Bayesian estimation of the generalised random dot product graph

Patrick Rubin-Delanchy & Francesco Sanna Passino

University of Bristol & Imperial College London

12th September, Warwick University

References

1. Patrick Rubin-Delanchy, Joshua Cape, Minh Tang, Carey E. Priebe. “A statistical interpretation of spectral embedding: the generalised random dot product graph” arxiv.org/abs/1709.05506 (JRSSB, to appear, 2022)
2. Rubin-Delanchy, Patrick. “Manifold structure in graph embeddings.” arXiv:2006.05168 (NeurIPS, 2020)
3. Nick Whiteley, Annie Gray, Patrick Rubin-Delanchy. “Matrix factorisation and the interpretation of geodesic distance” arXiv:2106.01260 (NeurIPS, 2021)
4. Fangzheng Xie, Yanxun Xu, “Optimal Bayesian Estimation for Random Dot Product Graphs” arxiv.org/abs/1904.12070 (Biometrika, 2020)
5. Fangzheng Xie, Yanxun Xu, “Efficient Estimation for Random Dot Product Graphs via a One-step Procedure” arxiv.org/abs/1910.04333 (JASA, 2021)
6. Dingbo Wu, Fangzheng Xie, “Statistical inference of random graphs with a surrogate likelihood function” arxiv.org/abs/2207.01702
7. Nick Whiteley, Annie Gray, Patrick Rubin-Delanchy, “Discovering latent topology and geometry in data: a law of large dimension” arxiv.org/abs/2208.11665

Graph spectral embedding

Consider an undirected graph on n nodes with symmetric adjacency matrix $\mathbf{A} \in \{0,1\}^{n\times n}$.

Definition (Spectral embedding)

The spectral embedding of $\mathbf{A}$ into D dimensions is defined as

$$\hat{\mathbf{X}} = [\hat{X}_1, \ldots, \hat{X}_n]^\top = \mathbf{U}|\mathbf{S}|^{1/2},$$

where $\mathbf{S} \in \mathbb{R}^{D\times D}$ is a diagonal matrix containing the D largest eigenvalues of $\mathbf{A}$ by magnitude, and $\mathbf{U} \in \mathbb{R}^{n\times D}$ is a matrix containing corresponding eigenvectors as columns.

Generalised random dot product graph

Let $\mathbf{I}_{p,q} = \text{diag}(1, \ldots, 1, -1, \ldots, -1)$, with p ones followed by q minus ones on the diagonal, where $p + q = D$.

Definition (GRDPG)

Let $X_1, \ldots, X_n \in \mathcal{X}$, for a valid set $\mathcal{X} \subset \mathbb{R}^D$. Then a generalised random dot product graph has a symmetric adjacency matrix satisfying

$$\mathbf{A}_{ij} \overset{ind}{\sim} \text{Bernoulli}\left\{X_i^\top \mathbf{I}_{p,q} X_j\right\},$$

for $i < j$, where $p + q = D$.

The random dot product graph (RDPG) corresponds to special case where $q = 0$.

Convergence of $\hat{X}_i$ to X_i

Theorem

For some indefinite orthogonal matrix $\mathbf{Q}$

1. *Asymptotically, conditional on X_i:*

$$\sqrt{n}(\mathbf{Q}\hat{X}_i - X_i) \overset{ind}{\sim} Normal\{\mathbf{0}, \mathbf{\Gamma}(X_i)\}$$

2. *The maximum error*

$$\max_{i \in \{1,\dots,n\}} \|\mathbf{Q}\hat{X}_i - X_i\| \to 0, \quad \textit{in probability}$$

Indefinite orthogonal group: $\{\mathbf{M} : \mathbf{M}^\top \mathbf{I}_{p,q} \mathbf{M} = \mathbf{I}_{p,q}\}$

Special case: degree-corrected stochastic block model

Definition (Degree-corrected stochastic block model)

- Let $Z_1, \ldots, Z_n \in \{1, \ldots, K\}$ denote the communities of the nodes
- Let $\mathbf{B} \in [0,1]^{K \times K}, \mathbf{B} = \mathbf{B}^\top$ denote the inter-community link probabilities.
- Let $w_1, \ldots, w_n \in [0,1]$ denote the node "popularities"

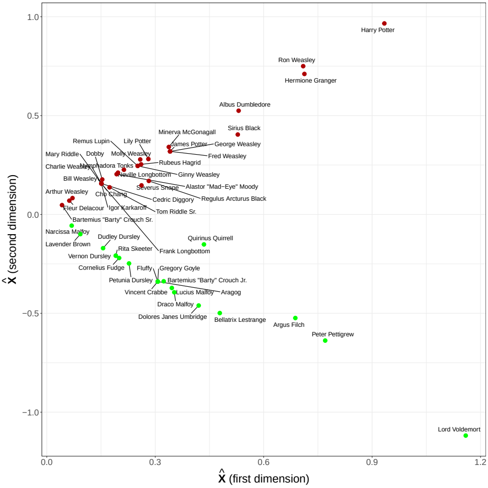
Then a graph follows a degree-corrected stochastic block model if

$$\mathbf{A}_{ij} \overset{ind}{\sim} \text{Bernoulli}(w_i w_j \mathbf{B}_{Z_i, Z_j}),$$

for $i < j$

To represent this as a GRDPG, find $\mathbf{V} = [V_1, \ldots, V_K]^\top$ such that $\mathbf{B} = \mathbf{V}\mathbf{I}_{p,q}\mathbf{V}^\top$, and set $X_i = w_i V_{Z_i}$.

