New Models for Multi-Class Networks

Omar De la Cruz Cabrera* Jiafeng Jin* Lothar Reichel*

March 30, 2021

Abstract

Many complex phenomena can be modeled by networks, that is, by a set of nodes connected by edges. Networks are represented by graphs, and several algebraic and analytical methods have been developed for their study. However, in order to obtain a more useful representation of a system, it is often appropriate to include more information about the nodes and/or edges, and those additions make it necessary to adapt or modify such methods of study.

Multi-class networks, in which the set of nodes and/or the set of edges are partitioned in two or more classes, are useful when different nodes and edges can play fundamentally distinct roles in the system. In this article we introduce new models and methods for multi-class networks, based on how the adjacency matrix is formed.

We apply this approach to obtain measures of node importance or centrality, in particular using the Perron eigenvector. Perturbation results shed light on how the relative importance of a node changes by the addition of a single edge, and experiments with both synthetic and real data sets illustrate features of the methods discussed.

1 Introduction

Many complex phenomena can be modeled by networks, that is, by a set of *nodes* and a set of connections among them, called *edges*. Network models simplify reality by ignoring some information. Indeed, in the simplest network models, neither nodes nor edges have attributes and only the connectivity between the nodes is modeled. This simplification makes it possible to apply graph theory and use powerful quantitative methods to extract information about complex systems that might not be easily accessible otherwise; see [12, 20] for many examples.

However, in order to obtain a more faithful representation of a system, it sometimes becomes necessary to include more information about the nodes and/or edges, and such additions usually make it necessary to adapt or modify the quantitative methods of analysis. An example is the inclusion of weights, in which edges and/or nodes are assigned a numerical value, representing a characteristic of interest for the modeler; see, e.g., [1, 19, 24]. Another example is furnished by *multi-class networks*, in which each node belongs to exactly one of k node classes, and each edge belongs to exactly one of d edge classes. This kind of model becomes necessary when different nodes can play fundamentally distinct roles in the system. It is the purpose of the present paper to discuss models for multi-class networks. Examples of multi-class networks include:

- **Bibliographic Network**: The node classes may be papers, authors, and journals. Classes of edges (relationships) may be citations (between papers), authorships (between papers and authors), and publications (between papers and journals).

*Department of Mathematical Sciences, Kent State University, Kent, OH 44242, USA. Email: odelacru@kent.edu (O. De la Cruz Cabrera), jjin3@kent.edu (Jiafeng Jin), reichel@math.kent.edu (L. Reichel)

- **Twitter Network**: The node classes may be users/accounts, tweets, and comments. Classes of edges (relationships) may be following (between users), tweet posting (between users and tweets), comment posting (between users and comments), and the relationship between comments and the tweets they refer to.

- **Healthcare Network**: The node classes may be patients, health conditions, treatments, providers, and insurers. Notice that "treatments" could be modeled simply as edges between providers and patients, but by modeling treatments as nodes, we can connect them to the patients, the providers, the treatment types, the health condition, and the insurer or insurers who is/are paying for it.

- **Genomic Network**: The node classes may be genes, diseases, chemical compounds, gene ontology categories, and tissues. Classes of edges (relationships) between nodes may be protein-protein interaction (between proteins), encoding (between genes and proteins), and up-regulation and down-regulation (between regulatory elements and genes).

Of the many properties of networks that can be studied quantitatively, in this paper we are interested in the ranking of nodes within node classes by their importance or centrality. Our work is inspired by previous results on 1-class, 2-class, and 3-class models by Bini, Del Corso, and Romani in two nice papers [2, 3], which they applied to the analysis of bibliographic networks. Our models are a modification of their approach, and can be extended in a natural way to k-class networks with $k > 3$. We will discuss differences and similarities of their and our approaches in detail.

This paper is organized as follows: Section 2 defines basic concepts. We review related work in Section 3, with particular focus on the work by Bini *et al.* [2, 3]. Sections 4-6 introduce our models, starting from one-class networks and increasing the complexity until reaching k-class networks. We include motivating examples and establish theoretical results based on perturbation theory in these sections. Section 7 presents computed results for some large-scale problems. In particular, these examples compare the orderings obtained with the models of the present paper to the ordering obtained with the models used by Bini *et al.* [2, 3]. Section 8 contains concluding remarks.

