Online Change Point Detection for Random Dot Product Graphs

Bernardo Marenco[1], Paola Bermolen[1], Marcelo Fiori[1], Federico Larroca[1], and Gonzalo Mateos[2]

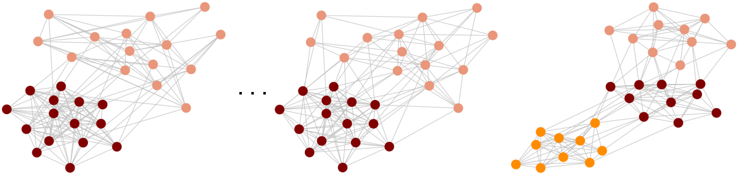
[1]Facultad de Ingeniería, Universidad de la República, Uruguay

[2]Dept. of Electrical and Computer Engineering, University of Rochester, Rochester, NY, USA.

Asilomar Conference on Signals, Systems, and Computers
Oct. 31st - Nov. 3rd, 2021

Problem statement

- **Given:** Stream of undirected graph observations



- **Model:** Random Dot Product Graph (RDPG) [Athreya et al. 2018]
- **Goal :** Detect in an online fashion when the underlying model changed

Contributions and impact

⇒ Marry sequential chage-point detection with graph representation learning

⇒ Explainable algorithm for (pseudo) real-time network monitoring

⇒ Guaranteed error-rate control, detection delay analysis

Random Dot Product Graphs (RDPGs)

- Node i has associated latent vector $\mathbf{x}_i \in \mathbb{R}^d$
- Edge (i, j) exists with probability $P_{ij} = \mathbf{x}_i^\top \mathbf{x}_j$
- Notation:

  $\Rightarrow$ n: number of nodes in all graphs
  $\Rightarrow$ $\mathbf{A} \in \{0, 1\}^{n \times n}$: adjacency matrix - hollow, symmetric
  $\Rightarrow$ $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]^\top \in \mathbb{R}^{n \times d}$: latent positions matrix
  $\Rightarrow$ $\mathbf{P} = \mathbf{X}\mathbf{X}^\top \in \mathbb{R}^{n \times n}$: matrix of edge probabilities

- Model is invariant to rotations in $\mathbf{X}$

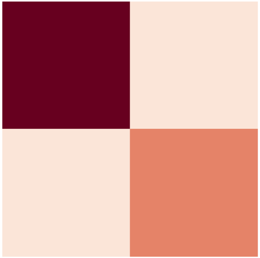
  $\Rightarrow$ $\mathbf{P} = \mathbf{X}\mathbf{X}^\top = \mathbf{X}\mathbf{W}(\mathbf{X}\mathbf{W})^\top$ for any orthogonal $\mathbf{W}$

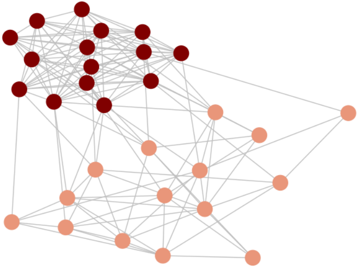
- Expressive model, SBM a special case of RDPG

- **Q:** Given a graph $\mathbf{A}$, how do we estimate the latent positions $\mathbf{X}$?

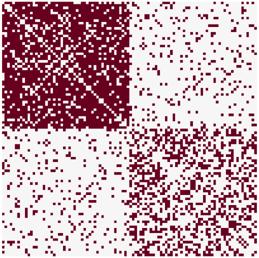
Adjacency Spectral Embedding

$\mathbf{P} = \mathbf{X}\mathbf{X}^\top$

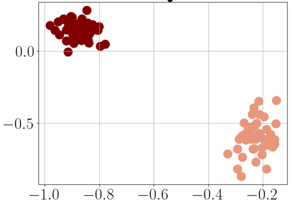




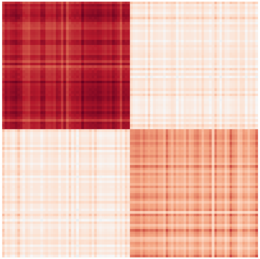
$\mathbf{A} = \mathbf{Q}\Lambda\mathbf{Q}^\top$



