Nonlinear Eigenvector Centrality for Networks

Francesco Tudisco[1,2] and Desmond J. Higham[1,a)]

[1]*School of Mathematics, University of Edinburgh, Edinburgh, EH9 3FD, UK*
[2]*School of Mathematics, Gran Sasso Science Institute, L'Aquila, 67100, Italy*

[a)]Corresponding author: d.j.higham@ed.ac.uk

Abstract. We develop, analyze and test a family of nonlinear generalisations to classical eigenvector network centrality. They are computable via a convergent nonlinear power method iteration, making them efficient for large, sparse networks. The nonlinearities may be chosen to meet domain-specific modeling requirements. Using transport data from the city of London, we show that appropriate nonlinearities improve predictions of passenger usage.

Motivation

Recent work has shown that classic Perron-Frobenius theory can be extended to nonlinear settings. This motivates our aim to define a collection of network centrality measures that build on eigenvector centrality and can be computed efficiently with nonlinear power iterations. We believe that the extra modeling flexibility arising from nonlinear iterative schemes has very wide applicability; we consider here the case of transport network data.

Set-up

Let $G = (V, E)$ be an undirected network with nodes $V = \{1, \dots, n\}$ and adjacency matrix A. We consider weighted networks, with $A_{ij} \geq 0$ measuring the strength of the pairwise relationship between nodes i and j. A centrality vector for G is a vector $\boldsymbol{u} \geq 0$ such that u_i quantifies the importance of the node i, according to a suitable model that exploits only the structure of the connections in G. A larger value indicates greater importance. We impose two natural assumptions:.

1. $\mathbb{1}^T \boldsymbol{u} = \|\boldsymbol{u}\|_1 = u_1 + \cdots + u_n = 1$, so that each u_i can be interpreted as a proportion of importance,
2. $\boldsymbol{u}$ is isomorphism invariant, i.e., the importance of node i does not change if we relabel the network nodes.

It is convenient to note at this stage that 2 has the following algebraic formulation: let $P \in \mathbb{R}^{n \times n}$ be any permutation matrix. If $\boldsymbol{u}$ is the centrality vector for the network with adjacency matrix A, then 2 holds if and only if $P\boldsymbol{u}$ is the centrality vector of the network with adjacency matrix PAP^T.

One widely used centrality model, called *eigenvector centrality* or *Bonacich centrality*, defines $\boldsymbol{u}$ as a Perron eigenvector of A, that is, a nonnegative $\boldsymbol{u}$ such that $A\boldsymbol{u} = \lambda \boldsymbol{u}$, with $\lambda > 0$ and $\mathbb{1}^T \boldsymbol{u} = 1$. We refer to [Newman(2010)] for a discussion of the ideas in the social science literature that led to the development of this centrality measure, and note that the concept can be traced back to algorithms used in the 19th century for ranking chess players, [Schäfermeyer(2019)]. The *linear* Perron–Frobenius theorem guarantees that such a vector exists and that it is uniquely defined if, for example, the graph G is connected. Moreover, recalling that $P^T P = PP^T = I$ for any permutation matrix, it easy to verify that 2 holds since $(PAP^T)(P\boldsymbol{u}) = \lambda (P\boldsymbol{u})$, for any permutation matrix P.

Nonlinear eigenvector centralities

By inspecting the entrywise relation defining the eigenvector centrality $\boldsymbol{u}$, we see that this model defines u_i to be proportional to the weighted sum of the centrality scores of the neighbours of i, that is,

$$u_i \propto \sum_j A_{ij} u_j = (A\boldsymbol{u})_i. \tag{1}$$

In this work, our aim is to generalize the linear relation (1) between the importance of i and the importance of its neighbourhood by means of *nonlinear* eigenvector centralities. This is a new concept that builds on modern nonlinear Perron–Frobenius theory [Lemmens and Nussbaum(2012), Gautier, Tudisco, and Hein(2019), Tudisco and Higham(2019), Sittoni and Tudisco(2024)] and recent extensions of classical eigenvector centrality scores to multilayer and higher-order networks [Arrigo and Tudisco(2019), Tudisco, Arrigo, and Gautier(2018), Tudisco and Higham(2021)].

Consider a mapping $f : \mathbb{R}^n \to \mathbb{R}^n$ such that $f(P\boldsymbol{x}) = Pf(\boldsymbol{x})$ for any $\boldsymbol{x} \in \mathbb{R}^n$ and any permutation matrix $P \in \mathbb{R}^{n\times n}$. We motivate a nonlinear eigenvector centrality for the node i via the proportional relation

$$u_i \propto g(\sum_j A_{ij}\, f(\boldsymbol{u})_j) = g(A\, f(\boldsymbol{u}))_i,$$

where $g : \mathbb{R} \to \mathbb{R}$ is a real valued function. Note that this relation is equivalent to the *nonlinear* eigenvector equation

$$g(A\, f(\boldsymbol{u})) = \lambda\, \boldsymbol{u}, \tag{2}$$

where we use notation that extends g to vectors in a componentwise manner, that is, $g(\boldsymbol{x})_i = g(x_i)$. Observe that this relation automatically implies property 2 for the nonlinear eigenvector $\boldsymbol{u}$. In fact, if (2) holds, we have

$$\lambda\, (P\boldsymbol{u}) = Pg(A\, f(\boldsymbol{u})) = g(PAP^T P\, f(\boldsymbol{u})) = g((PAP^T)\, f(P\boldsymbol{u}))\,.$$

Similarly to the linear case where both f and g are the identity map, if f and g are homogeneous or multihomogeneous, then the nonlinear Perron–Frobenius theorem [Lemmens and Nussbaum(2012), Gautier, Tudisco, and Hein(2019)] provides conditions that ensure existence and uniqueness of a nonnegative $\boldsymbol{u}$. However, for the transportation network application we have in mind, we are not necessarily interested in homogeneous mappings. For this reason, we propose below a new Perron–Frobenius type result that only relies on the ratio between the map and its derivative.