2 Basic Definitions

We model networks using the mathematical concept of a *graph* $\mathcal{G} = \langle V, E \rangle$, which is given by a pair of sets $V = \{v_1, \dots, v_n\}$, containing the *vertices* or *nodes*, and $E = \{e_1, \dots, e_m\}$, containing the *edges*. In a *directed* graph, at least one edge has a unique starting node and a unique ending node, while in an *undirected* graph all edges connect two nodes without a specified direction. For simplicity, we will not allow self-loops, and we will not allow more than one edge between a given pair of nodes, unless the edges are directed and go in opposite directions. Then the *adjacency matrix* for $\mathcal{G}$ is a matrix $A = [a_{ij}] \in \mathbb{R}^{n \times n}$ with $a_{ij} = 1$ if there is an edge from node v_i to node v_j, and $a_{ij} = 0$ otherwise. If $\mathcal{G}$ is undirected, then A is symmetric. We will identify a directed edge e starting at v_i and ending at v_j with the ordered pair (v_i, v_j). A walk of length k is a sequence $v_{i_1}, v_{i_2}, \dots, v_{i_{k+1}}$ of nodes and a sequence $e_{i_1}, e_{i_2}, \dots, e_{i_k}$ of edges such that e_{i_j} points from v_{i_j} to $v_{i_{j+1}}$. The nodes and edges of a walk are not required to be distinct. For further discussions on networks and graphs, we refer to [12, 20].

DEFINITION 1. A *multi-class network* is a directed graph $\mathcal{G} = \langle V, E \rangle$ together with a map $C : V \to \{1, \dots, k\}$ that assigns to each vertex one of k vertex classes, and a map $D : E \to \{1, \dots, d\}$ that assigns to each edge one of d edge classes, such that:

$$\text{For all } e = (v, w) \text{ and } e' = (v', w'), \text{ we have: } D(e) = D(e') \text{ iff } C(v) = C(v') \text{ and } C(w) = C(w') \tag{1}$$

REMARK 1. Notice that:

1. Equivalent definitions have been proposed elsewhere, e.g., in [23] under the term *information network*, further specified as *heterogeneous* if $k > 1$ and *homogeneous* otherwise.
2. The direct conditional in (1) is a natural condition, stating that edges of the same kind should connect nodes of the same types; one can always define more edge classes to make sure this holds.
3. The converse conditional in (1) is a technical requirement, which makes it easier to use block adjacency matrices (see below).
4. Definition 1 uses a directed graph for greater generality; if modeling considerations indicate that the relationship represented by a class of edges is symmetric, then this can be accommodated by adding a reverse class of edges containing the appropriate reverse edges.
5. Algebraically, Definition 1 means that there is a graph homomorphism (see [14]) from $\mathcal{G}$ onto a graph $\mathcal{S}$ with k vertices and d edges. One can allow self-loops in $\mathcal{S}$; a self-loop corresponds to a class of edges in $\mathcal{G}$ that connect nodes within the same node class.

3 Related work

An early example of ranking objects in a multi-class network is described by Law and Lodge [17], who are interested in ranking actors. The currently most popular ranking method, PageRank, was introduced by Brin and Page [6] and has spurred considerable related work; see [5], [20, Chapter 7], and references therein. Recently, tensor methods for multi-relation data have received considerable attention; see [7, 18, 21, 22]. We will comment on how our approach relates to the tensor method described by Ng et al. [21] at the end of Section 7. Similarly as Brin and Page, we will use the left Perron vector of an adjacency matrix to rank nodes, but the way we define the adjacency matrix is different.

The ranking of scientific publications has received considerable attention over the years. The simplest ranking method is to count the number of citations that each paper receives without considering its contents. The journal impact factor (JIF), introduced by Garfield [13], reflects the average number of citations that papers published in a journal receive over two years. This method ignores the quality of the papers and authors that provide citations. The JIF therefore might not provide an accurate measure of the quality of the papers in a journal; see Bini *et al.* [2, 3], as well as Del Corso and Romani [8] for discussions.

