.)

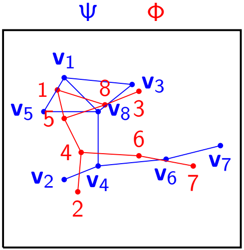
The reconstruction problem for deterministic UDG, is NP-hard Breu, Kirkpatrick, 1998.

Therefore, we only can aim to find an embedding $\Phi : V \to \mathcal{S}_n$ that yields a "good approximation" for the hidden (latent) Ψ

Lots of work done on the reconstruction and related problems, for different classes of graphs and using different techniques, on constant smooth.

Finding an RGG from its adjacency matrix

$$A_G = \begin{array}{c c} & \begin{array}{cccccccc} v_1 & v_2 & v_3 & v_4 & v_5 & v_6 & v_7 & v_8 \end{array} \\ \begin{array}{c} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ v_7 \\ v_8 \end{array} & \begin{pmatrix} 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \end{pmatrix} \end{array}$$



Displacement of Φ w.r.t. Ψ

Given A_G of a geometric graph G, our goal is to find an embedding Φ which is close to the hidden Ψ.

Given $\Phi, \Psi : V \to \mathcal{S}_n$, the sup distance is defined by

$$d_{\max}(\Phi, \Psi) = \max_{v \in V} d_E(\Phi(v), \Psi(v)).$$

As there are 8 symmetries σ of the square $\mathcal{S}_n$, define the symmetry-adjusted displacement d^* by

$$d^*(\Phi, \Psi) = \min_{\sigma} d_{\max}(\sigma \circ \Phi, \Psi).$$

Displacement is the most general measure of "closeness", when dealing with random graphs, but it is not the only one.

A solution to the reconstruction problem

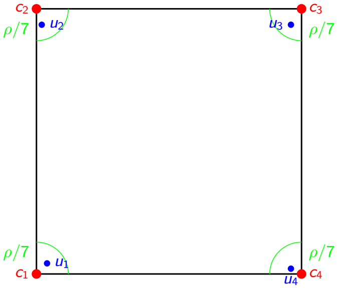
Díaz, McDiarmid, Mitsche (2019)

Given $G \in \mathcal{G}(\Psi, r)$ by its adjacency matrix A_G, we want to find an embedding Φ to the hidden Ψ, s.t. w.h.p. Φ approximates Ψ with minimal displacement, for the largest possible range of r.

Thm Given A_G for a hidden $G \in \mathcal{G}(\Psi, r)$, such that the range of the radius should be $n^{3/14} < r < \sqrt{n}$, fix $\varepsilon > 0$ be a small constant, there is an algorithm which in $O(n^2)$-time outputs a Φ such that w.h.p. $d^*(\Psi, \Phi) \leq (1+\varepsilon)r$.

Sketch of the proof

Using the vertex degrees, identify 4 vertices $C = \{u_1, u_2, u_3, u_4\}$ that w.h.p. are close to the 4 corners of $\mathcal{S}_n$

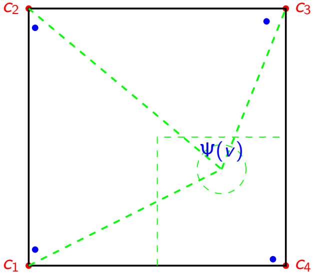


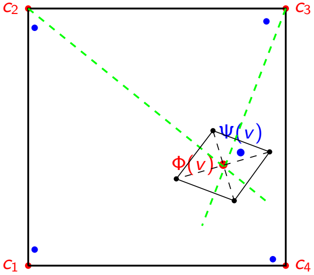
Algorithm

1. Pick u_1 as the vertex of min degree & place it in a corner. Mark it and all its neighbors.
2. Iteratively on the set of unmarked min-degree vertices, find the set C' with min-degree vertices.
3. Choose in C' the farthest vertex from u_1 and call it u_2, place it in opposite corner.
4. Place the remaining 2 vertices in the 2 remaining corners.

