36-781 Advanced Statistical Network Models
November 17 Lecture Notes

Abulhair Saparov
asaparov@andrew.cmu.edu

Continuous latent space models

Each vertex in a graph in this model is associated with a location $X_i \in \mathcal{M}$ where $\mathcal{M}$ is a well-behaved metric space with metric ρ. Some examples of metric spaces used in continuous latent space models are: (1) Euclidean space $\mathbb{R}^k$, (2) hyperbolic spaces $\mathbb{H}_k$, (3) other manifolds. In continuous latent space models, the probability of an edge between vertex i and j is given by

$$p(A_{ij} = 1) = w_n(\rho(X_i, X_j)),$$

where n is the number of vertices in the graph [5]. This dependence on n enables the model to vary with respect to n, but we will omit it henceforth for conciseness. Typically, latent space models also assume that w generally decreases for increasing ρ. In this model, the edges are independent conditioned on the latent vertex locations X_i. In many continuous latent space models, the latent vertex locations X_i are assumed to be distributed according to some distribution f on $\mathcal{M}$.

Hoff, Raftery, and Handcock [1] presented a continuous latent space model where $\mathcal{M}$ is a two or three-dimensional Euclidean space ($\mathbb{R}^2$ or $\mathbb{R}^3$), ρ is the standard Euclidean distance metric, and

$$w(\rho(X_i, X_j)) = \text{logit}^{-1}(\beta_0 - \rho(X_i, X_j)),$$

where $\text{logit}(p) = \log \frac{p}{1-p}$ and so $\text{logit}^{-1}(q) = (1 + e^{-q})^{-1}$. They also have the latent vertex locations distributed according to a standard Gaussian $f = \mathcal{N}(0, I)$.

Krioukov et al. [2, 4, 3] explored a different model where $\mathcal{M} = \mathbb{H}_2$ is the two-dimensional hyperbolic plane, ρ is the hyperbolic distance metric, and

$$w(\rho(X_i, X_j)) = \begin{cases} 1 & \text{if } \rho(X_i, X_j) < c, \\ 0 & \text{if } \rho(X_i, X_j) \geq c. \end{cases}$$

The latent vertex positions are distributed uniformly on a disc around the origin (uniform with respect to the Euclidean metric). The link function w was selected to simplify the mathematics in their derivations, but more sophisticated link functions can be used. The hyperbolic geometry naturally produces a "core-periphery" structure, since vertices near the "edge" of the hyperbolic space will be distant from each other, in terms of hyperbolic distance, and so they will tend not to connect with other points in the periphery. Points near the center of the disc will be part of the "core", and since they will tend to be closer to each other in terms of hyperbolic distance, they will tend to connect with other points in the core. This core-periphery structure is frequently encountered in routing networks on the internet. See figure 1 for a graphical depiction of this space.

Isometry

Def 1. *An* isometry *is an invertible map $\phi : \mathcal{M} \mapsto \mathcal{M}$ that preserves distances. That is, for any $x, y \in \mathcal{M}$, $\rho(x, y) = \rho(\phi(x), \phi(y))$.*

Isometries form a group since: (i) the identity map is an isometry, (ii) every isometry has an inverse which is also an isometry, and (iii) the composition of any two isometries is also an isometry (i.e., the group is closed

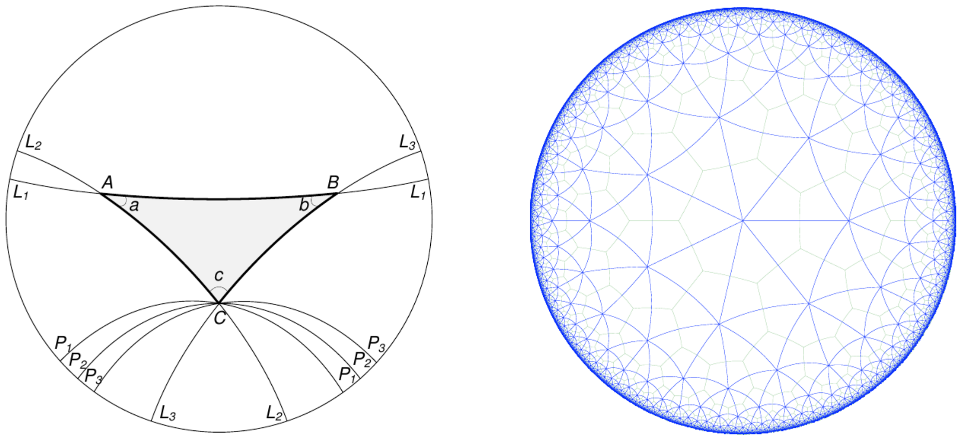


Figure 1: **(left)** A triangle drawn in the two-dimensional hyperbolic space $\mathbb{H}_2$ with three intersecting lines L_1, L_2, L_3. Three additional lines P_1, P_2, P_3 intersect at the vertex C. **(right)** A tesselation of $\mathbb{H}_2$ with equilateral triangles is shown in blue, and the dual tesselation with regular heptagons is shown in green. [3]