Bini, Del Corso, and Romani proposed an integrated model to evaluate papers, authors, and journals based on the quality of the papers, authors, and journals [2, 3]. We refer to this model as the BDR model. Bini et al. introduce a dummy paper, a dummy author, and a dummy journal to obtain an irreducible adjacency matrix. A dummy paper is a paper that cites all other papers and is cited by all other papers, a dummy author is the author who writes the dummy paper, and a dummy journal is a journal in which the dummy paper is published. Suitable normalization then yields a row-stochastic adjacency matrix, whose left unit eigenvector, in the Euclidean norm, associated with the largest eigenvalue can be scaled to have positive entries. This vector is unique and is referred to as the left Perron vector of the adjacency matrix. The relative importance of a node in its class is proportional to the relative size of the corresponding component of the left Perron vector.

Notice that the left Perron eigenvector of a row stochastic matrix can be interpreted, under suitable assumptions, as the steady distribution for a random walk on the nodes that follows the edges. However, the use of the Perron eigenvectors as measures of centrality does not depend on scaling the matrix to be row stochastic, but on the recursive notion that the centrality of a node is proportional to the sum of the centralities of its neighbors [4]. Whether an adjacency matrix should to be scaled or not is a modeling issue.

It follows that the relative importance of a paper in the BDR model is not merely based on the number of citations it receives, but also on the quality of the citations (the importance of the citing papers), the prestige of its author or co-authors, and the reputation of the journal in which it is published. Bini, Del Corso, and Romani [2, 3] proposed the following the models:

One-class model: Bini et al. [2, 3] consider n_1 papers and define an associated adjacency matrix $H = [h_{ij}] \in \mathbb{R}^{n_1 \times n_1}$, such that $h_{ij} = 1$ when paper i cites paper j, and $h_{ij} = 0$ otherwise. They assume the importance that paper i gives to other papers is scaled by the total number of papers it cites. This yields the row-stochastic matrix

$$\hat{H} = [\hat{h}_{ij}] \in \mathbb{R}^{n_1 \times n_1}, \qquad \hat{h}_{ij} = \frac{h_{ij}}{\sum_{k=1}^{n_1} h_{ik}}.$$

The above formula assumes that the denominator is positive. This is secured by letting one of the papers be the dummy paper that cites every other paper and is cited by every paper (except by itself). We henceforth will not explicitly discuss the dummy items in the BDR models. The entries of the unit left Perron vector of $\hat{H}$ yield the relative importance of the papers; large entries indicate high relative importance.

Two-class model: Bini et al. [2, 3] consider a model with n_2 authors and n_1 papers. It is determined by the adjacency matrix H defined above together with an additional adjacency matrix $K = [k_{ij}] \in \mathbb{R}^{n_2 \times n_1}$, such that $k_{ij} = 1$ if author i writes paper j, and $k_{ij} = 0$ otherwise. To determine the importance of papers and authors, Bini et al. use the model

$$S = \begin{bmatrix} KK^T & K \\ K^T & H \end{bmatrix}.$$

The $(ij)^{th}$ entry of the matrix $A = [a_{ij}] = KK^T \in \mathbb{R}^{n_2 \times n_2}$ indicates the number of papers that are co-authored by authors i and j. Bini et al. assume that the importance given by author i to his/her co-authors is scaled by the total number of papers that are co-authored by author i and his/her co-authors, and they also assume that the importance that paper i gives to its authors is scaled by the total number of authors who write paper i. They therefore row-normalize the matrices $A = KK^T$ and $P = [p_{ij}] = K^T$ such that

$$\hat{A} = [\hat{a}_{ij}], \quad \hat{a}_{ij} = \frac{a_{ij}}{\sum_{k=1}^{n_2} a_{ik}}, \qquad \hat{P} = [\hat{p}_{ij}], \quad \hat{p}_{ij} = \frac{p_{ij}}{\sum_{k=1}^{n_2} p_{ik}}.$$

To avoid that the importance of a paper is not proportional to the number of co-authors, Bini et al. column-normalization the matrix K such that

$$\hat{K} = [\hat{k}_{ij}], \quad \hat{k}_{ij} = \frac{k_{ij}}{\sum_{\ell=1}^{n_2} k_{\ell j}}$$

and then apply the following algorithm [2, 3] to obtain a row-stochastic matrix:

ALGORITHM 1. *For each $i \in \{1, \ldots, n_2\}$, compute $s_i = \sum_{j=1}^{n_1} \hat{k}_{ij}$. If $s_i \leqslant 1$, set $\tilde{k}_{ij} = \hat{k}_{ij}$, for $j = 1, \ldots, n_1 - 1$, and $\tilde{k}_{in_1} = 1 - \sum_{j=1}^{n_1 - 1} \hat{k}_{ij}$. Else divide the entries of the i^{th} row of $\hat{K}$ by s_i, that is, set $\tilde{k}_{ij} = \hat{k}_{ij}/s_i$. Output $\widetilde{K} = [\tilde{k}_{ij}]$.*

Finally, Bini et al. compute the left Perron vector of the matrix

$$\begin{bmatrix} \hat{A} & \widetilde{K} \\ \hat{P} & \hat{H} \end{bmatrix}.$$

The first n_2 components of the left Perron vector indicate the importance of the authors and the last n_1 components show the importance of the papers.

k-class	BDR model	Model of Section 5
$k = 2$	$\begin{bmatrix} KK^T & K \\ K^T & H \end{bmatrix}$	$\begin{bmatrix} H & K^T \\ K & 0 \end{bmatrix}$
Difference	Authors receive direct importance from their co-authors.	Authors receive indirect importance from their co-authors.

Table 1: Comparison of the BDR two-class model and the two-class model of Section 5. For the bibliographic network, we use H, K, and K^T to replace matrices H_{11}, H_{21}, and H_{12} respectively for the two-class model introduced in Section 5. The scaling of the matrices in both models has not been considered.

We say that object i receives *direct importance* from object j, if there is a walk of length one from object j to object i. Object i is said to receive *indirect importance* from object j, if there is a walk of length strictly larger than one from object j to object i, and there is no walk of length one from object j to object i. Figure 1 shows the difference in how importance is propagated in the BDR model and our model to be introduced in Section 5.

Table 1 displays the adjacency matrices for the two-class BDR model and our two-class model of Section 5. The BDR model assumes that the importance of author i is given by the direct importance of the paper(s) that he/she writes and the direct importance of his/her co-authors. However, in our model the importance of author i is given by the direct importance of the paper(s) he/she writes; the importance of paper j written by author i is given by the importance of the papers that cite paper j, and the importance of its authors. Therefore, author i also receives the indirect importance from his/her co-authors through the papers they co-author. In the BDR model, author i's work with well-known co-authors has a higher probability to be more important than author j's work with less well-known co-authors, even if author j's paper is more important than author i's paper. In our model, the importance of author i depends more on the importance of the papers he/she writes than on the reputation of his/her co-authors. Therefore, the authors give more importance to their co-authors in the BDR model than in our model. As already mentioned, the BDR model includes a dummy author.

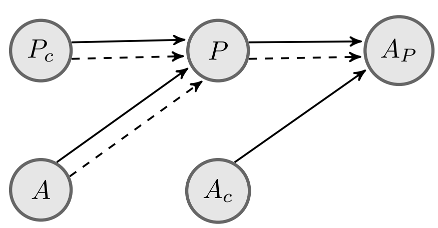

Figure 1: Comparison of the flow of importance of the BDR two-class model and the two-class model of Section 5. Solid lines show the flow of importance in the BDR two-class model and dashed lines show the flow of importance in the two-class model of Section 5. Here A stands for the author or set of co-authors of paper P, P_c stands for the set of papers that cite paper P, A_P stands for one of the authors of paper P, and A_c stands for the co-authors of author A_P.

The preferred choice of model for the ranking of papers, authors, and co-authors may depend on the context in which the results are used. Our model is an alternative to the BDR model. More details of the former are provided in Section 5. One of the features of the two-class model of this paper is that it naturally generalizes to k-class models. This is illustrated below.

Three-class model: Bini et al. [2, 3] consider the joint ranking of n_1 papers, n_2 authors, and n_3 journals by expanding their two-class model to a three-class model that in addition to the matrices H and K includes an adjacency matrix $F = [f_{ij}] \in \mathbb{R}^{n_3 \times n_1}$, such that $f_{ij} = 1$ if journal i publishes paper j, and $f_{ij} = 0$ otherwise. To determine the importance of these three objects, they use the adjacency matrix of the column labeled "BDR model" in Table 2.

k-class	BDR model	Our model
$k = 3$	$\begin{bmatrix} FHF^T & FK^T & F \\ KF^T & KK^T & K \\ F^T & K^T & H \end{bmatrix}$	$\begin{bmatrix} H & K^T & F^T \\ K & 0 & 0 \\ F & 0 & 0 \end{bmatrix}$
Difference 1	Authors receive direct importance from his/her co-authors and journals where he/she published papers.	Authors receive indirect importance from his/her co-authors and journals where he/she published papers.
Difference 2	Journal i receives direct importance from authors and journals that are associated with journal i.	Journal i receives indirect importance from authors and journals that are associated with journal i.