Sketch of the proof

Let $\mathcal{E}_1$ be the event in which the 4 vertices in C are placed near the corners in $\mathcal{S}_n$.
Conditioning on $\mathcal{E}_1$, for any $v \in V \backslash C$, we approximate $d_E(\Psi(v), \Psi(u_i))$ by using the $d_G(v, u_i)$ and then we place our estimate $\Phi(v)$ for $\Psi(v)$ at the intersection of the annuli centered on the 2 further away corners $\{u_i, u_j\}$.





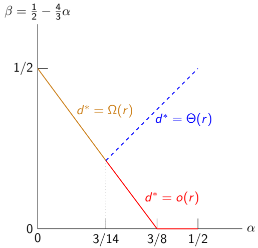
A better distortion for recovering $G \in \mathcal{S}_n$

(Dani, Díaz, Hayes, Moore (2021)

Thm. Given A_G for $r = n^{\alpha}$ $(0 < \alpha < 1/2)$, there is an $O(n^{2.373}\log(n))$ algorithm that w.h.p. reconstructs G, modulo the set of symmetries of $\mathcal{S}_n$ with $d^* = \Theta(\hat{\epsilon})$, i.e.

$$d^* = C \begin{cases} n^{\frac{1}{2}-\frac{4}{3}\alpha} & \text{if } \alpha < 3/8, \\ \sqrt{\log n} & \text{if } 3/8 \leq \alpha < 1/2. \end{cases}$$

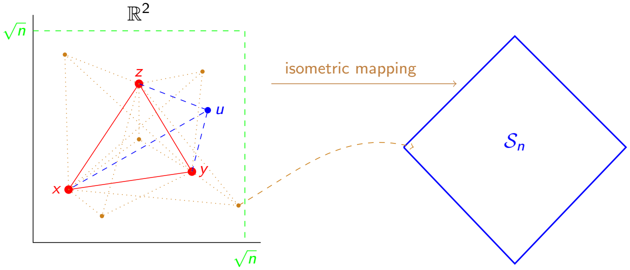
Notice that once we reconstruct the position of all vertices, we can get a good estimate on $d_E(u,v)$, $\forall u, v \in V$.



If $r = n^{\alpha}$ values of distortion d^* obtained by D,D,H,M-21 (red and blue) over D,McD,M-19 (dashed green) for the reconstruction problem.

Sketch of the Algorithm

- Using Seidel's APSP compute $d_G(u, v)$, $\forall u, v \in V$.
- In $\mathbb{R}^2$, choose deep x, y, z that form an acute triangle, with minimal length $\ell = \Omega(\sqrt{n})$. Estimate $\hat{d}(x, y), \hat{d}(y, z), \hat{d}(x, z)$, with error $\hat{\epsilon}$.
- For all other $u \in V \backslash \{x, y, z\}$, estimate their relative position with recpect x, y, z, with error $O(\hat{\epsilon})$.
- Do an isometric embedding from this graph into $\mathcal{S}_n$.



Complexity of the previous algorithm

The complexity of the algorithm is dominated by the computations of All Pairs Shortest Path using Seidel's randomized algorithm, which is $O(n^{\omega} \log n)$, where $\omega \sim 2.373$.

It should be possible to lower the complexity to $O(n^2)$, by avoiding using Seidel's APSP algorithm.

Reconstruction of RGG on $\mathbb{S}^2$

Scatter u.a.r. n points on surface of $\mathbb{S}^2$ in $\mathbb{R}^3$, according to a Poisson with $\lambda = 1$. Let $R = \sqrt{n/4\pi}$, so area $S_2 = n$.
For a given r, two points u, v on $\mathbb{S}^2$ are connected if $g(u, v) \leq r$, where $g(,)$ denotes the min geodesic distance.

