Links Between Kleinberg's Hubs and Authorities, Correspondence Analysis, and Markov Chains

Francois Fouss & Marco Saerens
Université Catholique de Louvain
Place des Doyens 1
B-1348 Louvain-la-Neuve, Belgium
{fouss, saerens}@isys.ucl.ac.be

Jean-Michel Renders
Xerox Research Center Europe
Chemin de Maupertuis 6
F-38240 Meylan (Grenoble), France
jean-michel.renders@xrce.xerox.com

Abstract

In this work, we show that Kleinberg's hubs and authorities model is closely related to both correspondence analysis, a well-known multivariate statistical technique, and a particular Markov chain model of navigation through the web. The only difference between correspondence analysis and Kleinberg's method is the use of the ***average*** *value of the hubs (authorities) scores for computing the authorities (hubs) scores, instead of the* ***sum*** *for Kleinberg's method. We also show that correspondence analysis and our Markov model are related to SALSA, a variant of Kleinberg's model.*

1. Introduction

Exploiting the graph structure of large document repositories, such as the web environment, is one of the main challenges of computer science and data mining today. In this respect, Kleinberg's proposition to distinguish web pages that are hubs and authorities (see [4]; the HITS algorithm) has been well-received in the community.

In this paper, we show that Kleinberg's hubs and authorities procedure [4] is closely related to both correspondence analysis (see for instance [3]), a well-known multivariate statistical analysis technique, and a particular Markov chain model of navigation through the web that provides the same results as correspondence analysis. We further show that correspondence analysis and the Markov model are related to SALSA [5], a variant of Kleinberg's model. This puts new lights on the interpretation of Kleinberg's procedure since correspondence analysis has a number of interesting properties that makes it well suited for the analysis of frequency tables. On the other hand, the proposed Markov model could easily be extended to more general structures, such as relational databases.

2. Kleinberg's procedure

In [4], Kleinberg introduced a procedure for identifying web pages that are good hubs or good authorities, in response to a given query. The following example is often mentioned. When considering the query "automobile makers", the home pages of Ford, Toyota and other car makers are considered as good authorities, while web pages that list these home pages are good hubs.

To identify good hubs and authorities, Kleinberg's procedure exploits the graph structure of the web. Each web page is a node and a link from page a to page b is represented by a directed edge from node a to node b. When introducing a query, the procedure first constructs a focused subgraph G, and then computes hubs and authorities scores for each node of G. Let n be the number of nodes of G. We now briefly describe how these scores are computed. Let $\mathbf{W}$ be the adjacency matrix of the subgraph G; that is, element w_{ij} (row i, column j) of matrix $\mathbf{W}$ is equal to 1 if and only if node (web page) i contains a link to node (web page) j; otherwise, $w_{ij} = 0$. We respectively denote by $\mathbf{x}^h$ and $\mathbf{x}^a$ the hubs and authorities $n \times 1$ column vector scores corresponding to each node of the subgraph.

Kleinberg uses an iterative updating rule in order to compute the scores. Initial scores at $k = 0$ are all set to 1, i.e. $\mathbf{x}^h = \mathbf{x}^a = \mathbf{1}$ where $\mathbf{1} = [1, 1, \ldots, 1]^{\mathrm{T}}$, a column vector made of 1 (T is the matrix transpose). Then, the following mutually reinforcing rule is used: the hub score for node i, x_i^h, is set equal to the normalized sum of the authority scores of all nodes pointed by i and, similarly, the authority score of node j, x_j^a, is set equal to the normalized sum of hub scores of all nodes pointing to j. This corresponds to the following updating rule:

$$\mathbf{x}^h(k+1) = \frac{\mathbf{W}\mathbf{x}^a(k)}{\|\mathbf{W}\mathbf{x}^a(k)\|_2} \tag{1}$$

$$\mathbf{x}^a(k+1) = \frac{\mathbf{W}^{\mathrm{T}}\mathbf{x}^h(k)}{\|\mathbf{W}^{\mathrm{T}}\mathbf{x}^h(k)\|_2} \tag{2}$$

where $\|\mathbf{x}\|_2$ is the Euclidian norm, $\|\mathbf{x}\|_2 = (\mathbf{x}^{\mathrm{T}}\mathbf{x})^{1/2}$.