Harry Potter enmity graph



Finite-rank latent position network models are high-dimensional GRDPGs

Definition (Latent position model)

Let $Z_1, \ldots, Z_n \in \mathcal{Z} \in \mathbb{R}^d$, $f : \mathcal{Z} \times \mathcal{Z} \rightarrow [0, 1]$ some kernel. Then the graph

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

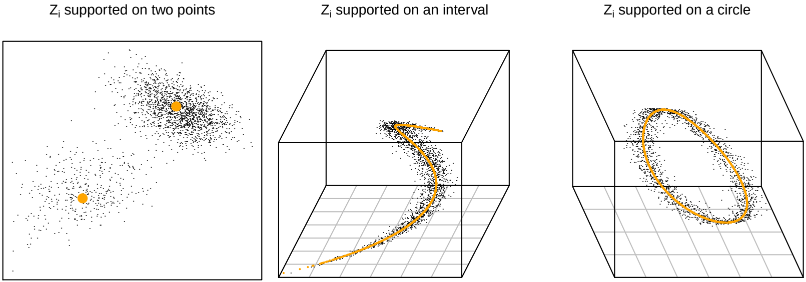
for $i < j$ is known as a latent position model

Then under regularity conditions — particularly that f has a finite rank D — we can find a map $\phi : \mathbb{R}^d \rightarrow \mathbb{R}^D$ such that

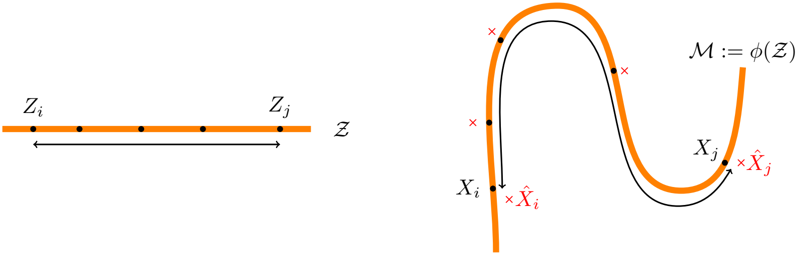
$$f(x, y) = \phi(x)^\top \mathbf{I}_{p,q} \phi(y),$$

for all $x, y \in \mathcal{Z}$.
In other words, the latent position model with positions Z_i is a GRDPG with positions $X_i = \phi(Z_i)$. Under various further conditions, the map is a homeomorphism, diffeomorphism, or even isometry...

Topological fidelity (simulated examples)

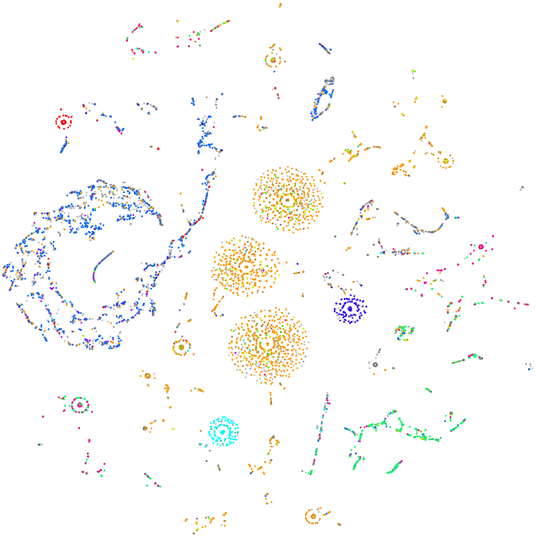


Illustration



$$\boxed{\|Z_i - Z_j\|_2 \propto d_{\mathcal{M}}\{X_i, X_j)\}}$$

Los Alamos National Lab computer network, spectral embedding followed by t-SNE



Problem

- In most cyber-security graphs, we need to fit a GRDPG not an RPDG (code and data for previous figure available).
- The spectral embedding procedure breaks down under e.g. sparsity and degree-heterogeneity, even where it's not evident that the GRDPG model is "wrong".
- It would be nice if we could fail in a more controlled way, e.g. with growing uncertainty bands around the estimated positions.
- In the case of the RDPG, Bayesian and likelihood-based procedures have been shown to have other advantages, e.g. reduced error variance. (Remark: I am not sure if this might not come at the expense of uniform consistency, which would be interesting but a definite disadvantage.)

Proposed project

- Extend Xie's Bayesian approach to the GRDPG; probably starting from paper (6), which has code and pseudo-code. This may involve solving or circumventing various mathematical questions, for example, finding a maximal set χ under the GRDPG and its volume. I have qualms about setting a "prior" on the latent positions, rather than their distribution, but maybe we live with this.
- Compute credible sets of latent positions. Again, this requires a little thought, even at a basic conceptual level, because of the unidentifiability by $\mathbf{Q}$. My proposed solution is to choose a default MAP configuration, and provide the credible set for any query node, with other nodes fixed to their MAP position.
- Investigate, even just empirically, whether we achieve reduced average error (over spectral embedding), and whether we achieve reduced *maximum* error (something I don't think has been commented on in Xie's papers).
- Comment on potential scalability.