For a vector $\boldsymbol{y}$ and a matrix M let us denote by $|\boldsymbol{y}|$ and $|M|$ the entrywise absolute value, i.e., the vector (resp. matrix) with components $(|\boldsymbol{y}|)_i = |y_i|$ (resp. $(|M|)_{ij} = |M_{ij}|$). Further, let us say that a map $f : \mathbb{R}^n \to \mathbb{R}^n$ is positive if $f(\boldsymbol{x}) > 0$ for all $\boldsymbol{x} > 0$, where inequalities are interpreted componentwise.

Theorem 1. *Let $f = (f_1, \dots, f_n)^T : \mathbb{R}^n \to \mathbb{R}^n$ be a positive differentiable mapping and let $J_f(\boldsymbol{x}) \in \mathbb{R}^{n\times n}$ be its Jacobian matrix at a point $\boldsymbol{x}$, entrywise defined by $J_f(\boldsymbol{x})_{ij} = \partial_j f_i(\boldsymbol{x})$. Let $\varepsilon > 0$ and let $g : \mathbb{R}^n \to \mathbb{R}^n$ be the componentwise power $g(\boldsymbol{x}) = \boldsymbol{x}^{1/\varepsilon}$. Define*

$$\kappa_+(A, f) = \max_{\boldsymbol{x}>0} \left\| \frac{A\,|J_f(\boldsymbol{x})|\,\boldsymbol{x}}{Af(\boldsymbol{x})} \right\|_\infty ,$$

where $\boldsymbol{x}/\boldsymbol{y}$ denotes componentwise division, that is, $(\boldsymbol{x}/\boldsymbol{y})_i = x_i/y_i$.

If $\kappa_+(A, f) < \varepsilon/2$ then the following properties hold:

1. *There exist unique $\boldsymbol{u}^* > 0$ and $\lambda > 0$ such that $\mathbb{1}^T\boldsymbol{u}^* = u_1^* + \cdots + u_n^* = 1$ and $g(A\, f(\boldsymbol{u}^*)) = \lambda\,\boldsymbol{u}^*$.*
2. *The sequence $\{\boldsymbol{u}_k\}_k$ generated by the following nonlinear power method*

$$\begin{cases} \boldsymbol{y}_k = g(A\, f(\boldsymbol{u}_k)) \\ \boldsymbol{u}_{k+1} = \boldsymbol{y}_k/(\mathbb{1}^T\boldsymbol{y}_k) \end{cases}, \qquad k = 0, 1, 2, 3, \dots \tag{3}$$

converges to $\boldsymbol{u}^*$, for any starting point $\boldsymbol{u}_0 > 0$.

We postpone the proof of the theorem to the final section. We continue by listing several important remarks.

1. If the map f is positively subhomogeneous we can give a simple upper bound for the ratio $\kappa_+(A, f)$ in terms of its degree of subhomogeneity. To this end, first recall that, by Euler's homogeneous function theorem, a differentiable $\phi : \mathbb{R}^n \to \mathbb{R}$ is positively homogeneous of degree $\alpha > 0$ if and only if $\nabla\phi(\boldsymbol{x})^T\boldsymbol{x} = \alpha\,\phi(\boldsymbol{x})$ for all positive vectors $\boldsymbol{x} > 0$. Therefore, we say that $f = (f_1, \dots, f_n)^T : \mathbb{R}^n \to \mathbb{R}^n$ is positively subhomogeneous of degrees $(\alpha_1, \dots, \alpha_n)$ if each $f_i : \mathbb{R}^n \to \mathbb{R}$ is such that $|\nabla f_i(\boldsymbol{x})|^T\boldsymbol{x} \le \alpha_i f_i(\boldsymbol{x})$, for all $\boldsymbol{x} > 0$, see also [Sittoni and Tudisco(2024)]. For any such f we get $(|J_f(\boldsymbol{x})|\boldsymbol{x})_i = |\nabla f_i(\boldsymbol{x})|^T\boldsymbol{x} \le \alpha_i f_i(\boldsymbol{x})$. Thus $(A|J_f(\boldsymbol{x})|\boldsymbol{x})_i \le (\max_k \alpha_k)(Af(\boldsymbol{x}))_i$ and we therefore obtain the following upper bound for $\kappa_+(A, f)$ in terms of the subhomogeneity degrees

$$\kappa_+(A, f) = \max_{\boldsymbol{x}>0}\max_{i=1,\dots,n}\frac{(A|J_f(\boldsymbol{x})|\boldsymbol{x})_i}{(Af(\boldsymbol{x}))_i} \le \max_k \alpha_k.$$