$\hat{\mathbf{X}} = \hat{\mathbf{Q}}\hat{\Lambda}^{1/2}$

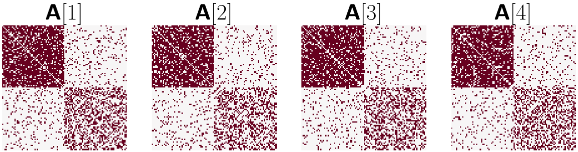


$\hat{\mathbf{P}} = \hat{\mathbf{X}}\hat{\mathbf{X}}^\top$

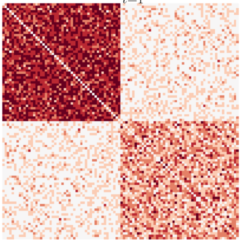


Training phase

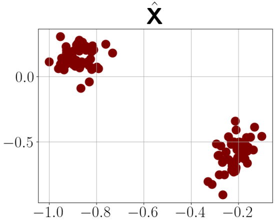
- **Idea:** Estimating function approach [Kirch and Tadjuidje Kamgaing 2015]
- Training set of m “clean” graphs with no change-point



$$\bar{\mathbf{A}} = \frac{1}{m}\sum_{t=1}^{m}\mathbf{A}[t]$$



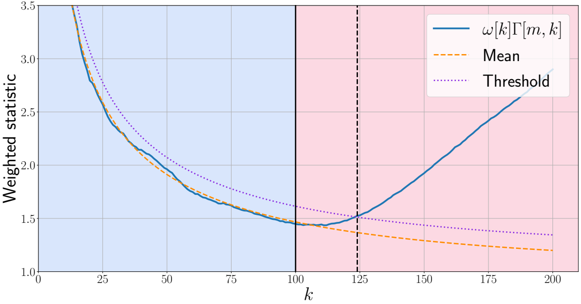
→



Operational phase

- Sequentially observe matrices $\mathbf{A}[m+1], \mathbf{A}[m+2], \ldots$
- Monitor the cumulative sum $\mathbf{S}[m,k] = \sum_{t=m+1}^{m+k} \left(\hat{\mathbf{X}}\hat{\mathbf{X}}^\top - \mathbf{A}[t] \right)$

Proposition: For large k, $\Gamma[m,k] := \|\mathbf{S}[m,k]\|^2$ has a generalized chi-squared distribution.

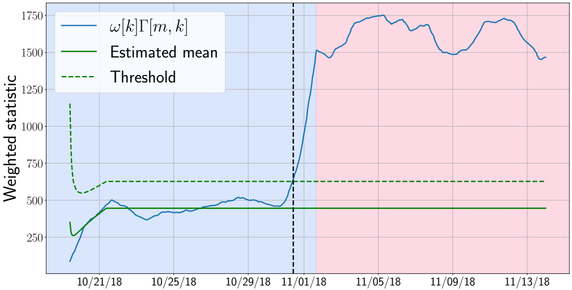


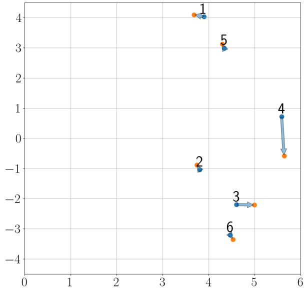
- Lightweight: $O(n^2)$ memory storage and computational complexity

Wireless network monitoring

- Extended RDPG to handle weighted, directed networks
- Real network of Wi-Fi APs. Hourly RSSI measurements for $n = 6$ nodes

  $\Rightarrow$ Ground-truth from network admin: *AP 4 was moved on 10/30*





- Explainability via interpretable ASE $\Rightarrow$ Identify source of change
- Reproducibility $\Rightarrow$ Try it @ `https://github.com/git-artes/cpd_rdpg`

References

Athreya, Avanti et al. (2018). "Statistical Inference on Random Dot Product Graphs: a Survey". In: *Journal of Machine Learning Research* 18.226, pp. 1–92. URL: http://jmlr.org/papers/v18/17-448.html.

Kirch, C. and J. Tadjuidje Kamgaing (2015). "On the use of estimating functions in monitoring time series for change points". In: *Journal of Statistical Planning and Inference* 161, pp. 25–49. ISSN: 0378-3758.