Kleinberg [4] showed that when following this update rule, $\mathbf{x}^h$ converges to the normalized principal (or dominant) right eigenvector of the symmetric matrix $\mathbf{W}\mathbf{W}^{\mathrm{T}}$, while $\mathbf{x}^a$ converges to the normalized principal eigenvector of the symmetric matrix $\mathbf{W}^{\mathrm{T}}\mathbf{W}$, provided that the eigenvalues are distinct.

Indeed, the equations (1), (2) result from the application of the power method, an iterative numerical method for computing the dominant eigenvector of a symmetric matrix [2], to the following eigenvalue problem:

$$\mathbf{x}^h \propto \mathbf{W}\mathbf{x}^a \Rightarrow \mathbf{x}^h = \mu\mathbf{W}\mathbf{x}^a \quad (3)$$
$$\mathbf{x}^a \propto \mathbf{W}^{\mathrm{T}}\mathbf{x}^h \Rightarrow \mathbf{x}^a = \eta\mathbf{W}^{\mathrm{T}}\mathbf{x}^h \quad (4)$$

where $\propto$ means "proportional to". This means that each hub node, i, is given a score, x_i^h, that is proportional to the sum of the authorities nodes scores to which it links to. Symmetrically, to each authorities node, j, we allocate a score, x_j^a, which is proportional to the sum of the hubs nodes scores that point to it. By substituting (3) in (4) and vice-versa, we easily obtain

$$\mathbf{x}^h = \mu\eta\mathbf{W}\mathbf{W}^{\mathrm{T}}\mathbf{x}^h = \lambda\mathbf{W}\mathbf{W}^{\mathrm{T}}\mathbf{x}^h$$
$$\mathbf{x}^a = \mu\eta\mathbf{W}^{\mathrm{T}}\mathbf{W}\mathbf{x}^a = \lambda\mathbf{W}^{\mathrm{T}}\mathbf{W}\mathbf{x}^a$$

which is an eigenvalue/eigenvector problem.

Many extensions of the updating rules (1), (2) were proposed. For instance, in [5] (the SALSA algorithm), the authors propose to normalise the matrices $\mathbf{W}$ and $\mathbf{W}^{\mathrm{T}}$ in (3) and (4) so that the new matrices verify $\mathbf{W}'\mathbf{1} = \mathbf{1}$ and $\mathbf{W}^{\mathrm{T}\prime}\mathbf{1} = \mathbf{1}$ (the sum of the elements of each row of $\mathbf{W}'$ and $\mathbf{W}^{\mathrm{T}\prime}$ is 1). In this case, (3) and (4) can be rewritten as

$$x_i^h \propto \frac{\sum_{j=1}^{n} w_{ij} x_j^a}{w_{i.}} \text{with } w_{i.} = \sum_{j=1}^{n} w_{ij} \quad (5)$$
$$x_j^a \propto \frac{\sum_{i=1}^{n} w_{ij} x_i^h}{w_{.j}} \text{with } w_{.j} = \sum_{i=1}^{n} w_{ij} \quad (6)$$

This normalization has the effect that nodes (web pages) having a large number of links are not privileged with respect to nodes having a small number of links. The relations (5) and (6) are not explicitely used in order to compute hubs and authorities scores (it would lead to the dominant right eigenvector, which is a trivial one, $\mathbf{1}$), but lead to an eigenvalue/eigenvector problem, as will become clear in next section. In the SALSA algorithm, the hubs and authorities scores are the "steady-state values" computed from the corresponding Markov model (see section 4).

3. Links with correspondence analysis

Correspondence analysis is a standard multivariate statistical analysis technique aiming to analyse frequency tables [3]. Suppose that we have a table of frequencies, $\mathbf{W}$, for which each cell, w_{ij}, represents the number of cases having both values i for the row variable and j for the column variable (we simply use the term "value" for the discrete value taken by a categorical variable). In our case, the records are the directed edges; the row variable represents the index of the origin node of the edge (hubs) and the column variable the index of the end node of the edge (authorities).

Correspondence analysis associates a score to the values of each of these variables. These scores relate the two categorical variables by what is called a "**reciprocal averaging**" relation [3]:

$$x_i^h \propto \frac{\sum_{j=1}^{n} w_{ij} x_j^a}{w_{i.}} \text{with } w_{i.} = \sum_{j=1}^{n} w_{ij} \quad (7)$$
$$x_j^a \propto \frac{\sum_{i=1}^{n} w_{ij} x_i^h}{w_{.j}} \text{with } w_{.j} = \sum_{i=1}^{n} w_{ij} \quad (8)$$

which is exactly the same as (5) and (6). This means that each hub node, i, is given a score, x_i^h, that is proportional to the average of the authorities nodes scores to which it links to. Symmetrically, to each authorities node, j, we allocate a score, x_j^a, which is proportional to the average of the hubs nodes scores that point to it.

