A fast and robust kernel optimization method for core--periphery detection in directed and weighted graphs

A fast and robust kernel optimization method for core–periphery detection in directed and weighted graphs

Francesco Tudisco[*] and Desmond J. Higham

[*]Correspondence:
tudisco.francesco@gmail.com
School of Mathematics, University
of Edinburgh, Peter Guthrie Tait
Road, EH9 3FD Edinburgh, UK
Full list of author information is
available at the end of the article

Abstract

Many graph mining tasks can be viewed as classification problems on high dimensional data. Within this class we consider the issue of discovering core-periphery structure, which has wide applications in the economic and social sciences. In contrast to many current approaches, we allow for weighted and directed edges and we do not assume that the overall network is connected. Our approach extends recent work on a relevant relaxed nonlinear optimization problem. In the directed, weighted setting, we derive and analyze a globally convergent iterative algorithm. We also relate the algorithm to a maximum likelihood reordering problem on an appropriate core-periphery random graph model. We illustrate the effectiveness of the new algorithm on a large scale directed email network.

Keywords: network; nonlinear Perron–Frobenius; power method; relaxation

1 Introduction

Graph theory gives a common framework for formulating and tackling a range of problems arising in data science. Many such tasks can be viewed in terms of categorizing nodes or discovering hidden substructures that relate them. Clustering, or community detection, is perhaps the most widely studied problem, and it forms the basis of many classification algorithms [1]. In this work we study the different, but closely related, issue of identifying *core–periphery* structure; we seek a set of nodes that are highly connected internally and with the rest of the network, forming the *core*, and a set of *peripheral* nodes that are strongly connected to the core but have only sparse internal connections. This kind of structure is important for a number of reasons. For example, identifying core–periphery structures can help in identifying and categorizing hubs, i.e., well-connected nodes. As noted in [2], such nodes often occur in real–world networks. This is an issue for some community detection methods, as hubs tend to be connected to many different communities and, thus, can be awkward to classify. Moreover, the set of core nodes can be used to identify internally cohesive subgraphs of highly central nodes. In fact, even though all core nodes typically have high centrality score, not all nodes with high centrality measures belong to the core and it is possible to find sparsely connected subgraphs of central nodes not belonging to the core [3].

The concept of the network core–periphery is closely related to the idea of rich-clubs, nested networks and onion network structures [4, 5, 6]. In particular, a number

of core–defining algorithms have been proposed in recent years, e.g., [7, 8, 9], following the seminal work by Borgatti and Everett [3]. Core–periphery structure has been detected and interpreted in many complex systems, including protein–protein interaction networks [10], metabolic and gene regulatory networks [11], social networks [3, 12], engineered networks (such as the Internet, power-grids or transportation networks) [9], and economic networks [13]. See also the review [14].

From a computational perspective, several recent works provide algorithms that apply to undirected networks. In particular, we have introduced in [9] a scalable nonlinear optimization method with global quality guarantees for core–periphery detection in binary, undirected and connected graphs. This method exploits an intriguing connection between optimization and nonlinear eigenproblems and allows for a fast and easily implementable iteration which guarantees to compute the global maximum of a highly nonconvex core–score quality function.

In this work we consider the core-periphery concept in the more general setting of directed, weighted and possibly disconnected networks and we extend the results of [9], both in terms of the algorithms and of the theoretical analysis, to this more challenging case.

In our directed case, we use the concept that a set of nodes forms a core if there are many core-to-core, core-to-periphery and periphery-to-core edges, with few periphery-to-periphery edges. Although the ideal core–periphery subdivision defines two well distinguished sets of nodes, in practice one often looks for a core–score vector $\boldsymbol{u} \geq 0$ such that a smaller value u_i indicates that node i is more peripheral. Such an assignment may be viewed as a type of node centrality measure [15]. Indeed the classic cases of degree centrality and eigenvector centrality have been proposed and tested in this context [3, 8, 9, 16], and, as we explain in Section 4, the approach we propose here may be viewed as a nonlinear generalization of both these cases.

The manuscript is organized as follows. Section 2 introduces some relevant notation. In Section 3 we express the core-periphery detection problem in terms of kernel-based optimization and in Section 4 we connect this idea with classical node centrality measures. Section 5 shows that another viewpoint is also relevant; the approach may be viewed as maximum likelihood reordering under a new random graph model that generates directed core-periphery structure. In Section 6 we study the nonlinear optimization problem and show that it may be solved via an inexpensive and globally convergent iteration. Section 7 illustrates the performance of the algorithm on a large scale email dataset and Section 8 gives some conclusions.

2 Notation

We consider directed and possibly weighted graphs $G = (V, E)$ with node set $V = \{1, \ldots, n\}$ and adjacency matrix $A = (A_{ij})$.

If node i does not point to node j then the entry a_{ij} is zero. Otherwise, a_{ij} takes a positive value, accounting for the strength of the directional tie from i to j.

We let $\mathbb{1}$ denote the column vector in $\mathbb{R}^n$ with all values equal to one, and define the in and out degree vectors as $\boldsymbol{d}^{\text{in}} = A\mathbb{1}$ and $\boldsymbol{d}^{\text{out}} = A^T\mathbb{1}$, respectively. Operations on and between vectors are to be interpreted in a componentwise sense, so that, for example, $\boldsymbol{x}^{p-1}$ has ith component given by x_i^{p-1} and $\boldsymbol{x}^{p-1}\boldsymbol{y}$ has ith component given

by $x_i^{p-1} y_i$. Inequalities involving vectors and matrices are also to be interpreted componentwise, so that, for example, $A \geq 0$ means $A_{ij} \geq 0$ for all i and j.

