Proceedings of Machine Learning Research vol 247:1–3, 2024 37th Annual Conference on Learning Theory

# Reconstructing the Geometry of Random Geometric Graphs (Extended Abstract)

**Han Huang** HHUANG@MISSOURI.EDU
*Department of Mathematics, University of Missouri, Columbia, MO 65203*

**Pakawut Jiradilok** PAKAWUT@MIT.EDU
*Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139*

**Elchanan Mossel** ELMOS@MIT.EDU
*Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139*

**Editors:** Shipra Agrawal and Aaron Roth

**Keywords:** Gromov–Hausdorff distance, manifold reconstruction, manifold learning, metric measure space, random geometric graph, Riemannian geometry.

The manifold assumption in machine learning is a popular assumption postulating that many models of data arise from distributions over manifolds, see e.g. Duchemin and De Castro (2023); Araya and De Castro (2019); Fefferman et al. (2021, 2020b, 2023); Atamanchuk et al. (2023) among many others. A major problem studied in this area is the inference problem of estimating an unknown manifold given data sampled from the manifold.

We are interested in a more difficult problem which arises in the context of random geometric graphs (see e.g. Penrose (2003); Duchemin and De Castro (2023) for surveys on the subject). Random geometric graphs are random graph models constructed by first sampling points from a metric space and then connecting each pair of sampled points with a probability that depends on their distance, independently among pairs.

In this setting, the sample points $X_1, X_2, \ldots, X_n$ are drawn from a manifold $M$ embedded in an ambient Euclidean space according to some probability measure $\mu$. The graph $G$ is constructed where each vertex corresponds to a latent sample point. Any two vertices $i$ and $j$ are connected by an edge with a probability determined by their Euclidean distance; specifically, the further apart $X_i$ and $X_j$ are, the less likely they are to be connected by an edge. This probability is represented by $\mathrm{p}(\|X_i - X_j\|)$, where $\mathrm{p} : [0, \infty) \to [0, 1]$ is a strictly decreasing function.

Rather than observing the data points $X_1, \ldots, X_n$, we ask if it is possible to infer the manifold $M$ by only observing the random geometric graph $G$.

In this work we demonstrate that under some mild assumptions on the manifold $M$, the probability measure $\mu$, and the function $\mathrm{p}$, with probability $1 - o_n(1)$, we can estimate both the Euclidean distance and the geodesic distance for every pair of latent points $(X_i, X_j)$ with an error of order $n^{-c/\dim(M)}$ for some constant $c > 0$. Furthermore, we can construct a discrete metric measure space $(\tilde{M}, \nu)$ that approximates $(M, \mu)$ in the Gromov–Hausdorff distance sense. All these results can be achieved by algorithms that have polynomial running time in $n$. If we combine our result with the work of Fefferman et al. (2020a, 2021), it is possible to recover a manifold $\hat{M}$ that is close to $M$.

Our work complements a large body of work on manifold learning, where the goal is to recover a manifold from sampled points sampled in the manifold along with their (approximate) distances.

. Extended abstract. Full version appears as [arXiv:2402.09591,v2]

## Acknowledgments

Han Huang and Pakawut Jiradilok were supported by Elchanan Mossel's Vannevar Bush Faculty Fellowship ONR-N00014-20-1-2826 and by Elchanan Mossel's Simons Investigator award (622132). Elchanan Mossel was partially supported by Bush Faculty Fellowship ONR-N00014-20-1-2826, Simons Investigator award (622132), ARO MURI W911NF1910217, NSF DMS-2031883, and NSF award CCF 1918421. We would like to thank Anna Brandenberger, Tristan Collins, Dan Mikulincer, Ankur Moitra, Greg Parker, and Konstantin Tikhomirov for helpful discussions.

## References

Ernesto Araya and Yohann De Castro. Latent distance estimation for random geometric graphs. *Advances in Neural Information Processing Systems*, 32, 2019.

Caelan Atamanchuk, Luc Devroye, and Gábor Lugosi. A note on estimating the dimension from a random geometric graph. *arXiv:2311.13059*, 2023.