Table 2: Comparison between the BDR three-class model and our three-class model. The scaling of the matrices in both models has not been considered.

Bini et al. modify the adjacency matrix shown in the column "BDR model" of Table 2 by rescaling before computing its Perron vector. As already mentioned, the BDR model includes a dummy journal. A new matrix $\hat{F}$ is obtained from F by dividing each entry by $\mu = \max\limits_{1 \leqslant i \leqslant n_3} (F \cdot \mathbf{e})$, that is, the maximum number among the number of papers published in each journal. Here and throughout this paper $\mathbf{e} = [\mathbf{1}, \mathbf{1}, \ldots, \mathbf{1}]^{\mathbf{T}}$ denotes a vector will all entries equal to one. Algorithm 1 is applied to a row-scaled version $\hat{F}$. The scaling is designed to make the matrix row-stochastic. This is the matrix used in the BDR model. This scaling has the effect that a journal that publishes many papers does not automatically give high importance to the papers published in it. The matrices FHF^T, FK^T and KF^T are normalized similarly as F. The matrix F^T is normalized similarly as K in the BDR two-class model. These matrices are used in the adjacency matrix of the BDR model. The (left) Perron vector of this scaled adjacency matrix determines the ranking of the papers, authors, and journals. Scaling details of our three-class model will be discussed in Section 7.

Table 2 shows the difference between the BDR three-class model and our three-class model. The BDR model assumes that the importance of author i is given by the direct importance of the papers he/she writes, the direct importance of his/her co-authors, and the direct importance of the journals where he/she publishes papers. The BDR model also assumes that the importance of journal i is given by the direct importance of journals whose papers cite papers published in the journal, the direct importance of authors who published their papers in journal i, and the direct importance of papers published in journal i. In our model, described in Section 6, we assume that the importance of author i is given by the direct importance from the papers he/she writes, indirect importance from journals through his/her papers, and indirect importance from co-authors through papers they co-authored. We also assume that the importance of journal i is given by the direct importance from the papers published in it, as well as indirect importance from authors published papers in it, and indirect importance from other journals through papers. Therefore, the proportion of importance that authors and journals give to an author or a journal in the BDR model is more than that in our model. That is, for some problems, it will over-accumulate the importance that an author or a journal receives from authors and journals.

The flow of importance of the BDR three-class model and our three class model is illustrated by Figure 2.

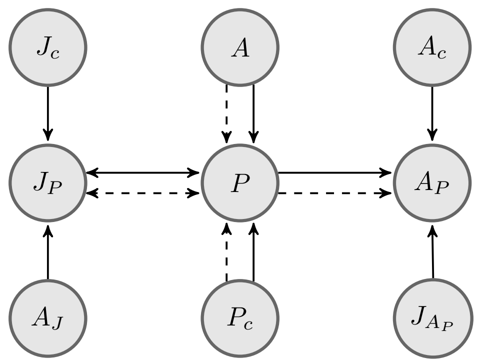

Figure 2: Comparison of the flow of importance of the BDR three-class model and our three-class model. Solid lines show the flow of importance in the BDR three-class model and dashed lines show the flow of importance in our three-class model. Here A denotes the author or set of co-authors of paper P, J_P denotes the journal where paper P is published, P_c denotes the set of papers that cite paper P, A_P stands for one of the authors of paper P, A_c stands for set of co-authors of author A_P, J_{A_P} denotes the set of journals where author A_P's papers are published, J_c denotes the set of journals that contain papers that cite papers published in journal J_P, and A_J denotes the set of authors who publish their papers in journal J_P.

4 One-class model

We will use the following notation:

- O_1: object class in the one-class model.
- n_1: number of objects (nodes) in O_1.
- $o_1, o_2, \ldots, o_{n_1}$: objects in the class O_1. For instance, o_i may stand for paper i or user i.