3 Core–periphery via functional kernel optimization

To search for the presence of a core and periphery we define a core–score vector, that is, a nonnegative vector $\boldsymbol{u}$ quantifying the coreness of the nodes, where $u_i > u_j$ indicates that node i is closer to the core than node j. We define our core–score vector as the solution to the following nonconvex and constrained core–periphery quality function maximization problem

$$
\begin{aligned}
\max \quad & f_\alpha(\boldsymbol{x}) \\
\text{s. t.} \quad & \boldsymbol{x} \geq 0 \text{ and } \|\boldsymbol{x}\| = 1
\end{aligned}
\tag{1}
$$

where, for some fixed real number $\alpha \in \mathbb{R}$ to be chosen, f_α is the core–quality function

$$
f_\alpha(\boldsymbol{x}) = \sum_{i,j=1}^n A_{ij} \, \kappa_\alpha(x_i, x_j), \qquad \kappa_\alpha(x, y) = \left(\frac{|x|^\alpha + |y|^\alpha}{2} \right)^{1/\alpha}.
\tag{2}
$$

Note that, since only relative values are important, a constraint of the form $\|\boldsymbol{x}\| = 1$ is very natural. However, there is no reason at this stage to prefer a particular norm over another. Therefore, we assume for now that $\| \cdot \|$ is any vector norm and consider the problem (1) in this general setting.

For $x, y \in \mathbb{R}$, the kernel $\kappa_\alpha(x, y)$ is the generalized (or Binomial) mean of the two nonnegative numbers $|x|$ and $|y|$. The case $\alpha \to \infty$ is particularly well-suited for core-periphery purposes. If $\boldsymbol{x}$ is a nonnegative vector, we have

$$
f_\infty(\boldsymbol{x}) := \lim_{\alpha \to \infty} f_\alpha(\boldsymbol{x}) = \sum_{ij=1}^n A_{ij} \max\{x_i, x_j\},
$$

and thus any nonnegative vector $\boldsymbol{x}$ for which $f_\infty(\boldsymbol{x})$ is large assumes a necessarily large value on the entries involving the nodes in the core and smaller values within the periphery. In fact, when $\| \cdot \|$ denotes a p-norm, any vector $\boldsymbol{x} \geq 0$, $\|\boldsymbol{x}\| = 1$ such that $f_\infty(\boldsymbol{x})$ is large assigns to each node a value x_i between zero and one so that each connection between two nodes i, j in the graph or, equivalently, each nonzero in the weight matrix A, involves at least one node such that x_i is large. We note that the relevance of $f_\infty(\boldsymbol{x})$ as a core–periphery quality function is highlighted for example in [8], and the relaxed version (1) involving α was considered in [9] for undirected graphs.

4 Connection with degree and eigenvector centralities

In the undirected case, $A = A^T$, it has been argued that both the degree vector and the eigenvector (or Bonacich) centrality vector carry interesting core–periphery information and are good candidates for core score vectors [3, 8, 9, 16]. In this section we show that when $\alpha = 1$ or $\alpha = 0$ the problem (1) admits an explicit solution that, even when the graph is directed, boils down to the degree and the eigenvector centrality, respectively.

When $\alpha = 1$ the function $f_\alpha(\boldsymbol{x})$ is linear, taking the form of the scalar product

$$f_1(\boldsymbol{x}) = \sum_{ij=1}^{n} A_{ij}\,\kappa_1(x_i, x_j) = \frac{1}{2}\mathbb{1}^T(A + A^T)\boldsymbol{x}\,.$$

As both the matrix $A + A^T$ and the vector $\boldsymbol{x}$ have nonnegative entries, using the Cauchy–Schwarz inequality, we have

$$f_1(\boldsymbol{x}) = \frac{1}{2}|\mathbb{1}^T(A + A^T)\boldsymbol{x}| \leq \frac{1}{2}\|\mathbb{1}^T(A + A^T)\|_2\|\boldsymbol{x}\|_2$$

and the inequality is always strict unless $\boldsymbol{x}$ is a multiple of $(A + A^T)\mathbb{1} = \boldsymbol{d}^{\text{in}} + \boldsymbol{d}^{\text{out}}$. Therefore, if we choose the norm constraint in (1) to be $\|\boldsymbol{x}\|_2 = 1$, we have that

$$\max_{\boldsymbol{x} \geq 0:\|\boldsymbol{x}\|_2 = 1} f_1(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{d}^{\text{in}} + \boldsymbol{d}^{\text{out}}\|_2$$

with maximizer $\boldsymbol{u}$ given by $\boldsymbol{u} = (A + A^T)\mathbb{1} = \boldsymbol{d}^{\text{in}} + \boldsymbol{d}^{\text{out}}$ properly normalized. This shows that the solution of (1) reduces to the degree vector when the graph is undirected and coincides with the sum of the incoming and outgoing degree vectors in the general case.

It is well-known that when $\alpha \to 0$, $\kappa_\alpha(x, y)$ converges to the geometric mean of $|x|$ and $|y|$. Thus, when $\alpha = 0$ and $\boldsymbol{x} \geq 0$, we have

$$f_0(\boldsymbol{x}) := \lim_{\alpha \to 0} f_\alpha(\boldsymbol{x}) = \sum_{ij=1}^{n} A_{ij}\,\sqrt{x_i x_j}\,.$$

As $\boldsymbol{x} \mapsto \sqrt{\boldsymbol{x}}$ is bijective on the set of vectors with nonnegative entries, we can change variable $\boldsymbol{y} = \sqrt{\boldsymbol{x}}$ in (1) and recast the problem (1) as

$$\max_{\boldsymbol{y} \geq 0} \sum_{ij} A_{ij}\, y_i y_j = \boldsymbol{y}^T A \boldsymbol{y}, \qquad \text{subject to } \|\boldsymbol{y}^2\| = 1.$$

Again, if we consider the 1-norm, we can write the constraint $\|\boldsymbol{y}^2\|_1 = 1$ as $\|\boldsymbol{y}^2\|_1 = \boldsymbol{y}^T\boldsymbol{y} = 1$ and problem (1) becomes

$$\max_{\boldsymbol{x} \geq 0} f_0(\boldsymbol{x}) = \max_{\boldsymbol{y} \geq 0} f_0(\boldsymbol{y}^2) = \max_{\boldsymbol{y} \geq 0} \frac{\boldsymbol{y}^T A \boldsymbol{y}}{\boldsymbol{y}^T \boldsymbol{y}}\,.$$

By the Perron–Frobenius theorem [17], the Rayleigh quotient $\boldsymbol{y}^T A \boldsymbol{y}/\boldsymbol{y}^T \boldsymbol{y}$ has a unique nonnegative maximizer $\boldsymbol{c}$, which coincides with the Perron eigenvector of the nonnegative matrix $A + A^T$. In other words, the core score that maximizes f_0 is the vector $\boldsymbol{u} = \boldsymbol{c}^2$, where $\boldsymbol{c}$ is the eigenvector centrality of the symmetrized network with weight matrix $A + A^T$.