Now, by defining the diagonal matrix $\mathbf{D}^h = \mathrm{diag}(1/w_{i.})$ and $\mathbf{D}^a = \mathrm{diag}(1/w_{.j})$ containing the number of links, we can rewrite (7) and (8) in matrix form

$$\mathbf{x}^h \propto \mathbf{D}^h\mathbf{W}\mathbf{x}^a \Rightarrow \mathbf{x}^h = \mu\mathbf{D}^h\mathbf{W}\mathbf{x}^a \quad (9)$$
$$\mathbf{x}^a \propto \mathbf{D}^a\mathbf{W}^{\mathrm{T}}\mathbf{x}^h \Rightarrow \mathbf{x}^a = \eta\mathbf{D}^a\mathbf{W}^{\mathrm{T}}\mathbf{x}^h \quad (10)$$

In the language of correspondence analysis, the row vectors of $\mathbf{D}^h\mathbf{W}$ are the hub **profiles**, while the row vectors of $\mathbf{D}^a\mathbf{W}^{\mathrm{T}}$ are the authorities **profiles**. These vectors sum to one. Notice that (9) and (10) differ from (3) and (4) only by the fact that we use the **average value** in order to compute the scores, instead of the sum.

Now, from (9), (10), we easily find

$$\mathbf{x}^h = \mu\eta\mathbf{D}^h\mathbf{W}\mathbf{D}^a\mathbf{W}^{\mathrm{T}}\mathbf{x}^h = \lambda\mathbf{D}^h\mathbf{W}\mathbf{D}^a\mathbf{W}^{\mathrm{T}}\mathbf{x}^h \quad (11)$$
$$\mathbf{x}^a = \mu\eta\mathbf{D}^a\mathbf{W}^{\mathrm{T}}\mathbf{D}^h\mathbf{W}\mathbf{x}^a = \lambda\mathbf{D}^a\mathbf{W}^{\mathrm{T}}\mathbf{D}^h\mathbf{W}\mathbf{x}^a \quad (12)$$

Correspondence analysis computes the **subdominant right eigenvector** of the matrices $\mathbf{D}^h\mathbf{W}\mathbf{D}^a\mathbf{W}^{\mathrm{T}}$ and $\mathbf{D}^a\mathbf{W}^{\mathrm{T}}\mathbf{D}^h\mathbf{W}$. Indeed the right principal eigenvector is a trivial one, $\mathbf{1} = [1, 1, \ldots, 1]^{\mathrm{T}}$, with eigenvalue $\lambda = 1$ (all the other eigenvalues are real positive and smaller that 1; see [3]) since the sum of the columns of $\mathbf{D}^h\mathbf{W}\mathbf{D}^a\mathbf{W}^{\mathrm{T}}$ (respectively $\mathbf{D}^a\mathbf{W}^{\mathrm{T}}\mathbf{D}^h\mathbf{W}$) is one for each row.

In standard correspondence analysis, this subdominant right eigenvector has several interesting interpretations in terms of "optimal scaling", of the "best approximation" in terms of chi-square distance to the original matrix, or of the linear combinations of the two sets of values that are "maximally correlated", etc. (see for instance [3]).

The next eigenvectors can be computed as well; they are related to the proportion of chi-square computed on the original table of frequencies that can be explained by the first m eigenvectors: they measure the departure to independence of the two discrete variables. Correspondence analysis is therefore often considered as an "equivalent" of principal components analysis for frequency tables.

4. A Markov chain model of web navigation

We now introduce a Markov chain model of random web navigation that provides the same results as correspondence analysis, and is therefore closely related to Kleinberg's procedure and SALSA. Hence, it provides a new interpretation for both correspondence analysis and Kleinberg's procedure. Notice that this random walk model is quite similar to the one proposed in SALSA ([5]; for other random walk models of web navigation, see PageRank [7] or [6]).