2. The differentiability requirement on f can be relaxed with almost no additional complication. It is enough to replace $J_f(\boldsymbol{x})$ in the definition of $\kappa_+(A, f)$ with any point in the subdifferential of f. More precisely, if $\partial_f(\boldsymbol{x})$ denotes the subdifferential of f at $\boldsymbol{x}$, then Theorem 1 holds unchanged if we define the ratio $\kappa_+(A, f)$ for a nondifferentiable f as

$$\kappa_+(A, f) = \max_{\boldsymbol{x}>0}\max_{M_{\boldsymbol{x}}\in\partial_f(\boldsymbol{x})}\left\|\frac{A\,|M_{\boldsymbol{x}}|\,\boldsymbol{x}}{Af(\boldsymbol{x})}\right\|_\infty. \tag{4}$$

3. The nonlinear power method in Theorem 1 is a natural generalization of the standard power method used to compute the Perron eigenvector of the adjacency matrix of the network, which is retrieved when both f and g are the identity map.

4. For typical choices of f, the computational cost of each iteration in (3) will be dominated by the cost of applying the adjacency matrix to a vector in forming $A\,f(\boldsymbol{u}_k)$. Hence the iteration will be feasible for large sparse networks.

In the next section we show that the novel and flexible centrality measure defined by (2) offers value in the transport setting.

London train stations usage

Tools from network science have proved to be valuable in the study of urban transport [Batty(2013)]. Here, we consider the use of network centrality measures in the case of the London train network. Nodes in the network represent stations, and we seek a centrality measure that correlates strongly with passenger usage. Such a measure, which requires only the topological connectivity structure, offers helpful information at the design stage. More importantly, it can be used in what-if-scenario testing, in order to predict the effect of changes, including unplanned network disruptions.

Algorithm design

We focus our analysis on the London train network [Domenico *et al.*(2014)Domenico, Solé-Ribalta, Gómez, and Arenas]. This is an undirected transportation network representing connections between tube train stations of the city of London. The graph we consider here is the aggregated version of the original multilayer network that refers to the London train network structure in 2016, restricted to the stations belonging to the underground rail network. It consists of one connected component with 271 nodes and 315 edges with nonzero weights. Each such weight takes an integer value of 1, 2, or 3 according to the number of different types of connection, from the three possibilities of underground, overground and Docklands Light Railway (DLR). The network data we used was collected from `https://manliodedomenico.com/data.php`

We collect additional passenger data from [Cipolla, Durastante, and Tudisco(2021)]. For each train station of the above network, we collect the number of passengers entering or exiting that station per year. We collect data for ten years: from 2008 to 2017. Our data and code are publicly available at `https://github.com/ftudisco/nlc`.

We then study the behavior of different nonlinear eigenvector centralities on this network data, motivated by the question of whether we can identify highly populated stations by exploiting only the topology of connections between stations.

To this end, we posit that passengers tend to use one station over another if

1. the station is well connected to important stations—as it is likely that a passenger origin or destination is an important (highly used) station, rather than a minor one,

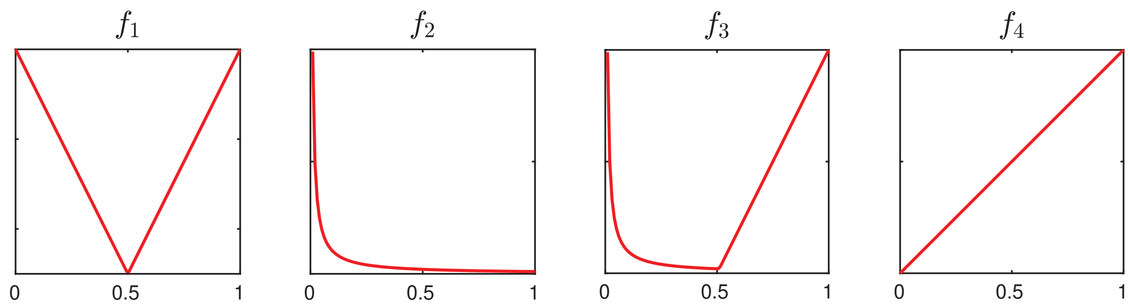

FIGURE 1. Visualization of the test functions $f_1, \ldots, f_4$ on a single variable $x \in \mathbb{R}$

2. the station is surrounded by stations of minor relevance (popularity)—in this way passengers are somewhat instinctively motivated to use that station rather than a less popular station in its neighborhood.

So, if u_i is a measure accounting for the popularity (in terms of passenger usage) of a station, then u_i should grow if either many popular stations or many unpopular stations are in the neighborhood of i.

To account for either or both of points 1 and 2, we consider the following four choices for the function f:

- $f_1(\boldsymbol{x}) = |\boldsymbol{x} - 1/2| + 1/2$
- $f_2(\boldsymbol{x}) = 1/(100\boldsymbol{x})$
- $f_3(\boldsymbol{x}) = \max\{1/(100\boldsymbol{x}), 2\boldsymbol{x} - 1\}$
- $f_4(\boldsymbol{x}) = \boldsymbol{x}$

and we fix $g(\boldsymbol{x}) = \boldsymbol{x}^{1/\varepsilon}$, with $\varepsilon > 0$. Figure 1 shows the behavior of each f_i on a single variable in $[0, 1]$. Here, to allow the overall shapes to be compared visually we have removed the scaling information from the vertical axes. We also note that the factors of 100 in f_2 and f_3 are arbitrary, in the sense that the resulting node rankings would be the same if any other positive value were used.