In Section 6 we show that both these two cases are actually a special case of a more general setting. We prove that for any $\alpha \geq 0$ the solution of (1) is the Perron eigenvector of a nonlinear core–periphery operator. In particular, this implies that the solution $\boldsymbol{u}$ to (1) is unique for any $\alpha \geq 0$, with an appropriate normalization. Moreover, using nonlinear Perron–Frobenius theory, this further allows us to introduce an iterative algorithm that computes $\boldsymbol{u}$, with global convergence guarantees.

5 Logistic core-periphery random model for directed graphs

We introduced in [9] a random graph model for undirected and unweighted graphs that can be used to artificially generate networks with a planted core–periphery structure. This model, unlike more classical block-based versions, is based on the logistic sigmoid function $1/(1 + e^{-x})$ rather than a Heaviside step function and allows a smooth transition between the set of core nodes and the set of peripheral ones. We refer to it as the *logistic core–periphery random model*. We notice that similar logistic function based random models have been considered in [18, 19, 20].

Here we extend the model to the case of directed graphs and we prove that the method of maximum likelihood applied to this random model coincides with the core–periphery quality function maximization problem (1), which provides the core–periphery analogue of a known phenomenon for stochastic block models in the community detection case [21].

Consider a core–ranking assignment, that is, a nonnegative permutation vector $\boldsymbol{\pi}$ that assigns a distinct integer π_i between 1 and n to each vertex i. The closer π_i is to 1, the higher the rank of i as a member of the core. For convenience, we shift-and-scale the core–ranking vectors via the affine transform $\boldsymbol{u} \mapsto \mathbb{1} - \boldsymbol{\pi}/n$. Hence, we consider the set

$$\mathrm{CR}(n) = \left\{ \boldsymbol{u} \in \mathbb{R}^n : u_i = 1 - \pi_i/n, \ \boldsymbol{\pi} \text{ is a permutation of } \{1, \ldots, n\} \right\},$$

so that, similarly to a core–score assignment, $\boldsymbol{u} \in CR(n)$ has values in $[0, 1]$ and larger values of $\boldsymbol{u}$ correspond to higher positions in the core ranking.

Now, given $\boldsymbol{u} \in \mathrm{CR}(n)$, the logistic core–periphery random model generates an edge from node i to node j with independent probability given by

$$\Pr(i \to j) = \frac{1}{1 + e^{-\kappa_\alpha(u_i, u_j)}} = p_{ij}(\boldsymbol{u}). \tag{3}$$

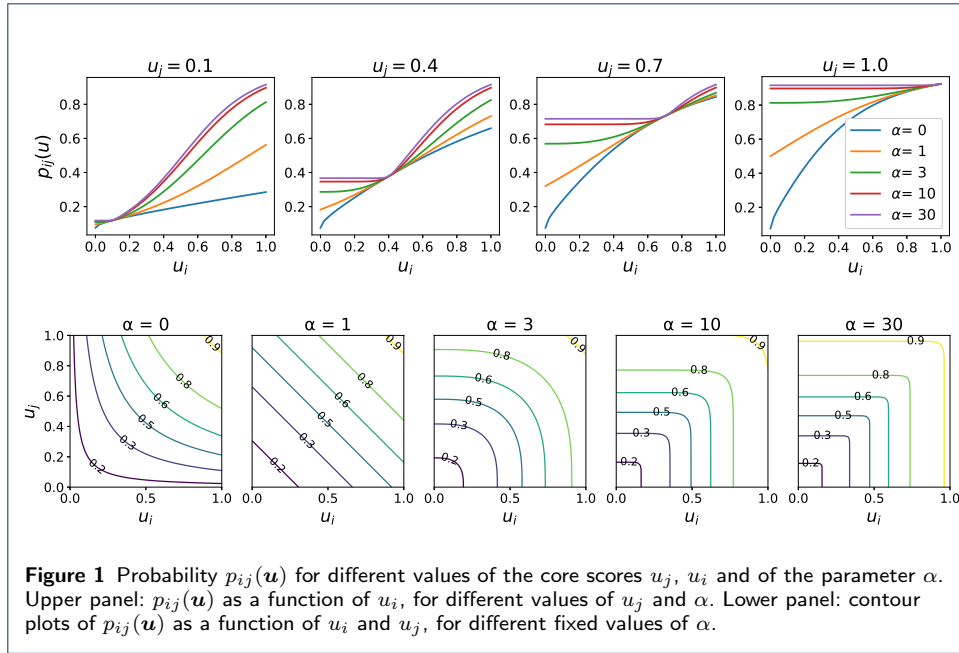Note that for $\alpha_2 \geq \alpha_1 \geq 0$ we have

$$\sqrt{|xy|} = \kappa_0(x, y) \geq \kappa_{\alpha_1}(x, y) \geq \kappa_{\alpha_2}(x, y) \geq \kappa_\infty(x, y) = \max\{|x|, |y|\}.$$

Thus, for any $\alpha \geq 0$, the probability $p_{ij}(\boldsymbol{u})$ tends to be large if at least one of the nodes i and j has a high core rank and this effect increases as α grows, as shown by Figure 1.

Suppose we are given a network with the nodes in arbitrary order and wish to find the best core ranking assignment based on the logistic random model (3). From a maximum likelihood perspective, this corresponds to maximizing the log-likelihood

$$L_\alpha(\boldsymbol{u}) = \sum_{ij \in E} \log p_{ij}(\boldsymbol{u}) + \sum_{ij \notin E} \log(1 - p_{ij}(\boldsymbol{u})) \tag{4}$$

among all possible $\boldsymbol{u} \in \mathrm{CR}(n)$. In other words, assuming that the given network is a sample from the logistic core–periphery random model (3) with the node labels shuffled arbitrarily, this is most likely to be the correct reordering.

Figure 1 Probability $p_{ij}(\boldsymbol{u})$ for different values of the core scores u_j, u_i and of the parameter α. Upper panel: $p_{ij}(\boldsymbol{u})$ as a function of u_i, for different values of u_j and α. Lower panel: contour plots of $p_{ij}(\boldsymbol{u})$ as a function of u_i and u_j, for different fixed values of α.