We first define a **Markov chain** in the following way. We associate a state of the Markov chain to every hub and every authority node ($2n$ in total); we also define a random variable, $s(k)$, representing the state of the Markov model at time step k. Moreover, let S^h be the subset of states that are hubs and S^a be the subset of states that are authorities. We say that $s^h(k) = i$ (respectively $s^a(k) = i$) when the Markov chain is in the state corresponding to the i^{th} hub (authority) at time step k. As in [5], we define a random walk on these states by the following single-step transition probabilities

$$\mathrm{P}(s^h(k+1) = i | s^a(k) = j) = \frac{w_{ij}}{w_{.j}} \tag{13}$$

$$\mathrm{P}(s^a(k+1) = j | s^h(k) = i) = \frac{w_{ij}}{w_{i.}} \tag{14}$$

All the other transitions being impossible: $\mathrm{P}(s^a(k+1) = j | s^a(k) = i) = 0$ and $\mathrm{P}(s^h(k+1) = j | s^h(k) = i) = 0$, for all i, j. In other words, to any hub page, $s^h(k) = i$, we associate a non-zero probability of jumping to an authority page, $s^a(k+1) = j$, pointed by the hub page (equation 14), which is inversely proportional to the number of directed edges leaving $s^h(k) = i$. Symmetrically, to any authority page $s^a(k) = i$, we associate a non-zero probability of jumping to a hub page $s^h(k+1) = j$ pointing to the authority page (equation 13), which is inversely proportional to the number of directed edges pointing to $s^a(k) = i$.

We suppose that the Markov chain is irreducible, that is, any state can be reached from any other state. If this is not the case, the Markov chain can be decomposed into closed sets of states which are completely independent (there is no communication between them), each closed set being irreducible. In this situation, our analysis can be performed on these closed sets instead of the full Markov chain.

Now, if we denote the probability of being in a state by $x_i^h(k) = \mathrm{P}(s^h(k) = i)$ and $x_i^a(k) = \mathrm{P}(s^a(k) = i)$, and we define $\mathbf{P}^h$ as the transition matrix whose elements are $p_{ij}^h = \mathrm{P}(s^h(k+1) = j | s^a(k) = i)$ and $\mathbf{P}^a$ as the transition matrix whose elements are $p_{ij}^a = \mathrm{P}(s^a(k+1) = j | s^h(k) = i)$, from equations (13) and (14),

$$\mathbf{P}^h = \mathbf{D}^a \mathbf{W}^{\mathrm{T}} \tag{15}$$

$$\mathbf{P}^a = \mathbf{D}^h \mathbf{W} \tag{16}$$

The evolution of the Markov model is characterized by

$$\mathbf{x}^h(k+1) = (\mathbf{P}^h)^{\mathrm{T}} \mathbf{x}^a(k) \tag{17}$$

$$\mathbf{x}^a(k+1) = (\mathbf{P}^a)^{\mathrm{T}} \mathbf{x}^h(k) \tag{18}$$

It is easy to observe that the Markov chain is periodic with period 2: each hub (authority) state could potentially be reached in one jump from an authority (hub) state but certainly not from any other hub (authority) state. In this case, the set of hubs (authorities) corresponds to a subset which itself is an irreducible and aperiodic Markov chain whose evolution is given by

$$\mathbf{x}^h(k+2) = (\mathbf{P}^h)^{\mathrm{T}} (\mathbf{P}^a)^{\mathrm{T}} \mathbf{x}^h(k) = (\mathbf{Q}^h)^{\mathrm{T}} \mathbf{x}^h(k)$$

$$\mathbf{x}^a(k+2) = (\mathbf{P}^a)^{\mathrm{T}} (\mathbf{P}^h)^{\mathrm{T}} \mathbf{x}^a(k) = (\mathbf{Q}^a)^{\mathrm{T}} \mathbf{x}^a(k)$$

where $\mathbf{Q}^h$ and $\mathbf{Q}^a$ are the transition matrices of the corresponding Markov models for the hubs and authorities. These Markov chains are aperiodic since each link (corresponding to a transition) can be followed in both directions (from hub to authority and from authority to hub) so that, when starting from a state, we can always return to this state in two steps. Hence, all the diagonal elements of $\mathbf{Q}^h$ and $\mathbf{Q}^a$ are non-zero and the Markov chains are aperiodic.