Note that Theorem 1 applies immediately to f_2, and f_4, both of which are differentiable on the set of positive vectors. It is not difficult to observe that $\kappa_+(A, f_2) = 1$. In fact, the Jacobian of f_2 is the diagonal matrix

$$J_{f_2}(\boldsymbol{x}) = -\frac{1}{100}\begin{pmatrix} x_1^{-2} & & \\ & \ddots & \\ & & x_n^{-2} \end{pmatrix}$$

and thus $|J_{f_2}(\boldsymbol{x})|\boldsymbol{x} = f_2(\boldsymbol{x})$. The same property holds for f_4.

Unlike f_2 and f_4, the functions f_1 and f_3 are not differentiable. However, a similar argument using the definition in equation (4), shows that $\kappa_+(A, f_3) = 1$, as for any $M_{\boldsymbol{x}} \in \partial_{f_3}(\boldsymbol{x})$ we have $|M_{\boldsymbol{x}}|\boldsymbol{x} = f_3(\boldsymbol{x})$ and thus $\frac{A\,|M_{\boldsymbol{x}}|\,\boldsymbol{x}}{A f_3(\boldsymbol{x})} = 1$. Finally, concerning $\kappa_+(A, f_1)$, one can easily observe that $|M_{\boldsymbol{x}}|\boldsymbol{x} = \boldsymbol{x}$, for any $M_{\boldsymbol{x}} \in \partial_{f_1}(\boldsymbol{x})$ and thus

$$\begin{aligned}
\kappa_+(A, f_1) &= \max_{\boldsymbol{x}>0} \max_{i=1,\ldots,n} \frac{\sum_j A_{ij}x_j}{\sum_j A_{ij}|x_j - 1/2| + 1/2\sum_j A_{ij}} \\
&\leq \max_{\boldsymbol{x}>0} \max_{i,j=1,\ldots,n} \frac{A_{ij}x_j}{A_{ij}|x_j - 1/2| + 1/2A_{ij}} \\
&= \max_{\boldsymbol{x}>0} \max_{i,j=1,\ldots,n} \frac{x_j}{|x_j - 1/2| + 1/2} \leq 1,
\end{aligned}$$

where the inequality follows by the well known inequality $\sum_i a_i / \sum_i b_i \leq \max_i a_i/b_i$, that holds for any set of nonnegative numbers a_i, b_i.

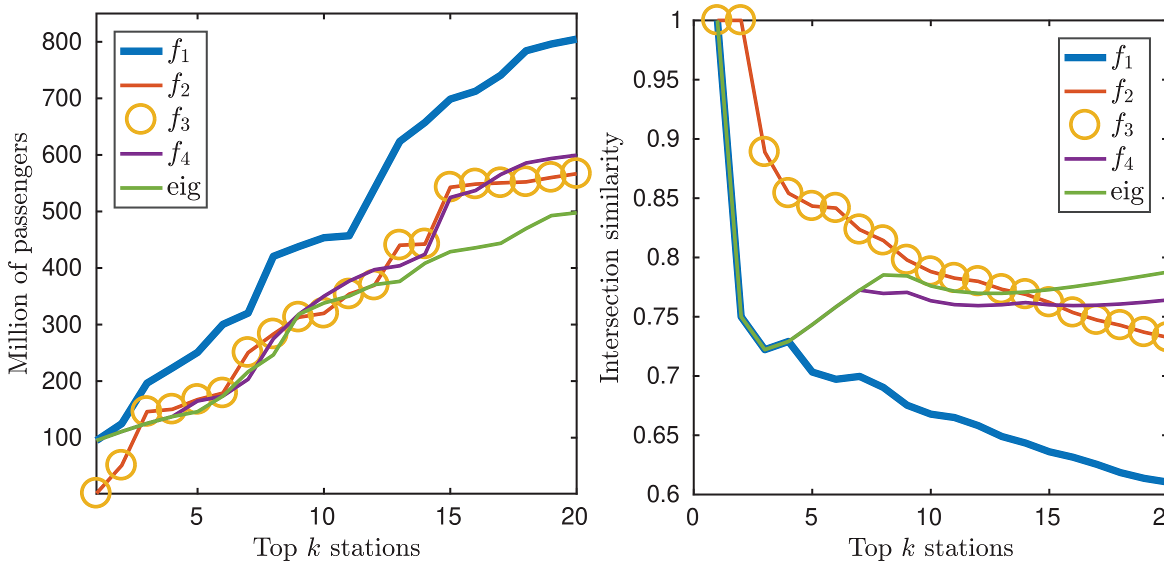

FIGURE 2. Left: Cumulative sum of the number of passengers that in 2016 used the top 20 train stations identified by the (linear) eigenvector centrality "eig" and the nonlinear eigenvector centrality (2) associated with $f_1, \ldots, f_4$. Right: Intersection similarity between the sequence of stations sorted according to the actual number of passengers (in 2016) and the sequence of highest ranked stations according to the five centrality models.