A key observation here is that, for ranking vectors $\boldsymbol{u} \in \mathrm{CR}(n)$, this maximum likelihood approach is equivalent to assigning a core–score to the nodes which maximizes a logistic core quality function f_α. In fact, the following extension of Theorem 3.1 in [9] holds.

Theorem 1 *Let G be a directed unweighted graph. For any $\alpha \geq 0$, a vector $\boldsymbol{u}^\star \in \mathrm{CR}(n)$ is solution of*

$$\max_{\boldsymbol{u} \in \mathrm{CR}(n)} L_\alpha(\boldsymbol{u})$$

if and only if it is solution of

$$\max_{\boldsymbol{u} \in \mathrm{CR}(n)} f_\alpha(\boldsymbol{u}).$$

This equivalence provides further justification for the kernel optimization approach. It also suggests that the logistic core–periphery random model (3) is a useful resource for testing core–periphery detection algorithms in this directed setting.

We also note that a closely related generative random graph model for core–periphery networks was proposed in [20]. That work focused on the undirected case and aimed to incorporate additionally available spatial information.

6 Core-periphery nonlinear operator

A study of the Hessian of f_α reveals that f_α is neither convex nor concave in general. This makes the solution of (1) particularly challenging. However, here we show that this optimization problem can be re-cast in terms of the Perron eigenvector of a nonlinear operator. We then show how its solution is always achievable via a generalization of the classical power method from numerical linear algebra.

Given $\alpha \geq 1$, consider the nonlinear core-periphery operator $\Phi_\alpha : \mathbb{R}^n \rightarrow \mathbb{R}^n$, entrywise defined as follows

$$\boldsymbol{x} \mapsto \Phi_\alpha(\boldsymbol{x})_i = |x_i|^{\alpha-2} x_i \sum_{j=1}^n \frac{A_{ij} + A_{ji}}{\kappa_\alpha(x_i, x_j)^{\alpha-1}} \quad i = 1, \ldots, n.$$

Given $p > 1$, we consider the following nonlinear eigenvalue problem for Φ_α

$$\Phi_\alpha(\boldsymbol{x}) = \lambda \, \boldsymbol{x}^{p-1} \,. \tag{5}$$

One easily realizes that Φ_α is linear if and only if $\alpha = 1$ in which case that operator degenerates into the map such that $\Phi_1(\boldsymbol{x}) = \boldsymbol{d}^{\mathrm{in}} + \boldsymbol{d}^{\mathrm{out}}$, for any nonnegative vector $\boldsymbol{x} \geq 0$. In this setting it is easily seen that the only nonnegative solution of (5) is $\boldsymbol{x} = \boldsymbol{d}^{\mathrm{in}} + \boldsymbol{d}^{\mathrm{out}}$ with $\lambda = 1$ and $p = 2$. Combined with the discussion of Section 4, this shows that for the case $\alpha = 1$ the unique nonnegative solution of the eigenvalue problem (5) coincides with the maximizer of (1). We can retrieve the same analogy for $\alpha \rightarrow 0$ and $p = 1$. In that case we have $\lim_{\alpha \rightarrow 0} \Phi_\alpha(\boldsymbol{x})_i = \Phi_0(\boldsymbol{x})_i = \frac{1}{\sqrt{x_i}} \sum_{j=1}^n (A_{ij} + A_{ji}) \sqrt{x_j}$ whereas (5) becomes $\Phi_0(\boldsymbol{x}) = \lambda \mathbb{1}$. Arguing as in Section 4, again, we deduce that for $\alpha = 0$ and $p = 1$ a nonnegative solution of the eigenvalue problem (5) coincides with a maximizer of (1).

When $\alpha \neq 0, 1$ the question of existence and uniqueness of a solution to (5) is less trivial. The following theorem gives a full answer and shows that the same one-to-one correspondence between (5) and (1) holds.

Theorem 2   *Let $\alpha \geq 0$ and $p > \max\{1, \alpha\}$. Then the eigenvalue equation (5) has a unique nonnegative solution $\boldsymbol{u} \geq 0$ such that $\|\boldsymbol{u}\|_p := (|u_1|^p + \cdots + |u_n|^p)^{1/p} = 1$ which is also the unique solution of (1), provided that $\| \cdot \| = \| \cdot \|_p$. Moreover $\boldsymbol{u}$ is positive if and only if the network has no isolated nodes, i.e., all nodes have at least one outgoing or one incoming edge.*

Note that, as no assumption on the connectedness of the graph is made, the eigenvector centrality, i.e., the nonnegative solution of (5) for $\alpha = 0$ and $p = 1$, is not uniquely defined. Instead, Theorem 2 shows that the core–score assignment is always unique when $p > \max\{1, \alpha\}$ and $\alpha \geq 0$. The relevance of Theorem 2 is not only theoretical. In fact, it comes together with the following corollary which shows the global convergence to $\boldsymbol{u}$ of a simple iterative scheme.

Corollary 3   *Given an initial guess $\boldsymbol{u}_0 > 0$ and parameters $\alpha \geq 0$, $p > \max\{1, \alpha\}$ and $q = p/(p-1)$, consider the following iterative method*

$$\begin{cases} \boldsymbol{v}_{k+1} = \Phi_\alpha(\boldsymbol{u}_k) \\ \boldsymbol{u}_{k+1} = \|\boldsymbol{v}_{k+1}\|_q^{1-q} |\boldsymbol{v}_{k+1}|^{q-2} \boldsymbol{v}_{k+1} \end{cases} , \quad k = 0, 1, 2, 3, \ldots$$

Then $\boldsymbol{u}_k \geq 0$ for all $k \geq 0$ and $\|\boldsymbol{u}_k - \boldsymbol{u}\| = \mathcal{O}\big((\frac{\alpha-1}{p-1})^k\big)$, i.e., $\boldsymbol{u}_k$ converges to the unique solution $\boldsymbol{u} \geq 0$ of (1) and (5).