Sébastien Bubeck, Jian Ding, Ronen Eldan, and Miklós Z Rácz. Testing for high-dimensional geometry in random graphs. *Random Structures & Algorithms*, 49(3):503–532, 2016.

Quentin Duchemin and Yohann De Castro. Random geometric graph: some recent developments and perspectives. In *High dimensional probability IX—the ethereal volume*, volume 80 of *Progr. Probab.*, pages 347–392. Birkhäuser/Springer, Cham, 2023.

Ronen Eldan, Dan Mikulincer, and Hester Pieters. Community detection and percolation of information in a geometric setting. *Combin. Probab. Comput.*, 31(6):1048–1069, 2022. ISSN 0963-5483,1469-2163. doi: 10.1017/s0963548322000098. URL https://doi.org/10.1017/s0963548322000098.

Charles Fefferman, Sergei Ivanov, Yaroslav Kurylev, Matti Lassas, and Hariharan Narayanan. Reconstruction and interpolation of manifolds. I: The geometric Whitney problem. *Found. Comput. Math.*, 20(5):1035–1133, 2020a. ISSN 1615-3375,1615-3383. doi: 10.1007/s10208-019-09439-7. URL https://doi.org/10.1007/s10208-019-09439-7.

Charles Fefferman, Sergei Ivanov, Matti Lassas, and Hariharan Narayanan. Reconstruction of a Riemannian manifold from noisy intrinsic distances. *SIAM J. Math. Data Sci.*, 2(3):770–808, 2020b. ISSN 2577-0187. doi: 10.1137/19M126829X. URL https://doi.org/10.1137/19M126829X.

Charles Fefferman, Sergei Ivanov, Matti Lassas, Jinpeng Lu, and Hariharan Narayanan. Reconstruction and interpolation of manifolds II: Inverse problems for Riemannian manifolds with partial distance data. *arXiv:2111.14528*, 2021.

Charles Fefferman, Sergei Ivanov, Matti Lassas, and Hariharan Narayanan. Fitting a manifold to data in the presence of large noise. *arXiv:2312.10598*, 2023.

Mikhael Gromov, Misha Katz, Pierre Pansu, and Stephen Semmes. *Metric structures for Riemannian and non-Riemannian spaces*, volume 152. Springer, 1999.

Adel Javanmard and Andrea Montanari. Localization from incomplete noisy distance measurements. *Found. Comput. Math.*, 13(3):297–345, 2013. ISSN 1615-3375,1615-3383. doi: 10.1007/s10208-012-9129-5. URL https://doi.org/10.1007/s10208-012-9129-5.

Shuangping Li and Tselil Schramm. Spectral clustering in the gaussian mixture block model. *arXiv:2305.00979*, 2023.

Siqi Liu, Sidhanth Mohanty, Tselil Schramm, and Elizabeth Yang. Testing thresholds for high-dimensional sparse random geometric graphs. In *STOC '22—Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, pages 672–677. ACM, New York, 2022.

Sewoong Oh, Andrea Montanari, and Amin Karbasi. Sensor network localization from local connectivity: Performance analysis for the mds-map algorithm. In *2010 IEEE Information Theory Workshop on Information Theory (ITW 2010, Cairo)*, pages 1–5. IEEE, 2010.

Mathew Penrose. *Random geometric graphs*, volume 5 of *Oxford Studies in Probability*. Oxford University Press, Oxford, 2003. ISBN 0-19-850626-0. doi: 10.1093/acprof:oso/9780198506263.001.0001. URL https://doi.org/10.1093/acprof:oso/9780198506263.001.0001.

Takashi Shioya. *Metric measure geometry*, volume 25 of *IRMA Lectures in Mathematics and Theoretical Physics*. EMS Publishing House, Zürich, 2016. ISBN 978-3-03719-158-3. doi: 10.4171/158. URL https://doi.org/10.4171/158. Gromov's theory of convergence and concentration of metrics and measures.

Daniel L. Sussman, Minh Tang, and Carey E. Priebe. Universally consistent latent position estimation and vertex classification for random dot product graphs. *arXiv:1207.6745*, 2012.