Therefore, the transition matrices of the corresponding Markov chains for hubs and authorities are

$$\mathbf{Q}^h = \mathbf{P}^a \mathbf{P}^h = \mathbf{D}^h \mathbf{W} \mathbf{D}^a \mathbf{W}^{\mathrm{T}} \tag{19}$$

$$\mathbf{Q}^a = \mathbf{P}^h \mathbf{P}^a = \mathbf{D}^a \mathbf{W}^{\mathrm{T}} \mathbf{D}^h \mathbf{W} \tag{20}$$

The matrices appearing in these equations are equivalent to the ones appearing in (11), (12). Now, it is well-know that the subdominant (the dominant right eigenvector is trivially $\mathbf{1}$) right eigenvector – as computed in correspondence analysis – of the transition matrix, $\mathbf{Q}$, of an irreducible, aperiodic, Markov chain measures the departure of each state from the "equilibrium position" or "steady-state" probability vector (for a precise definition, see the appendix A or [9]), $\boldsymbol{\pi}$, which is given by the first left eigenvector of the transition matrix $\mathbf{Q}$: $\mathbf{Q}^{\mathrm{T}} \boldsymbol{\pi} = \boldsymbol{\pi}$, subject to $\sum_{i=1}^{n} \pi_i = 1$, with eigenvalue $\lambda = 1$. This principal left eigenvector, $\boldsymbol{\pi}$, is unique and positive and is called the "steady state" vector; it corresponds to the probability of finding the Markov chain in state $s = i$ in the long-run behavior, $\lim_{k \to \infty} \mathrm{P}(s(k) = i) = \pi_i$, and is independent of the initial distribution of states at $k = 0$. The elements of the subdominant right eigenvector of $\mathbf{Q} = \mathbf{Q}^h$ or $\mathbf{Q}^a$ can thus be interpreted as a kind of "distance" from each state to its "steady-state" value. The states of the Markov chain are often classified by means of the values of this subdominant eigenvector as well as the few next eigenvectors [9].

Lempel and Moran [5], in the SALSA algorithm, propose, as hubs and authorities scores, to compute the steady-state vectors, $\boldsymbol{\pi}^h$ and $\boldsymbol{\pi}^a$, corresponding to the hubs and the authorities transition matrices, $\mathbf{Q}^h$, $\mathbf{Q}^a$. We propose instead (or maybe in addition) to use the subdominant right eigenvector, which produces the same results as correspondence analysis, and which is often used in order to characterise the states of the Markov chain, as already mentionned. Eventually, the next eigenvectors/eigenvalues could be computed as well; they correspond to higher-order corrections.

5. Conclusions

We showed that Kleinberg's method for computing hubs and authorities scores is closely related to correspondence analysis, a well-known multivariate statistical analysis method. This will allow to provide new interpretations of Kleinberg's method. We then introduce a random walk model of navigation through the web, related to SALSA, and we show that this model is equivalent to correspondence analysis. This random walk model has an important advantage: it could easily be extended to more complex structures, such as relational databases.

References

[1] P. Bremaud. *Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues*. Springer-Verlag New York, 1999.

[2] G. H. Golub and C. F. V. Loan. *Matrix Computations, 3th Ed.* The Johns Hopkins University Press, 1996.

[3] M. J. Greenacre. *Theory and Applications of Correspondence Analysis*. Academic Press, 1984.

[4] J. M. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46(5):604–632, 1999.

[5] R. Lempel and S. Moran. Salsa: The stochastic approach for link-structure analysis. *ACM Transactions on Information Systems*, 19(2):131–160, 2001.

[6] A. Y. Ng, A. X. Zheng, and M. I. Jordan. Link analysis, eigenvectors and stability. *International Joint Conference on Artificial Intelligence (IJCAI-01)*, 2001.

[7] L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: Bringing order to the web. *Technical Report, Computer System Laboratory, Stanford University*, 1998.

[8] A. Papoulis and S. U. Pillai. *Probability, Random Variables and Stochastic Processes*. McGraw-Hill, 2002.

[9] W. J. Stewart. *Introduction to the Numerical Solution of Markov Chains*. Princeton University Press, 1994.

A. Appendix: Distance to steady state vector

In this appendix, we show that the entries of the subdominant right eigenvector of the transition matrix $\mathbf{Q}$ of the aperiodic, irreducible Markov chains for hubs and authorities can be interpreted as a distance to the "steady-state" vector, $\boldsymbol{\pi}$. From (19), (20), we can easily show that $\mathbf{Q}$ is positive semidefinite so that all its eigenvalues are positive real and its eigenvectors are real. Moreover, since $\mathbf{Q}$ is stochastic nonnegative, all the eigenvalues are ≤ 1, and the eigenvalue 1 has multiplicity one. The proof is adapted from [8], [9], [1].