Note that on sparse networks the method scales linearly with the number of nodes. In fact, each iteration requires $\mathcal{O}(|E|)$ floating point operations, where $|E|$ is the number of edges in the graph. Moreover, the free parameter p allows us to tune the overall number of iterations $k^\star = \mathcal{O}(\ln \varepsilon / \ln \frac{\alpha-1}{p-1})$ required to achieve the precision $\|\boldsymbol{u}_{k^\star} - \boldsymbol{u}\| = \mathcal{O}(\varepsilon)$. We will refer to the iteration in Corollary 3 as the Nonlinear Spectral Method (NSM).

From Corollary 3 the choice $p \gg \alpha$ appears to be attractive, since it leads an extremely rapid (linear) convergence rate. However, in choosing values for p and α, we must take account of two further issues.

1. As we argued in Section 3, a larger value of α gives a kernel that more closely matches the ideal of $\max\{x_i, x_j\}$.

2. A larger value of p produces a relaxed problem that is less likely to distinguish between the nodes. (Note that in the extreme case of $p = \infty$, the constraint $\|\boldsymbol{x}\|_\infty = 1$ allows for the obvious solution $\boldsymbol{x} = \mathbb{1}$, which assigns the same score to all nodes.)
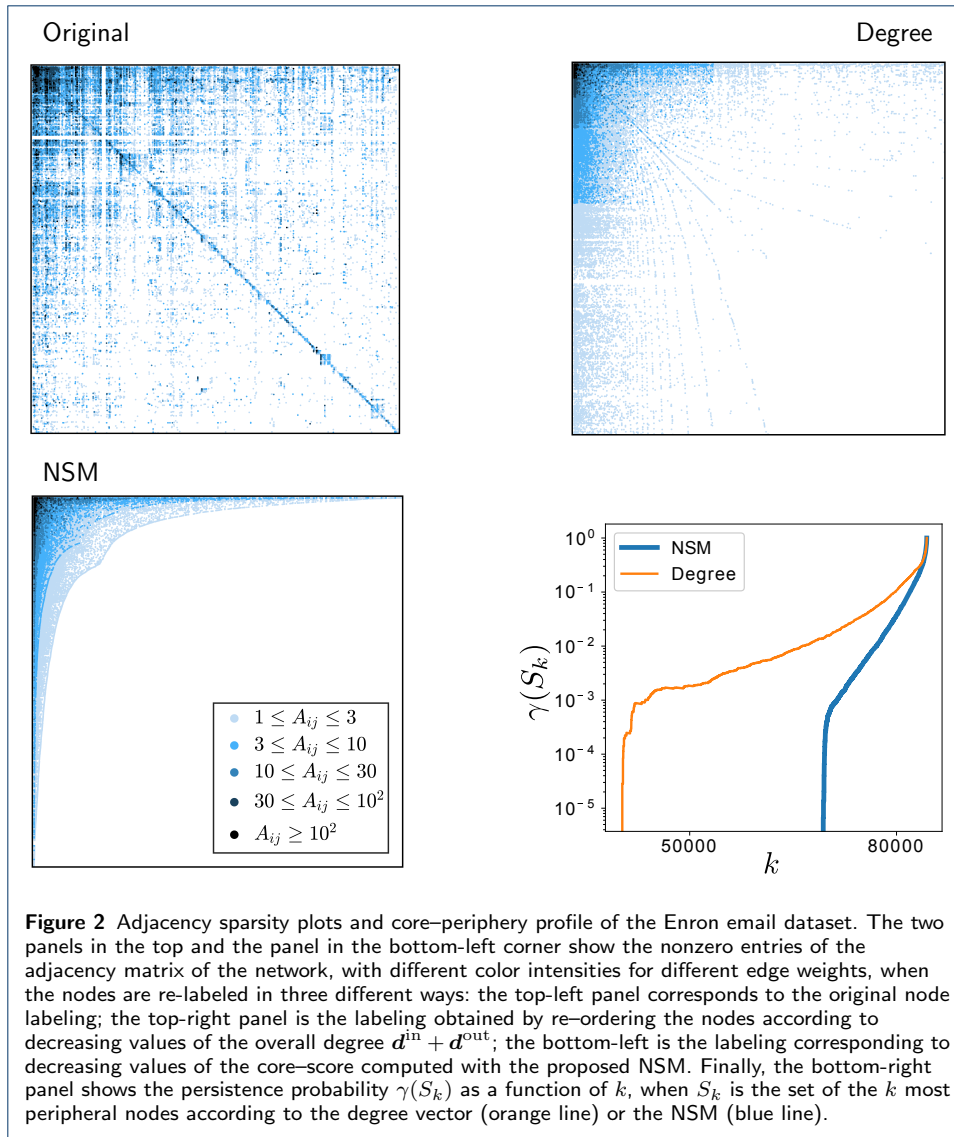
Combining points 1 and 2 with Corollary 3, we must compromise between a large parameter α in the kernel and a not-too-large value p for the vector norm, while keeping $p > \alpha$ to maintain convergence. In practice, we found that changing the value of p did not significantly affect the core–periphery structure output of the algorithm, which was instead governed by the value of α. In our experiments we chose $p = 2\alpha$ and $\alpha = 10$, as this produced good results with guaranteed fast convergence. Moreover, we observed that larger values of α did not produce a noticeable change in the core–periphery structure identified.

7 Enron dataset

The Enron email network consists of 1,148,072 emails sent between 87,273 employees of Enron between 1999 and 2003. Nodes in the network are individual employees and weighted directed edges, with weights ranging from 1 to 3,904, count the number of emails sent from one employee to another. It is possible to send an email to oneself, and thus this network contains self–loops. Note that this network is not strongly (or even weakly) connected. The data has been collected from [22].

