MUTUAL REINFORCEMENT AT SECOND ORDER

Francesca Arrigo (Strathclyde) , Desmond J. Higham (Edinburgh), and Francesco Tudisco (GSSI)

SIAM Workshop on Network Science 2020
July 9–10 · Toronto

Summary

We propose and analyse a general tensor-based framework for incorporating second order features into spectral network measures. This approach allows us to combine traditional pairwise links with information that records whether triples of nodes are involved in triangles. Our treatment covers classical spectral methods and recently proposed cases from the literature that incorporate simplex information. We also identify interesting extensions. In particular, we define a mutually-reinforcing (spectral) version of the classical clustering coefficient, where triangles are deemed to be important if they intersect with other important triangles. The underlying object of study is a constrained nonlinear eigenvalue problem associated with a cubic tensor. Using recent results from nonlinear Perron–Frobenius theory, we establish existence and uniqueness under mild conditions, and show that the new spectral measures can be computed efficiently and robustly with a nonlinear power method.

Background and Motivation

Network science is grounded in the concepts of nodes and edges. In particular, the adjacency matrix efficiently encodes the topology of a graph and is easily manipulated using linear algebraic techniques. Here, if there are n nodes, $A \in \mathbb{R}^{n\times n}$ and $A_{ij} = 1$ if nodes i and j are connected, with $A_{ij} = 0$ otherwise. We assume that the connections are undirected, so $A_{ij} = A_{ji}$. This network representation is at the heart of many centrality and clustering algorithms [8]. We motivate our work with *eigenvector centrality*, which quantifies the importance of node i by y_i, where $\mathbf{y} \in \mathbb{R}^n$ is the Perron eigenvector of A [5, 14]:

$$y_i \propto \sum_{j=1}^{n} A_{ij} y_j, \quad \mathbf{y} \geq 0.$$

This centrality measure was popularized in the social network literature in the 1970s, [5], but can be traced back to algorithms proposed in the 1890s for ranking players involved in chess tournaments [14]. Eigenvector centrality is *mutually reinforcing*, in the sense that the importance of node i is proportional to the importance of its neighbours. We note that the concept of mutual reinforcement is also at the heart of Google's PageRank measure [12].

It is becoming apparent, however, that many important network features arise from the interaction of larger groups of nodes [2, 11, 13]. Information on higher-order interactions among nodes is *indirectly* used in many network science algorithms by considering traversals around the network. However, recent work [2, 3, 4, 7, 11] has shown that there is benefit in *directly* taking this information into account when designing algorithms. In this talk we discuss a general tensor-based framework for incorporating second-order features, i.e., triangles, into network measures [1].

Nonlinear Eigenvalue Framework

Our object of study is a new constrained nonlinear eigenvalue problem associated with a cubic tensor $\boldsymbol{T}$:

$$\alpha A\mathbf{x} + (1-\alpha)\boldsymbol{T}_p(\mathbf{x}) = \lambda\,\mathbf{x}, \qquad \mathbf{x} \geq 0, \tag{1}$$

where $\alpha \in [0, 1]$ and

$$\boldsymbol{T}_p(\mathbf{x})_i = \sum_{j,k=1}^{n} \boldsymbol{T}_{ijk} \left(\frac{|x_j|^p + |x_k|^p}{2}\right)^{1/p}. \tag{2}$$

For example, we may introduce the *binary triangle tensor*, where $\boldsymbol{T}_{ijk} = 1$ if nodes i, j, k form a triangle and $\boldsymbol{T}_{ijk} = 0$ otherwise. In this new measure, a node inherits extra importance from sharing triangles with nodes that themselves take part in important triangles.

We note that the power mean parameter p in (2) allows us to move, for example, between $\max\{|x_j|, |x_j|\}$, given by the limit $p \to \infty$, and the geometric mean, $\sqrt{|x_i x_j|}$ given by the limit $p \to 0$. The parameter α interpolates between traditional edge-based eigenvector centrality, $\alpha = 1$, and a purely second-order version, $\alpha = 0$.

A novel mutually-reinforcing (spectral) version of the classical Watts-Strogatz clustering coefficient [15] arises when we take $\alpha = 1$ in and set $\boldsymbol{T}_{ijk} = 1/(d_i(d_i - 1))$ if nodes i, j, k form a triangle and $\boldsymbol{T}_{ijk} = 0$ otherwise, where d_i denotes the degree of node i.

Theoretical and Experimental Results

Using recent advances in nonlinear Perron–Frobenius theory [9], we are able to establish existence of a unique solution $\mathbf{x}$ to (1) under mild conditions on the network topology. Moreover, these new spectral measures can be computed efficiently and robustly using a nonlinear power method. Just as for the standard power method, (a) the iteration is guaranteed to converge from any positive starting vector, (b) the convergence rate is linear, and (c) the main computational cost at each iteration is a matrix-vector multiplication with A, and hence the method is well-suited to the large-scale sparse networks arising in many applications.