Let $\mathbf{e}_l = [0, 0, \ldots, 1, \ldots, 0]^{\mathrm{T}}$ be the column vector with the l^{th} component equal to 1, all others being equal to 0. $\mathbf{e}_l$ will denote that, initially, the system starts in state l. After one time step, the probability density of finding the system in one state is $\mathbf{x}(1) = \mathbf{Q}^{\mathrm{T}}\mathbf{e}_l$, and after k steps, $\mathbf{x}(k) = (\mathbf{Q}^{\mathrm{T}})^k \mathbf{e}_l$. Now, the idea is to compute the distance

$$d_l(k) = || \left(\mathbf{Q}^{\mathrm{T}}\right)^k \mathbf{e}_l - \boldsymbol{\pi} ||_2 \qquad (21)$$

in order to have an idea of the rate of convergence to the steady state when starting from a particular state $s = l$.

Let $(\lambda_i, \mathbf{u}_i)$, $i = 1, 2, \ldots n$ represent the n right eigenvalue/eigenvectors pairs of $\mathbf{Q}$ in decreasing order of λ_i. Thus $\mathbf{QU} = \mathbf{U\Lambda}$ where $\mathbf{U}$ is the $n \times n$ matrix made of the column vectors $\mathbf{u}_i$ which form a basis of $\Re^n$, $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n]$, and $\mathbf{\Lambda} = \mathrm{diag}(\lambda_i)$. Hence,

$$\mathbf{Q} = \mathbf{U\Lambda U}^{-1} = \mathbf{U\Lambda V} \qquad (22)$$

where we set $\mathbf{V} = \mathbf{U}^{-1}$. We therefore obtain $\mathbf{VQ} = \mathbf{\Lambda V}$ where $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n]^{\mathrm{T}}$, so that the column vectors $\mathbf{v}_i$ are the left eigenvectors of $\mathbf{Q}$, $\mathbf{v}_i^{\mathrm{T}}\mathbf{Q} = \lambda_i \mathbf{v}_i^{\mathrm{T}}$: the rows of $\mathbf{V}$ are $\mathbf{v}_i^{\mathrm{T}}$. Moreover, since $\mathbf{VU} = \mathbf{I}$, we have $\mathbf{v}_i^{\mathrm{T}}\mathbf{u}_j = \delta_{ij}$.

Hence, from (22),

$$\begin{aligned} \mathbf{Q}^k &= \mathbf{U\Lambda}^k\mathbf{V} = \textstyle\sum_{i=1}^n \lambda_i^k \mathbf{u}_i \mathbf{v}_i^{\mathrm{T}} \\ &= \mathbf{1}\boldsymbol{\pi}^{\mathrm{T}} + \textstyle\sum_{i=2}^n \lambda_i^k \mathbf{u}_i \mathbf{v}_i^{\mathrm{T}} \\ &= \mathbf{1}\boldsymbol{\pi}^{\mathrm{T}} + \lambda_2^k \mathbf{u}_2 \mathbf{v}_2^{\mathrm{T}} + O((n-2)\lambda_3^k) \qquad (23) \end{aligned}$$

since $\lambda_i < 1$ for $i > 1$ and the eigenvalues/eigenvectors are sorted in decreasing order of eigenvalue. Let us now return to (21)

$$\begin{aligned} d_l(k) &= \left\| \left(\mathbf{Q}^{\mathrm{T}}\right)^k \mathbf{e}_l - \boldsymbol{\pi} \right\|_2 \\ &\simeq \left\| \left(\boldsymbol{\pi}\mathbf{1}^{\mathrm{T}} + \lambda_2^k \mathbf{v}_2 \mathbf{u}_2^{\mathrm{T}}\right) \mathbf{e}_l - \boldsymbol{\pi} \right\|_2 \\ &\simeq \left\| \boldsymbol{\pi} + \lambda_2^k \mathbf{v}_2 \mathbf{u}_2^{\mathrm{T}} \mathbf{e}_l - \boldsymbol{\pi} \right\|_2 \\ &\simeq \lambda_2^k \, \|\mathbf{v}_2\|_2 \, u_{2l} \end{aligned}$$

where u_{2l} is l^{th} component of $\mathbf{u}_2$. Since the only term that depends on the initial state, l, is u_{2l}, the eigenvector $\mathbf{u}_2$ can be interpreted as a distance to the steady-state vector. ■