Three plots in Figure 2 display the network by means of colored adjacency sparsity plots. Here, each nonzero entry in the adjacency matrix is shown with an intensity that corresponds to the edge weight (the darker the dot the larger the weight on the corresponding edge). These plots correspond to three different node labelings: the first one (top–left corner) is the original node labeling; the second plot (top–right corner) is the labeling, somewhat corresponding to a rich–club paradigm, obtained by re-ordering the nodes according to decreasing values of the overall degree $\boldsymbol{d}^{\mathrm{in}} + \boldsymbol{d}^{\mathrm{out}}$; the third one (bottom–left corner) is the labeling corresponding to decreasing values of the core–score computed with the NSM using parameters $\alpha = 10$ and $p = 20$. This latter figure clearly shows that the Enron email dataset contains a strong core–periphery structure, which was less prevalent initially. This is further confirmed by the core–periphery profile in the bottom–right plot, which shows the behavior of

$$\gamma(S_k) = \frac{\sum_{i,j \in S_k} A_{ij}}{\sum_{i \in S_k} d_i^{\mathrm{in}} + d_i^{\mathrm{out}}}$$

Figure 2 Adjacency sparsity plots and core–periphery profile of the Enron email dataset. The two panels in the top and the panel in the bottom-left corner show the nonzero entries of the adjacency matrix of the network, with different color intensities for different edge weights, when the nodes are re-labeled in three different ways: the top-left panel corresponds to the original node labeling; the top-right panel is the labeling obtained by re–ordering the nodes according to decreasing values of the overall degree $d^{\text{in}} + d^{\text{out}}$; the bottom-left is the labeling corresponding to decreasing values of the core–score computed with the proposed NSM. Finally, the bottom-right panel shows the persistence probability $\gamma(S_k)$ as a function of k, when S_k is the set of the k most peripheral nodes according to the degree vector (orange line) or the NSM (blue line).

against k, where S_k is the set of k most peripheral nodes corresponding to a core–score assignment. As k varies from 1 to $n = 87,273$, $\gamma(S_k)$ varies from 0 to 1 and measures the ratio of periphery–periphery links to periphery–all links, if S_k were to be chosen to be the periphery set. Thus a network has a strong core–periphery structure revealed by a core–score vector if the corresponding profile $\gamma(S_k)$ takes small values as k increases from zero and then grows dramatically as k crosses some threshold value.

For undirected networks, the profile $\gamma(S_k)$ was proposed in [23] as a means to visualize core-periphery structure. In this case, $\gamma(S_k)$ coincides with the persistence probability of the set S_k, i.e., the probability that a random walker who is currently in any of the nodes of S_k remains in S_k at the next time step. For directed strongly connected networks, the persistence probability of S_k would instead be given by $\sum_{ij \in S_k} y_i P_{ij} / \sum_{j \in S_k} y_j$, where $\boldsymbol{y}$ is the stationary distribution of the random walk with transition matrix $P_{ij} = A_{ij}/d_i^{\text{out}}$. However, as the Enron dataset we
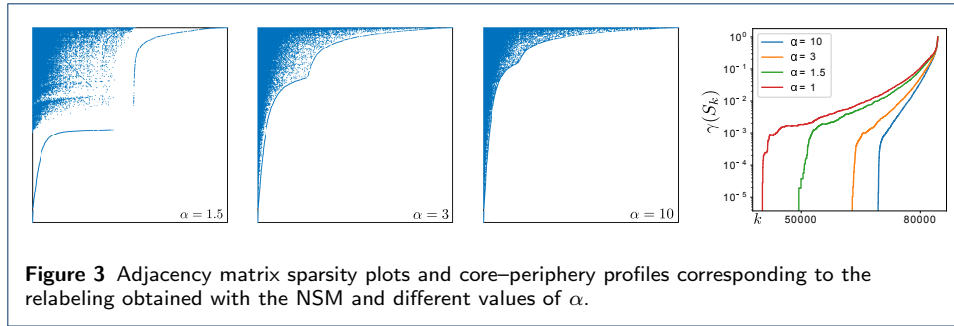
Figure 3 Adjacency matrix sparsity plots and core–periphery profiles corresponding to the relabeling obtained with the NSM and different values of α.

are considering is not connected, $\boldsymbol{y}$ is not well defined, and we compute $\gamma(S_k)$ in its place.

Finally, in order to show how the parameter α affects the core–periphery assignment obtained with NSM on this dataset, we show in Figure 3 the core–periphery structure and core-periphery profile using three different values of α, namely $\alpha \in \{1.5, 3, 10\}$. While changing the value of p does not effect the core–periphery structure output of the algorithm, small values of α show a weaker core–periphery structure, which is consistent with the fact that our model ideally works best when $\alpha \to \infty$. However, in practice $\alpha = 10$ performs well and have observed that larger values of α do not result in any significant change.

Since the network is not strongly connected, we do not show plots corresponding to the eigenvector centrality—this is not uniquely defined and, in our tests, different runs of Julia's `Arpack.eigs` gave rise to very different reorderings.

For this network the NSM with parameters $\alpha = 10$ and $p = 20$ computed the solution to 9 digits of precision in less than 5 seconds on a standard i7 single core laptop, using Julia 1.0. Our code in both Matlab and Julia is available online at the address https://github.com/ftudisco/nonlinear-core-periphery.

8 Conclusion

Our main aim in this work was to show that the attractive properties of the nonlinear spectral method proposed in [9] can almost completely be transferred to the directed, weighted and unconnected setting. In particular we show that for the core–periphery kernel quality function (1), proposed for example in [8, 9], there is always a unique solution for $\alpha \geq 0$ and $p > \max\{1, \alpha\}$, and this solution can be computed via a nonlinear spectral method whenever it is feasible to form matrix-vector products based on the network weight matrix. The proposed method, which exploits an intriguing connection between optimization and eigenproblems, generalizes the classical power method in order to compute the global maximum of a highly nonconvex function; thus it may also be of interest in other machine learning contexts.

Appendix. Theorem proofs

Proof of Theorem 1. First note that, by adding and removing $\sum_{ij \in E} \log(1 - p_{ij}(\boldsymbol{u}))$, the log likelihood (4) can be equivalently written as

$$L_\alpha(\boldsymbol{u}) = \sum_{ij \in E} \log\left(\frac{p_{ij}(\boldsymbol{u})}{1 - p_{ij}(\boldsymbol{u})}\right) + \sum_{i,j=1}^{n} \log(1 - p_{ij}(\boldsymbol{u})) =: S_1(\boldsymbol{u}) + S_2(\boldsymbol{u}).$$

