Final Honour School of Mathematics Part C

C5.4 Networks MINI-PROJECT

There are two alternative options. Students may submit a Mini-Project addressing either option, not both.

Option 1
Suppose we are given a network with n vertices, labelled v_i for $i = 1, \ldots, n$. We often wish to embed such a network within a suitable K-dimensional Euclidean space, $\mathbb{R}^K$, called the "embedding space". This corresponds to a mapping of each vertex v_i onto a chosen $\mathbf{x}_i \in \mathbb{R}^K$. It is desirable that the better connected are any two vertices then the closer they are located within the embedding space.

The resulting vertices form a cloud of n points, to which topological data analysis methods (including persistent homology) might be applied. For example the GLEE method set out in L. Torres, K.S. Chan, T. Eliassi-Rad, GLEE: Geometric Laplacian Eigenmap Embedding, Journal of Complex Networks, Volume 8, Issue 2, April 2020, cnaa007,`https://tinyurl.com/32xyjtp9`.

How should K be set? Using a spectral method, based on the graph Laplacian, one might look at the spectrum and identify any *gaps* at which one might fix or limit the embedding.

An alternative method is given by E. Estrada, The communicability distance in graphs Linear Algebra and its Applications Volume 436, Issue 11, 1 June 2012, Pages 4317-4328, `https://tinyurl.com/3z47kmju` involving the spectral decomposition of the adjacency matrix. Other methods are also common and advanced within the literature.

How might we estimate or set K, to be as small as possible for a given graph?

More specifically, suppose that we generate a test network using n given (random) locations, say $\mathbf{y} \in \mathbb{R}^J$ for some J, and we set up the adjacency matrix via $A_{ij} = 1$ iff $||\mathbf{y}_i - \mathbf{y}_j|| < \delta$ for some suitable $\delta > 0$; and $A_{ij} = 0$ otherwise. Then, given A, without any knowledge of the $\mathbf{y}$'s, how well might these embedding methods suggest an embedding dimension K which is close to the *actual value*, J?

Alternatively given any cloud of points (an embedding for the network's vertices) within a larger dimensional Euclidean space one may estimate the intrinsic dimension of that point cloud using the "two-nearest -neighbour methods" method introduced by Facco, E., d'Errico, M., Rodriguez, A., Laio, A. (2017), Estimating the *intrinsic dimension*, D_i, of dataset, by a minimal neighbourhood information. Sci Rep 7, 12140, `https://doi.org/10.1038/s41598-017-11873-y`. Of course the points may all lie on a D_i dimensional curved or folded manifold that is embedded within a much higher dimensional Euclidean space. How does such an estimate compare to estimators of K, from the embedding process?

Write a report that seeks to addresses this problem, that is posed within the general framework of network embedding methods. This includes, but is not limited to, the concept of Laplacian embedding, the embedding dimension, and the intrinsic dimension of the resulting embedded point cloud.

Your report must include some original numerical simulations (which you produce) and a description in your own words of the mathematical methods. Your project will be evaluated based on its technical quality, its originality, its breadth, and the quality of its presentation, each of these criteria having the same weight.

Option 2
We have seen that for the triadic closure mechanism (where friends of friends are more likely to become friends), together with the random, and independent, formation and loss of pairwise friendships results in a stochastic dynamical system for the friendship adjacency network, evolving in discrete time steps (see also M.C. Parsons, P. Grindrod, D.J. Higham, Bistability through Triadic Closure, Vol. 8, Issue 4, 2012 November 30, 2012 EDT `https://tinyurl.com/3y4fye5d`).

Generalise this model to the case where there are two schools in a town, with n_1 and n_2 pupils (fixed and large) respectively; and where friendships are formed and lost randomly, yet independently and distinctly, between pairs of pupils who attend the same schools and who attend different schools, together with the triadic closure mechanism being active.

How do simulations compare with the mean field deterministic approximation for different parameter values? Can one demonstrate multiple distinct stable states for the mean field deterministic dynamics, for a fixed set of parameters?

Begin by assuming there are no interactions at all between pupils from different schools, so you have two bistable copies of the single population dynamics previously studied. Then increase the inter-school friendships in a suitable way. Justify your own modelling assumptions,

Write a report that seeks to addresses this problem, that is posed within the general framework of evolving friendship networks in discrete time. This should include, but not be limited to, the concepts of triadic closure, Markov models, and the mean field approximation.

Your report must include some original numerical simulations (which you produce) and a description in your own words of the mathematical methods. Your project will be evaluated based on its technical quality, its originality, its breadth, and the quality of its presentation, each of these criteria having the same weight.

For either option, your report should be in the format and style of an article for the journal Proceedings of the National Academy of Sciences, whose template is available here: `https://tinyurl.com/4xn6u85k`

The report must include all sections (abstract, significance statement, etc.) in 6 pages, which corresponds to approximately to 4,000 words, with up to 50 references, and up to 4 medium-size graphical elements (i.e., figures and tables).

Your report need not contain original research results, though you must use some original research papers (not just review articles or books) as resources.