Computational results

Our aim is now to test the extent to which the network centrality measures developed above can identify nodes that perform well in terms of total passenger usage.

The left panel of Figure 2 shows the cumulative sum of the number of passengers using the top twenty stations identified by the standard (linear) eigenvector centrality and the nonlinear centrality model $(Af(\boldsymbol{u}))^{1/\varepsilon} = \lambda \boldsymbol{u}$, for the four choices of f described above. For these experiments we chose $\varepsilon = 2.1 \geq 2\kappa_+(A, f_i) + 0.1$, for $i = 1, \ldots, 4$.

The right panel of Figure 2 compares the top twenty stations identified by the same 5 models with the "ground truth", i.e., the actual twenty stations with the highest number of passengers. This comparison is made by means of the intersection similarity between the sequence of top ranked stations according to each centrality and the sequence of most populated stations. For completeness, we recall that the intersection similarity is a measure to compare two ranked lists that may not contain the same elements. It is defined as follows [Fagin, Kumar, and Sivakumar(2003)]: let $\boldsymbol{r}$ the ranking corresponding to one of the centrality models and let $\boldsymbol{\rho}$ be the ground truth ranking. The intersection similarity of $\boldsymbol{r}$ and $\boldsymbol{\rho}$ is the vector ISIM with entries

$$\mathrm{ISIM}_k = \frac{1}{k} \sum_{j=1}^{k} \frac{\Big|\Delta\Big((r_1, \ldots, r_j), (\rho_1, \ldots, \rho_j)\Big)\Big|}{2j},$$

where $|S|$ denotes the cardinality of the set S and Δ is the symmetric difference operator. When the first k entries in $\boldsymbol{r}$ and $\boldsymbol{\rho}$ are completely different ISIM_k is equal to 1, whereas $\mathrm{ISIM}_k = 0$ if and only if the first k entries in $\boldsymbol{r}$ and $\boldsymbol{\rho}$ coincide. Hence, a lower value indicates better performance.

The data illustrated in Figure 2 refers to passenger usages during 2016 . This analysis supports the modeling based on points 1 and 2 above, as $f_1 = |x - \frac{1}{2}| + \frac{1}{2}$ identifies roughly 40% more passengers than other models. We focus on 2016 because the network data we are using refers to the London train network structure from that year. However, a similar behavior is observed across a ten–year span usage data. This is seen in Table 1 where we compare the number of passengers using the top $k = 5, 10, 15, 20$ stations identified by the linear eigenvector centrality and the four nonlinear eigenvector centrality models, for each of the ten years $2008, 2009, \ldots, 2017$.

It is also worth noting that the centrality ranking obtained with models corresponding to f_2 and f_3 always coincide, even though f_2 and f_3 significantly differ in the interval $[1/2, 1]$. This phenomenon may be explained by the fact

TABLE 1. Millions of passengers per year using the top $k = 5, 10, 15, 20$ stations ranked according to the nonlinear centrality model (2), for the four choices of functions $f_1, \ldots, f_4$, and the standard eigenvector centrality, across the 10 years range 2017–2008. Here largest is best. Largest value for each test is shown in bold.

	k	2017	2016	2015	2014	2013	2012	2011	2010	2009	2008
f_1	5	**243.4**	**250.8**	**243.7**	**243.4**	**222.4**	**213.4**	**208.8**	**202.8**	**193.7**	**198.9**
	10	**435.6**	**453.6**	**441.9**	**439.2**	**414.5**	**400.3**	**391.7**	**380.7**	**362.3**	**367.4**
	15	**676.7**	**698.6**	**694.4**	**704.2**	**656.0**	**633.5**	**616.8**	**593.3**	**572.0**	**579.9**
	20	**779.1**	**804.7**	**796.0**	**803.4**	**750.5**	**724.6**	**703.5**	**677.2**	**652.1**	**657.6**
f_2 & f_3	5	169.1	167.4	164.7	163.7	156.6	149.1	145.3	138.9	129.9	129.0
	10	315.8	319.8	313.1	313.7	294.4	281.5	274.0	264.7	252.6	257.5
	15	526.5	542.7	539.1	534.1	504.1	484.3	471.5	453.4	435.3	440.2
	20	551.2	566.7	563.7	555.3	529.0	508.5	496.2	477.7	460.1	465.1
f_4	5	170.9	164.8	169.4	167.3	150.8	143.2	137.5	129.3	123.0	128.7
	10	349.5	350.1	351.3	349.0	316.9	303.9	294.5	280.4	267.9	274.0
	15	509.3	524.7	514.9	506.1	471.1	453.4	437.8	419.5	402.5	410.3
	20	581.0	599.4	585.9	576.8	538.6	517.6	501.1	478.9	461.0	467.8
eig	5	152.1	145.9	151.7	149.8	137.0	129.9	123.4	115.1	108.7	113.5
	10	337.8	338.1	342.3	342.1	311.4	298.6	288.4	276.5	265.6	270.4
	15	426.0	428.8	428.9	425.0	390.4	374.5	364.3	345.1	333.4	338.5
	20	496.0	497.7	495.2	491.0	454.5	435.5	422.8	402.7	389.9	393.9

that $f_3(x_i)$ is unbounded near zero and grows quickly when x_i is smaller than $1/2$ and thus the contribution of small values of x_i dominates the overall centrality model.