under composition). In this group, the identity is the identity map $\phi(x) = x$, and the group operation is composition. For any isometries ϕ_1, ϕ_2, ϕ_3, we have $(\phi_1 \circ \phi_2) \circ \phi_3 = \phi_1 \circ (\phi_2 \circ \phi_3)$ and so the composition operation is associative. But it is not necessarily commutative, and so the group is not Abelian.

For Euclidean geometry, isometries can be represented as a composition of three elementary transformations: translations, rotations, and reflections. The group of Euclidean isometries has two subgroups: (1) the set of isometries composed by translations and rotations, and (2) the set of isometries composed by translations, rotations, and a single reflection. Thus, the isometry group has two connected components.

Two sets of points $\boldsymbol{X} = \{X_1, \ldots, X_n\}$ and $\boldsymbol{Y} = \{Y_1, \ldots, Y_n\}$ are called *isometric* if there is an isometry ϕ such that $X_i = \phi(Y_i)$ for all i. So if $\boldsymbol{X}$ and $\boldsymbol{Y}$ are isometric, then

$$\rho(X_i, X_j) = \rho(Y_i, Y_j) \text{ for all } i, j.$$

This implies that for any i, j,

$$p(A_{ij} = 1 | \boldsymbol{X}) = w(\rho(X_i, X_j)) = w(\rho(Y_i, Y_j)) = p(A_{ij} | \boldsymbol{Y}).$$

Therefore, the distinction between $\boldsymbol{X}$ and $\boldsymbol{Y}$ is *unidentified* by the data.

Def 2. *For a set of points $\boldsymbol{X} = \{X_1, \ldots, X_n\}$, we define the equivalence class $[\boldsymbol{X}]$ as the set of all point sets $\boldsymbol{Y} \in \mathcal{M}^n$ such that for some isometry ϕ, for all $i = 1, \ldots, n$, $X_i = \phi(Y_i)$.*

The best we can hope to recover from data is $[\boldsymbol{X}]$, not $\boldsymbol{X}$, regardless of the size of n.

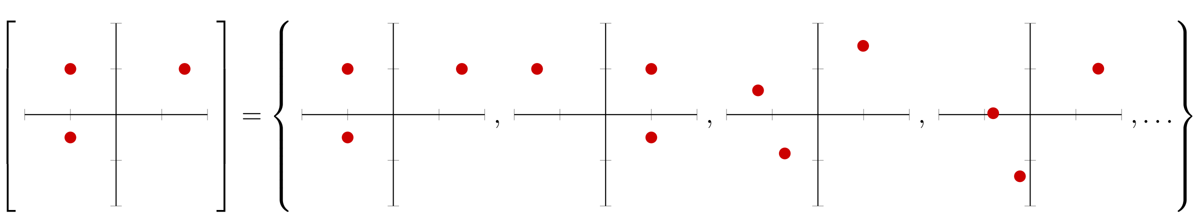


Figure 2: An example of an equivalence class of a set of three points under isometric transformations in $\mathbb{R}^2$.

The log-likelihood of a graph with adjacency matrix A under a continuous latent space model is

$$p(A|\boldsymbol{X}) = \prod_{i,j} p(A_{ij} = 1|X_i, X_j)^{A_{ij}} p(A_{ij} = 0|X_i, X_j)^{1-A_{ij}},$$

$$l(\boldsymbol{X}) \triangleq \log p(A|\boldsymbol{X}) = \sum_{i,j} A_{ij} \log w(\rho(X_i, X_j)) + (1 - A_{ij}) \log(1 - w(\rho(X_i, X_j))),$$

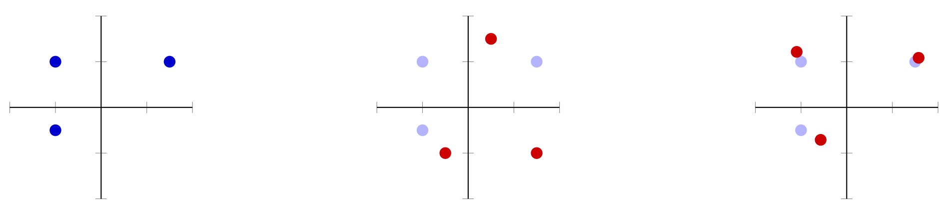
$$= \sum_{i,j} \log(1 - w(\rho(X_i, X_j))) + A_{ij} \log \frac{w(\rho(X_i, X_j))}{1 - w(\rho(X_i, X_j))},$$

$$= \sum_{i,j} \log(1 - w(\rho(X_i, X_j))) + A_{ij} \text{logit}(w(\rho(X_i, X_j))).$$