To illustrate the type of computational result that will be presented in the talk, Figure 1 shows a box and whiskers plot for the ratio of link prediction success with and without second order information. Here, we used a seeded PageRank algorithm [10, 16] on a citation network with 233 nodes and 994 edges. We repeatedly removed 10% of the edges at random for the algorithms to predict. The prevalence of ratios larger than one indicates that the use of second order information is beneficial.

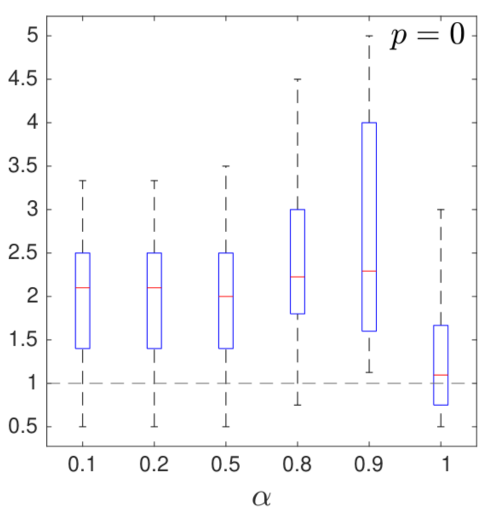

Figure 1: Box and whisker plot showing median, and 25th and 75th percentiles, for the ratio of success in predicting randomly removed edges with and without second order information. Here, success is measured by the number of correct entries among the predicted top 10%. Ratios greater than one indicate that second order information is beneficial. We used $p = 0$ in (2) and show results for various α in (1).

Acknowledgements

The work of FA was supported by fellowship ECF-2018-453 from the Leverhulme Trust. The work of DJH was supported by EPSRC/RCUK Established Career Fellowship EPM00158X/1 and by EPSRC Programme Grant EP/P020720/1. The work of FT was partially supported by INdAM–GNCS and EPSRC Programme Grant EP/P020720/1.

References

[1] F. Arrigo, D. J. Higham and F. Tudisco. "A framework for second-order eigenvector centralities and clustering coefficients". Submitted (2019).

[2] A. R. Benson, D. F. Gleich and J. Leskovec. "Higher-order organization of complex networks." Science, 353 (2016), pp. 163–166.

[3] A. R. Benson, R. Abebe, M. T. Schaub, A. Jadbabaie and J. Kleinberg. "Simplicial closure and higher- order link prediction." Proceedings of the National Academy of Sciences, 115 (2018), pp. E11221–E11230.

[4] A. R. Benson. "Three Hypergraph Eigenvector Centralities". SIAM Journal on Mathematics of Data Science, 1(2019), pp. 293–312.

[5] P. Bonacich. "Power and centrality: a family of measures". American Journal of Sociology, 92 (1987), pp. 1170–1182.

[6] E. Estrada and F. Arrigo. "Predicting triadic closure in networks using communicability distance functions". SIAM Journal on Applied Mathematics, 75(2015), pp. 1725–1744.

[7] N. Eikmeier and D. F. Gleich, "Classes of Preferential Attachment and Triangle Preferential Attachment Models with Power-law Spectra". Journal of Complex Networks, (to appear).

[8] E. Estrada and D. J. Higham. "Network properties revealed through matrix functions". SIAM Review, 52 (2010), pp. 696–671.

[9] A. Gautier, F. Tudisco and M. Hein. "A unifying Perron–Frobenius theorem for nonnegative tensors via multihomogeneous maps". SIAM Journal on Matrix Analysis and Applications 40 (2019), pp. 1206–1231.

[10] D. Gleich and K. Kloster. "Seeded PageRank solution paths." European Journal of Applied Mathematics 27 (2016), 812-–845.

[11] A. P. Kartun–Giles and G. Bianconi. "Beyond the clustering coefficient: A topological analysis of node neighbourhoods in complex networks". Chaos, Solitons & Fractals: X, (in press).

[12] L. Page, S. Brin, R. Motwani and T. Winograd. "The PageRank citation ranking: Bringing order to the web". Technical Report. Stanford InfoLab, (1999).

[13] V. Salnikov, D. Cassese and R. Lambiotte. "Simplicial complexes and complex systems". European Journal of Physics, 40 (2019), 014001.

[14] J. P. Schäfermeyer. "On Edmund Landau's contribution to the ranking of chess players". Unpublished manuscript (2019).

[15] D. J. Watts and S. H. Strogatz. "Collective dynamics of 'small-world' networks". Nature, 393 (1998), pp. 440–442.

[16] G. Jeh and J. Widom J. "Scaling personalized web search". In Proceedings of the 12th International Conference on World Wide Web, ACM (2003), 271–279.