Let us analyze the two terms S_1 and S_2 individually. If $a = \frac{1}{1+e^{-b}}$, we have

$$\frac{a}{1-a} = \left(\frac{1}{1+e^{-b}}\right)\left(\frac{e^{-b}}{1+e^{-b}}\right)^{-1} = e^b$$

which, from (3), implies that

$$S_1(\boldsymbol{u}) = \sum_{ij \in E} \log\left(\frac{p_{ij}(\boldsymbol{u})}{1 - p_{ij}(\boldsymbol{u})}\right) = \sum_{ij \in E} \kappa_\alpha(u_i, u_j) = \sum_{i,j=1}^n A_{ij}\kappa_\alpha(u_i, u_j),$$

i.e., $S_1(\boldsymbol{u}) = f_\alpha(\boldsymbol{u})$. Now note that, if $\boldsymbol{u}, \boldsymbol{v} \in \mathrm{CR}(n)$ then there exists a permutation σ of $\{1, \ldots, n\}$ such that $u_i = v_{\sigma(i)}$ for all i. Therefore,

$$S_2(\boldsymbol{u}) = \sum_{i,j=1}^n \log\left(\frac{1}{1 + e^{-\kappa_\alpha(u_i, u_j)}}\right) = \sum_{i,j=1}^n \log\left(\frac{1}{1 + e^{-\kappa_\alpha(v_{\sigma(i)}, v_{\sigma(j)})}}\right) = S_2(\boldsymbol{v}),$$

which implies that S_2 is constant on $\mathrm{CR}(n)$. Thus $\boldsymbol{u}$ maximizes $L_\alpha(\boldsymbol{u})$ if and only if it maximizes $S_1(\boldsymbol{u})$ and the proof is complete. $\qquad\square$

Proof of Theorem 2 and Corollary 3. This proof is based on the proof of Theorem 4.5 in [9] and the lemmas therein proved. For convenience, let us denote by $\mathbb{R}_+^n$ the cone of vectors with nonnegative entries. Since we are interested in a nonnegative maximizer of $f_\alpha(\boldsymbol{x})$ constrained on the sphere $\|\boldsymbol{x}\|_p = 1$, we can equivalently look for a maximizer of $f_\alpha(\boldsymbol{x}/\|\boldsymbol{x}\|_p)$ on the whole cone of nonnegative vectors $\mathbb{R}_+^n$. Now, notice that f_α is positively 1-homogeneous, that is $f_\alpha(a\boldsymbol{x}) = af_\alpha(\boldsymbol{x})$ holds for any real number $a \geq 0$. Therefore we can further change our problem into the global maximum on $\mathbb{R}_+^n$ of $g(\boldsymbol{x}) = f_\alpha(\boldsymbol{x})/\|\boldsymbol{x}\|_p$, without losing any generality. The critical point condition for g implies the equivalence with the eigenvalue problem (5), i.e., $\boldsymbol{x}$ is a stationary point for g if and only if it is such that $\Phi_\alpha(\boldsymbol{x}) = \lambda\boldsymbol{x}^{p-1}$. As $p > 1$, we can equivalently write $\widetilde{\Phi}(\boldsymbol{x}) = \mu\boldsymbol{x}$, with $\mu = \lambda^{\frac{1}{p-1}}$ and $\widetilde{\Phi}(\boldsymbol{x}) = \Phi_\alpha(\boldsymbol{x})^{\frac{1}{p-1}}$. We now show that there can only be one nonnegative $\boldsymbol{x}$ such that $\|\boldsymbol{x}\|_p = 1$ and $\widetilde{\Phi}(\boldsymbol{x}) = \mu\boldsymbol{x}$.

To this end, note that $\widetilde{\Phi}(\boldsymbol{x}) \geq 0$ for any $\boldsymbol{x} \geq 0$. Thus if $\boldsymbol{x} \geq 0$ and $\|\boldsymbol{x}\|_p = 1$, then $\mu > 0$ and we have

$$\mu = \|\mu\boldsymbol{x}\|_p = \|\widetilde{\Phi}(\boldsymbol{x})\|_p = \|\Phi_\alpha(\boldsymbol{x})\|_q^{q-1},$$

where q is such that $1/p + 1/q = 1$. Therefore any $\boldsymbol{x} \geq 0$, $\|\boldsymbol{x}\|_p = 1$ solution of (5) is a fixed point of the map

$$\Psi : \mathbb{R}_+^n \to \mathbb{R}_+^n, \qquad \Psi(\boldsymbol{x}) = \frac{\widetilde{\Phi}(\boldsymbol{x})}{\|\widetilde{\Phi}(\boldsymbol{x})\|_p} = \frac{\Phi_\alpha(\boldsymbol{x})^{q-1}}{\|\Phi_\alpha(\boldsymbol{x})\|_q^{q-1}}.$$

Let $\tau(\boldsymbol{x}, \boldsymbol{y}) = \|\ln\boldsymbol{x} - \ln\boldsymbol{y}\|_\infty$. Lemma 4.4 of [9] implies that

$$\frac{\tau(\Psi(\boldsymbol{x}), \Psi(\boldsymbol{y}))}{\tau(\boldsymbol{x}, \boldsymbol{y})} \leq \frac{|\alpha - 1|}{p - 1},$$

for any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}_+^n$. As τ is a complete metric on the cone $\mathbb{R}_+^n$ (see, for example, [24]), this shows that Ψ is a contraction and thus it has a unique fixed point.

We conclude that, when $\alpha > 0$ and $p > \max\{1, \alpha\}$, the eigenvalue equation (5) has a unique nonnegative solution $\boldsymbol{u} \geq 0$ such that $\|\boldsymbol{u}\|_p = 1$, which is also the unique solution of (1) when we choose $\|\cdot\| = \|\cdot\|_p$. Note moreover that if we start with a positive $\boldsymbol{u}_0 > 0$ and apply Ψ iteratively to give $\boldsymbol{u}_{k+1} = \Psi(\boldsymbol{u}_k)$ we obtain $\tau(\boldsymbol{u}_{k+1}, \boldsymbol{u}_k) \leq (\frac{|\alpha-1|}{p-1})^k \tau(\boldsymbol{u}_1, \boldsymbol{u}_0)$. This, together with the inequality $\tau(\boldsymbol{u}_1, \boldsymbol{u}_0) \leq \gamma$, proved for example in Corollary 4.6 of [9], completes the proof of Corollary 3.