We observe that the likelihood is unchanged if $\boldsymbol{X}$ is transformed by any isometry ϕ. As such, latent vertex positions are unidentified by data. We also observe that it is often times simpler to write parametric models in terms of logit transforms.

Parametric inference with MCMC

With some prior f on the latent coordinates, there are a number ways of performing inference: (i) Pick an initial set of coordinates for the latent vertex locations X_i, (ii) the posterior density is proportional to $p(A|\boldsymbol{X}) \prod_{i=1}^n f(X_i)$, and (iii) use Markov chain Monte Carlo to sample from this posterior density. The posterior is not necessarily invariant under isometries since the prior is not. The proposal step of MCMC is often times modified to deal with isometry. This modification is to solve a *Procrustes problem* at each step: find the isometry ϕ such that the distance between $\boldsymbol{X}$ and $\phi(\boldsymbol{X})$ is minimized.



$\boldsymbol{X}$ before sampling step   newly sampled $\boldsymbol{X}$ solution to Procrustes problem

Nonparametric inference

To make nonparametric inference feasible, we need to make additional assumptions:

1. For any two point sets $\boldsymbol{X}$ and $\boldsymbol{Y}$, if $\rho(X_i, X_j) = \rho(Y_i, Y_j)$ for all i, j, then there exists an isometry ϕ such that $X_i = \phi(Y_i)$ for all i.

2. The set of isometries has a finite number of connected components, which we call B. In Euclidean space, for example, $B = 2$.

3. w is logit-bounded. That is, $|\text{logit}(w_n(\rho(X_i, X_j)))| \leq v_n$ and $n/v_n \to \infty$ as $n \to \infty$.

4. All derivatives of w exist and are finite.

Def 3. *The expectation of the log-likelihood is* $\bar{l}(\boldsymbol{X}) \triangleq \mathbb{E}[l(\boldsymbol{X})]$*, where the expectation is taken with respect to the distribution of graphs* A*.*

Under these assumptions, the set of maxima of $\bar{l}$ is exactly $[\boldsymbol{X}^*]$ where $\boldsymbol{X}^*$ are the true underlying coordinates. Furthermore, for each $\boldsymbol{X} \in \mathcal{M}^n$,

$$p(|l(\boldsymbol{X}) - \bar{l}(\boldsymbol{X})| > \epsilon) \leq 2 \exp\left\{-\frac{n(n-1)\epsilon^2}{v_n^2}\right\},$$

which follows from the bounded difference inequality. To obtain uniform convergence, we need to bound

$$p\left(\sup_{\boldsymbol{X}\in\mathcal{M}^n} |l(\boldsymbol{X}) - \bar{l}(\boldsymbol{X})| > \epsilon\right),$$

and so we need the effective size of the set of likelihood functions. It can be shown that the effective number of likelihood functions is exponential in n and B. But since the absolute error in the likelihood is bounded by the exponential of the square of n, and so the error in the likelihood will go to zero exponentially quickly.

References

[1] Peter D. Hoff, Adrian E. Raftery, and Mark S. Handcock. Latent space approaches to social network analysis. *Journal of the American Statistical Association*, 97(460):1090–1098, 2002.

[2] Dmitri V. Krioukov, Fragkiskos Papadopoulos, MariÃąn BoguÃśÃą, and Amin Vahdat. Efficient navigation in scale-free networks embedded in hyperbolic metric spaces. *CoRR*, abs/0805.1266, 2008.

[3] Dmitri V. Krioukov, Fragkiskos Papadopoulos, Maksim Kitsak, Amin Vahdat, and MariÃąn BoguÃśÃą. Hyperbolic geometry of complex networks. *CoRR*, abs/1006.5169, 2010.

[4] Dmitri V. Krioukov, Fragkiskos Papadopoulos, Amin Vahdat, and MariÃąn BoguÃśÃą. On curvature and temperature of complex networks. *CoRR*, abs/0903.2584, 2009.

[5] Cosma Rohilla Shalizi and Dena Asta. Consistency of maximum likelihood for continuous-space network models. Submitted, 2016.