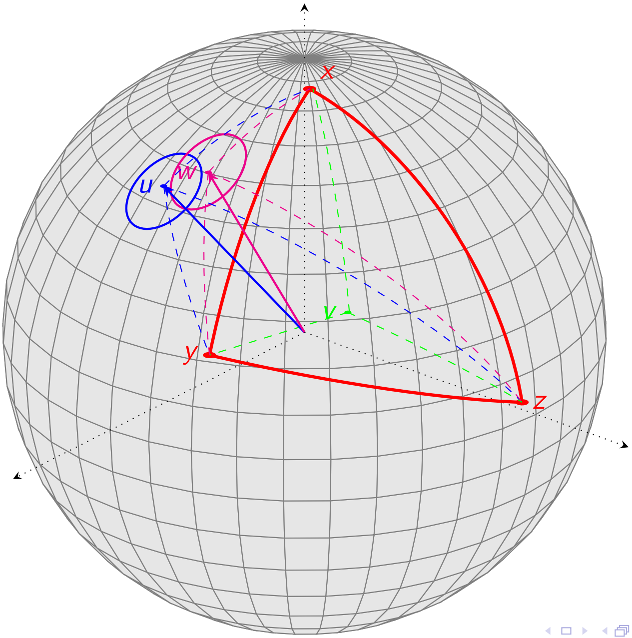
For early definitions of RGG on $\mathbb{S}^2$ see for ex. Bubeck, Ding, Eldan, Rácz,2016

Thm. Let $r = n^\alpha$ for $0 < \alpha < 1/2$. There is an $O(n^{3.37}\log(n))$ algorithm that w.h.p. reconstructs the vertex positions of a RGG, modulo symmetries of S_2, with

$$d^* = C \begin{cases} n^{\frac{1}{2}-\frac{4}{3}\alpha} & \text{if } \alpha < 3/8, \\ \sqrt{\log n} & \text{if } 3/8 \leq \alpha < 1/2. \end{cases}$$

Conjecture: Similar techniques can work on most d-dimensional curved manifolds, for fixed d.

Reconstruction of RGG on $\mathbb{S}^2$



Reconstruction of RGG in the d-dimensional hypercube

Consider the d-dimensional hypercube $H = [0, n^{1/d}]^d$, for d fixed:

- Define and compute the volume of the d-dim lens and lunes,
- define waypoints as the vertices in a d-simplex.

Thm. Let G be a RGG in H, given by A_G, let $r = n^\alpha$, for $0 < \alpha < 1/d$. There is an algorithm with running time $O(n^{2.37} \log n)$, that w.h.p. reconstructs G, modulo symmetries of the hypercube, with distortion

$$d^* \leq C_d \begin{cases} n^{\frac{1}{d} - \frac{2d}{d+1}\alpha} & \alpha < \frac{d+1}{2d^2}, \\ \sqrt{\log n} & \frac{d+1}{2d^2} \leq \alpha < \frac{1}{d}. \end{cases}$$

Related models (1)

On the Estimation of Latent Distances Using Graph Distances.
E.Arias-Castro, A.Channarond, Pelletier, N.Verzelen (2018)

Given $V = \{x_1, \ldots, x_n\}$ latent points u.a.r. on $\mathbb{S}^{d-1}$ (d fixed) define a random graph G on V by the adjacency matrix A, where the probability of having $a_{ij} = 1$ depends of a link function $\Phi(d_E(x_i, x_j))$, where $\Phi : [0, \infty) \to [0, 1]$.

We want to approximate distances, reconstructing latent points, etc..

For RGG, given an r, the link function is defined by

$$\Phi(d_E) = \mathbb{1}\{d_E \leq r\}.$$

The error of distances, and therefore recovering points is $\Theta(r)$.

Related models (2)

Latent Distance Estimation for Random Geometric Graphs.
E.Araya, Y. De Castro (2019)

- Having $|V| = n$ latent points on $\mathbb{S}^{d-1}$ they want to approximate $d_E(x_i, x_j)$ for any two points.
- Also uses the link function ϕ is given by graphon function on $\mathbb{S}^{d-1}$.
- They sparsify the graph by giving every node a small probability to other points and a great probability to be connected to near nodes. The resulting graph is not a RGG.
- They use harmonic analysis on the $\mathbb{S}^{d-1}$ to have a nice characterization on the graphon spectrum.
- Their main result is an $O(\log n)$ approximation for pairwise Euclidean distance between points.

As a byproduct, their method can also estimate the dimension d of the latent space.

Thank you for your attention