Next we prove that $\boldsymbol{u}$ has a zero component i if and only if i is an isolated node, i.e., it has no incoming nor outgoing links. To this end, let $\Omega_A^+ \subseteq \mathbb{R}_+^n$ be the set of vectors

$$\Omega_A^+ = \{\boldsymbol{x} \geq 0 : x_i = 0 \text{ if and only if } i \text{ is isolated}\}.$$

Note that, equivalently, $x_i = 0$ for $\boldsymbol{x} \in \Omega_A^+$ if and only if $A_{ij} + A_{ji} = 0$ for all $j = 1, \ldots, n$. Now note that if $\boldsymbol{x} \in \Omega_A^+$ then $\Phi_\alpha(\boldsymbol{x}) \in \Omega_A^+$. In fact, from its definition

$$\Phi_\alpha(\boldsymbol{x})_i = x_i^{\alpha-1} \sum_{j=1}^n \frac{A_{ij} + A_{ji}}{\kappa_\alpha(x_i, x_j)^{\alpha-1}}$$

we see that, if i is isolated, $x_i = 0$ and thus $\Phi_\alpha(\boldsymbol{x})_i = 0$, whereas if $x_i > 0$, then $\Phi_\alpha(\boldsymbol{x})_i > 0$ as $\kappa_\alpha(x_i, x_j) \geq x_i > 0$ and there exists at least one j^* such that $A_{i,j^*} + A_{j^*,i} > 0$, which implies $\Phi_\alpha(\boldsymbol{x})_i \geq x_i^{\alpha-1}(A_{i,j^*} + A_{j^*,i})\kappa_\alpha(x_i, x_{j^*})^{1-\alpha} > 0$. Note that the same conclusion holds for any initial positive vector; that is, $\boldsymbol{x} > 0$ implies $\Phi_\alpha(\boldsymbol{x}) \in \Omega_A^+$. Therefore the iterative method of Corollary 3 converges to a vector in Ω_A^+ for any starting point, or, equivalently, any nonnegative solution of (5) must be in Ω_A^+. Since there exists only one such solution, the proof is complete. $\square$

Declarations

References

1. Bertozzi, A.L., Luo, X., Stuart, A.M., Zygalakis, K..C.: Uncertainty quantification in the classification of high dimensional data. SIAM-ASA J. Uncertainty Quantification **6**, 568–595 (2018). doi:10.1137/17M1134214
2. Rombach, M.P., Porter, M.A., Fowler, J.H., Mucha, P.J.: Core-periphery structure in networks. SIAM J. Applied Math. **74**, 167–190 (2014)
3. Borgatti, S.P., Everett, M.G.: Models of core/periphery structures. Social Networks **21**, 375–395 (2000)
4. Zhou, S., Mondragón, R.J.: The rich-club phenomenon in the internet topology. IEEE Communications Letters **8**(3), 180–182 (2004)
5. Bascompte, J., Jordano, P., Melián, C.J., Olesen, J.M.: The nested assembly of plant–animal mutualistic networks. Proceedings of the National Academy of Sciences **100**(16), 9383–9387 (2003)
6. Schneider, C.M., Moreira, A.A., Andrade, J.S., Havlin, S., Herrmann, H.J.: Mitigation of malicious attacks on networks. Proceedings of the National Academy of Sciences **108**(10), 3838–3841 (2011)
7. Cucuringu, M., Rombach, P., Lee, S.H., Porter, M.A.: Detection of core–periphery structure in networks using spectral methods and geodesic paths. European Journal of Applied Mathematics **27**(6), 846–887 (2016)
8. Rombach, M.P., Porter, M.A., Fowler, J.H., Mucha, P.J.: Core-periphery structure in networks (revisited). SIAM Review **59**, 619–646 (2017)
9. Tudisco, F., Higham, D.J.: A nonlinear spectral method for core-periphery detection in networks. SIAM. J. Math. Data Science **269**–292 (2019)
10. Kim, P.M., Korbel, J.O., Gerstein, M.B.: Positive selection at the protein network periphery: evaluation in terms of structural constraints and cellular context. Proceedings of the National Academy of Sciences **104**(51), 20274–20279 (2007)
11. Sandhu, K.S., Li, G., Poh, H.M., Quek, Y.L.K., Sia, Y.Y., Peh, S.Q., Mulawadi, F.H., Lim, J., Sikic, M., Menghi, F., *et al.*: Large-scale functional organization of long-range chromatin interaction networks. Cell reports **2**(5), 1207–1219 (2012)
12. Baños, R., Borge-Holthoefer, J., Wang, N., Moreno, Y., González-Bailón, S.: Diffusion dynamics with changing network composition. Entropy **15**(11), 4553–4568 (2013)
13. Tomasello, M.V., Napoletano, M., Garas, A., Schweitzer, F.: The rise and fall of R & D networks. Industrial and corporate change **26**(4), 617–646 (2017)
14. Csermely, P., London, A., Wu, L.-Y., Uzzi, B.: Structure and dynamics of core/periphery networks. Journal of Complex Networks **1**(2), 93–123 (2013)
15. Newman, M.E.J.: Networks: an Introduction. Oxford University Press, (2011)
16. Mondragón, R.J.: Network partition via a bound of the spectral radius. Journal of Complex Networks **5**(4), 513–526 (2016)
17. Horn, R.A., Johnson, C.R.: Matrix Analysis. Cambridge University Press, (1990)
18. O'Connor, L., Médard, M., Feizi, S.: Maximum likelihood latent space embedding of logistic random dot product graphs. arXiv preprint arXiv:1510.00850 (2015)
19. Hoff, P.D., Raftery, A.E., Handcock, M.S.: Latent space approaches to social network analysis. Journal of the american Statistical association **97**(460), 1090–1098 (2002)
20. Jia, J., Benson, A.R.: Random spatial network models with core-periphery structure. In: Proc. ACM International Conf. on Web Search and Data Mining (WSDM) (2018)
21. Newman, M.E.: Equivalence between modularity optimization and maximum likelihood methods for community detection. Physical Review E **94**(5), 052315 (2016)
22. Klimt, B., Yang, Y.: The Enron corpus: A new dataset for email classification research. In: Proc. European Conf. on Machine Learning, pp. 217–226 (2004)
23. Della Rossa, F., Dercole, F., Piccardi, C.: Profiling core-periphery network structure by random walkers. Scientific Reports **3**, 1467 (2013)
24. Lemmens, B., Nussbaum, R.D.: Nonlinear Perron-Frobenius Theory. Cambridge University Press, (2012)