Our overall conclusion from these tests is that, in terms of predicting passenger usage, standard *linear* eigenvector centrality can be improved by the introduction of customized, and rigorously analyzed nonlinear measures, with f in (3) taking the form of f_1 from Figure 1 proving particularly effective.

Discussion

We have shown that the widely used concept of eigenvector centrality may be viewed as a special case of a general iteration scheme (3) involving two nonlinear functions. Moreover, the nonlinearities may be chosen to meet application–specific modeling requirements without sacrificing uniqueness or positivity and without affecting the dominant “adjacency matrix times vector” computational cost of each iteration. For the case of passenger usage in an urban transport system, we showed that customized nonlinearities can offer significant benefits. In future work we plan to study connections between this approach and (a) random graph models and (b) network reordering techniques.

Proof of Theorem 1

The proof of Theorem 1 is inspired by the nonlinear Perron–Frobenius type theorems developed in [Gautier, Tudisco, and Hein(2019), Tudisco and Higham(2019), Gautier, Nguyen, and Hein(2016)] and requires a number of preliminary results that we prove first.

Let $\nabla\phi : \mathbb{R}^n \to \mathbb{R}^n$ denote the gradient of a differentiable function $\phi : \mathbb{R}^n \to \mathbb{R}$ and for two vectors $\boldsymbol{x}, \boldsymbol{y}$, let $\boldsymbol{xy}$ denote componentwise multiplication, so $(\boldsymbol{xy})_i = x_i y_i$.

Lemma 2. *Let $\phi : \mathbb{R}^n \to \mathbb{R}^n$ be a differentiable function and, for $\boldsymbol{x} \in \mathbb{R}^n$, let $\varphi(\boldsymbol{x}) = \phi(\boldsymbol{x})/\|\phi(\boldsymbol{x})\|_1$. Then*

$$\sum_{k=1}^{n} \left| \frac{\partial_k \varphi_i(\boldsymbol{x}) x_k}{\varphi_i(\boldsymbol{x})} \right| \le 2 \max_{i=1,\ldots,n} \frac{|\nabla\phi_i(\boldsymbol{x})|^T \boldsymbol{x}}{|\phi_i(\boldsymbol{x})|},$$

for any $\boldsymbol{x} > 0$ and all $i, k = 1, \ldots, n$.

Proof. From the chain rule we have

$$\frac{\partial_k \varphi_i(\boldsymbol{x})}{\varphi_i(\boldsymbol{x})} = \frac{\|\phi(\boldsymbol{x})\|_1 \, \partial_k \phi_i(\boldsymbol{x}) - \phi_i(\boldsymbol{x})\{\partial_k|\phi_1(\boldsymbol{x})| + \cdots + \partial_k|\phi_n(\boldsymbol{x})|\}}{\|\phi(\boldsymbol{x})\|_1 \, \phi_i(\boldsymbol{x})}$$
$$= \frac{\partial_k \phi_i(\boldsymbol{x})}{\phi_s(\boldsymbol{x})} - \sum_{s=1}^n \frac{|\phi_i(\boldsymbol{x})|}{\|\phi(\boldsymbol{x})\|_1} \frac{\partial_k|\phi_s(\boldsymbol{x})|}{|\phi_s(\boldsymbol{x})|}.$$

Multiplying the latter formula by x_k and summing over k gives us

$$\sum_{k=1}^n \left| \frac{\partial_k \varphi_i(\boldsymbol{x}) x_k}{\varphi_i(\boldsymbol{x})} \right| \le \sum_{k=1}^n \left| \frac{\partial_k \phi_i(\boldsymbol{x}) x_k}{\phi_i(\boldsymbol{x})} \right| + \sum_{s=1}^n \frac{|\phi_s(\boldsymbol{x})|}{\|\phi(\boldsymbol{x})\|_1} \sum_{k=1}^n \frac{\partial_k|\phi_s(\boldsymbol{x})| x_k}{|\phi_s(\boldsymbol{x})|}$$
$$= \frac{|\nabla \phi_i(\boldsymbol{x})|^T \boldsymbol{x}}{|\phi_i(\boldsymbol{x})|} + \sum_{s=1}^n \frac{|\phi_s(\boldsymbol{x})|}{\|\phi(\boldsymbol{x})\|_1} \frac{|\nabla \phi_s(\boldsymbol{x})|^T \boldsymbol{x}}{|\phi_s(\boldsymbol{x})|}$$
$$\le \max_{i=1,\dots,n} \frac{|\nabla \phi_i(\boldsymbol{x})|^T \boldsymbol{x}}{|\phi_i(\boldsymbol{x})|} \left(1 + \sum_{s=1}^n \frac{|\phi_s(\boldsymbol{x})|}{\|\phi(\boldsymbol{x})\|_1} \right),$$

which concludes the proof. □

Lemma 3. *Let $\varphi = (\varphi_1, \dots, \varphi_n) : \mathbb{R}^n \to \mathbb{R}^n$ be defined as*

$$\varphi_i(\boldsymbol{x}) = \frac{(Af(\boldsymbol{x}))_i^{1/\varepsilon}}{\|(Af(\boldsymbol{x}))^{1/\varepsilon}\|_1}.$$

Then

$$\frac{\varepsilon}{2} \left\| \frac{\nabla \varphi_i(\boldsymbol{x}) \boldsymbol{x}}{\varphi_i(\boldsymbol{x})} \right\|_1 \le \kappa_+(A, f)$$

for all positive vectors $\boldsymbol{x} \in \mathbb{R}^n$ and all $i = 1, \dots, n$.

Proof. For $i = 1, \dots, n$, let $\phi_i : \mathbb{R}^n \to \mathbb{R}$ be defined as $\phi_i(\boldsymbol{x}) = (Af(\boldsymbol{x}))_i^{1/\varepsilon}$. From Lemma 2, for any positive $\boldsymbol{x} \in \mathbb{R}^n$, we have

$$\left\| \frac{\nabla \varphi_i(\boldsymbol{x}) \boldsymbol{x}}{\varphi_i(\boldsymbol{x})} \right\|_1 \le 2 \max_{i=1,\dots,n} \frac{|\nabla \phi_i(\boldsymbol{x})|^T \boldsymbol{x}}{|\phi_i(\boldsymbol{x})|}$$
$$= \frac{2}{\varepsilon} \max_{i=1,\dots,n} \frac{|(Af(\boldsymbol{x}))_i^{\frac{1}{\varepsilon}-1} A \nabla f_i(\boldsymbol{x})|^T \boldsymbol{x}}{(Af(\boldsymbol{x}))_i^{\frac{1}{\varepsilon}}}$$
$$\le \frac{2}{\varepsilon} \max_{i=1,\dots,n} \frac{(A|\nabla f_i(\boldsymbol{x})|)^T \boldsymbol{x}}{(Af(\boldsymbol{x}))_i}$$
$$= \frac{2}{\varepsilon} \max_{i=1,\dots,n} \frac{(A|J_f(\boldsymbol{x})|\boldsymbol{x})_i}{(Af(\boldsymbol{x}))_i} = \frac{2}{\varepsilon} \kappa_+(A, f),$$

which concludes the proof. □

Proof of Theorem 1. Consider the map $F = (F_1, \dots, F_n) : \mathbb{R}^n \to \mathbb{R}^n$ given by $F(\boldsymbol{x}) = g(Af(\boldsymbol{x}))/\|g(Af(\boldsymbol{x}))\|_1$. Since f is positive, F maps the sphere–slice $S_1^+ = \{\boldsymbol{x} > 0 : \|\boldsymbol{x}\|_1 = 1\}$ into itself. Then, on S_1^+ consider the following Thompson's distance $d(\boldsymbol{x}, \boldsymbol{y}) = \|\ln \boldsymbol{x} - \ln \boldsymbol{y}\|_\infty$. Recall that, from the Mean Value Theorem, for any differentiable $\phi : \mathbb{R}^n \to \mathbb{R}$ we have

$$\phi(\boldsymbol{x}) - \phi(\boldsymbol{y}) = \nabla \phi(\boldsymbol{\xi})^T (\boldsymbol{x} - \boldsymbol{y})$$

for a point $\boldsymbol{\xi}$ in the line segment joining $\boldsymbol{x}$ and $\boldsymbol{y}$. Applying this theorem to the function $\phi_i(\boldsymbol{x}) = (\ln F(e^{\boldsymbol{x}}))_i$, for any two $\boldsymbol{x}$ and $\boldsymbol{y}$ we get

$$d(F(e^{\boldsymbol{x}}), F(e^{\boldsymbol{y}})) = \max_{i=1,\dots,n} |\phi_i(\boldsymbol{x}) - \phi_i(\boldsymbol{y})| \le \Big(\max_{i=1,\dots,n} \max_{\boldsymbol{\xi} > 0} \|\nabla \phi_i(\boldsymbol{\xi})\|_1 \Big) \|\boldsymbol{x} - \boldsymbol{y}\|_\infty .$$

Since the exponential maps positive vectors into positive vectors bijectively, from the chain rule we have

$$\max_{\xi>0} \nabla\phi_i(\boldsymbol{\xi}) = \max_{\xi>0} \frac{\nabla F_i(e^{\xi})e^{\xi}}{F_i(e^{\xi})} = \max_{\xi>0} \frac{\nabla F_i(\boldsymbol{\xi})\boldsymbol{\xi}}{F_i(\boldsymbol{\xi})}$$

and thus, from Lemma 3, we obtain

$$\max_{i=1,\dots,n} \max_{\xi>0} \|\nabla\phi_i(\boldsymbol{\xi})\|_1 \leq \frac{2}{\varepsilon}\kappa_+(A, f)\,.$$

Now, for any two $\boldsymbol{x}$ and $\boldsymbol{y}$ in S_1^+ let $\widetilde{\boldsymbol{x}} > 0$ and $\widetilde{\boldsymbol{y}} > 0$ be such that $\boldsymbol{x} = e^{\widetilde{\boldsymbol{x}}}$ and $\boldsymbol{y} = e^{\widetilde{\boldsymbol{y}}}$. We have

$$d(F(\boldsymbol{x}), F(\boldsymbol{y})) = d(F(e^{\widetilde{\boldsymbol{x}}}), F(e^{\widetilde{\boldsymbol{y}}})) \leq \frac{\kappa_+(A, f)}{\varepsilon/2}\|\widetilde{\boldsymbol{x}} - \widetilde{\boldsymbol{y}}\|_\infty = \frac{\kappa_+(A, f)}{\varepsilon/2} d(\boldsymbol{x}, \boldsymbol{y})\,.$$

As $\kappa_+(A, f) < \varepsilon/2$, we deduce that $F : S_1^+ \to S_1^+$ is a contraction with respect to the distance d. Finally, since (S_1^+, d) is a complete metric space, we deduce that there exists a unique fixed point $\boldsymbol{u}^*$ of F in S_1^+. This implies that there exists a unique scalar $\lambda > 0$ such that $g(Af(\boldsymbol{u}^*)) = \lambda\boldsymbol{u}^*$. This shows the first point of the theorem.

For the second point, note that the sequence $\{\boldsymbol{u}_k\}$ generated by the nonlinear power method can be equivalently written as $\boldsymbol{u}_{k+1} = F(\boldsymbol{u}_k)$. Thus, we have

$$\begin{aligned} d(\boldsymbol{u}_{k+1}, \boldsymbol{u}^*) &= d(F(\boldsymbol{u}_k), F(\boldsymbol{u}^*)) \\ &\leq \frac{\kappa_+(A, f)}{\varepsilon/2} d(\boldsymbol{u}_k, \boldsymbol{u}^*) \leq \cdots \leq \left(\frac{\kappa_+(A, f)}{\varepsilon/2}\right)^{k+1} d(\boldsymbol{u}_0, \boldsymbol{u}^*) \end{aligned}$$

for any $\boldsymbol{u}_0 > 0$. This implies that $d(\boldsymbol{u}_{k+1}, \boldsymbol{u}^*) \to 0$ as $k \to \infty$, that is $\boldsymbol{u}_k \to \boldsymbol{u}^*$. □

Data Statement

All data and code related to this work are publicly available at the GitHub online repository https://github.com/ftudisco/nlc. The London tube data was collected from [Cipolla, Durastante, and Tudisco(2021)] and [Domenico *et al.*(2014)Domenico, Solé-Ribalta, Gómez, and Arenas].

Funding

The work of FT is supported by the PRIN projects FIN4GEO and MOLE. FT is a member of INdAM-GNCS. The work of DJH is supported by a Fellowship from the Leverhulme Trust.

REFERENCES

[Newman(2010)] M. E. J. Newman, *Networks: an Introduction* (Oxford University Press, Oxford, 2010).

[Schäfermeyer(2019)] J. P. Schäfermeyer, Unpublished manuscript (2019).

[Lemmens and Nussbaum(2012)] B. Lemmens and R. Nussbaum, *Nonlinear Perron–Frobenius Theory*, Vol. 189 (Cambridge University Press, Cambridge, 2012).

[Gautier, Tudisco, and Hein(2019)] A. Gautier, F. Tudisco, and M. Hein, SIAM Journal Matrix Analysis and Applications **40**, 1179—1205January (2019).

[Tudisco and Higham(2019)] F. Tudisco and D. J. Higham, SIAM J. Mathematics of Data Science **1**, 269–292 (2019).

[Sittoni and Tudisco(2024)] P. Sittoni and F. Tudisco, "Subhomogeneous deep equilibrium models," in *International Conference on Machine Learning (ICML)* (2024).

[Arrigo and Tudisco(2019)] F. Arrigo and F. Tudisco, "Multi-dimensional, multilayer, nonlinear and dynamic hits," in *Proceedings of the SIAM International Conference on Data Mining* (2019).

[Tudisco, Arrigo, and Gautier(2018)] F. Tudisco, F. Arrigo, and A. Gautier, SIAM Journal on Applied Mathematics **78**, 853–876 (2018).

[Tudisco and Higham(2021)] F. Tudisco and D. J. Higham, Communications Physics **4**, 1–10 (2021).

[Batty(2013)] M. Batty, *The New Science of Cities* (The MIT Press, Boston, 2013).

[Domenico *et al.*(2014)Domenico, Solé-Ribalta, Gómez, and Arenas] M. D. Domenico, A. Solé-Ribalta, S. Gómez, and A. Arenas, Proceedings of the National Academy of Sciences **111**, 8351–8356 (2014).

[Cipolla, Durastante, and Tudisco(2021)] S. Cipolla, F. Durastante, and F. Tudisco, ESAIM Mathematical Modelling and Numerical Analysis **55**, 77–97 (2021).

[Fagin, Kumar, and Sivakumar(2003)] R. Fagin, R. Kumar, and D. Sivakumar, SIAM Journal on Discrete Mathematics **17**, 134–160 (2003).

[Gautier, Nguyen, and Hein(2016)] A. Gautier, Q. N. Nguyen, and M. Hein, "Globally optimal training of generalized polynomial neural networks with nonlinear spectral methods," in *Advances in Neural Information Processing Systems* (2016), pp. 1687–1695.