Suppose that we are interested in studying the ranking of the objects in O_1 in a homogeneous network. Let $H = [h_{ij}] \in \mathbb{R}^{n_1 \times n_1}$ be the adjacency matrix for the network, such that $h_{ij} = 1$ if there is a link from o_i to o_j, and $h_{ij} = 0$ otherwise. The importance of o_j is given by the j^{th} entry of the left Perron vector x of H. Thus, x satisfies $x^T H = \rho x^T$, where ρ is the spectral radius of H. For different kinds of problems, normalization of the rows of H may have to be considered. More details will be introduced in the examples of this section.

Some rows or columns of the adjacency matrix H may only have zero entries. This occurs when some node o_i does not have any links (edges) to other nodes, in which case o_i is said to be a *sink*, or when a node o_i does not have any links from other nodes, in which case o_i is referred to a *source*. The Perron vector is unique when H is nonnegative and irreducible, but this is not the case in the presence of a sink or source.

The adjacency matrix can be modified so that it is nonnegative and irreducible by allowing "teleporting." We consider three teleporting methods. They depend on a parameter $0 < \epsilon < 1$, which determines the probability of teleporting. The first two teleporting methods below are related to the teleporting used by the Google PageRank model [6, 16], while the last approach generalizes the one used by Bini et al. [2, 3].

- TM_1: For $\epsilon \in (0, 1)$, we define the new weighted adjacency matrix $\hat{H} = H + \epsilon \cdot [\mathbf{e}\mathbf{e}^T - I]$, where $\mathbf{e} = [1, 1, \ldots, 1]^T$. The identity matrix $I \in \mathbb{R}^{n_1 \times n_1}$ is subtracted since we do not allow self-loops.
- TM_2: For $\epsilon \in (0, 1)$, we introduce the weighted adjacency matrix $\hat{H} = H + \epsilon \cdot [\mathbf{e}\mathbf{e}^T - H - I]$. The matrix H is subtracted from the teleporting matrix because we allow teleporting only between nodes with no citation path.
- TM_3: For $\epsilon \in (0, 1]$, we introduce a dummy node o_{n_1+1} such that it has bi-directional links with all other nodes $o_1, o_2, \ldots, o_{n_1}$. This gives the new weighted adjacency matrix $\hat{H} \in \mathbb{R}^{(n_1+1)\times(n_1+1)}$, whose leading $n_1 \times n_1$ principal submatrix is H. Its last diagonal entry vanishes and the remaining entries of $\hat{H}$ in the last row and column equal ϵ.

We first illustrate the one-class model with a simple undirected network, in which the nodes represent users and the links represent friendships.

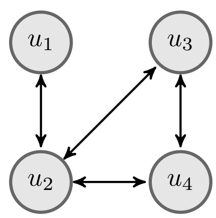

Figure 3: Undirected network of Example 4.1.

EXAMPLE 4.1. Consider the network of Figure 3 with four users. The associated adjacency matrix is given by

$$H = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}.$$

Thus, H is nonnegative and irreducible. Its left Perron vector

$$x^T = \begin{bmatrix} 0.2818 & 0.6116 & 0.5227 & 0.5227 \end{bmatrix}$$

shows that User 2 has the highest rank. This depends on that he/she has three friends. Users 3 and 4 have the same, but lower rank, because both of them have two friends, only. $\square$

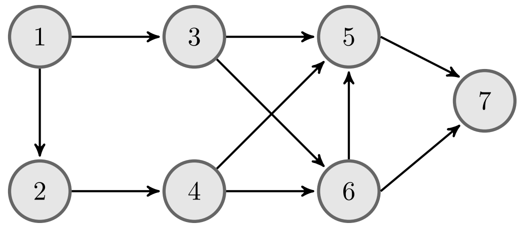

Figure 4: Directed network of Example 4.2.

EXAMPLE 4.2. Regard the directed network of Figure 4. The nodes represent papers and the edges stand for citations. Paper 7 is a sink since it does not cite any paper. To obtain a unique unit left Perron vector, we complement the network with a teleporting strategy. To understand the meaning of teleportation, imagine that a student is conducting a research project. He starts reading a paper that interests him and wants to look at other interesting papers. He can either pick a new paper by following the citation path of papers or randomly look for another paper (teleporting). The parameter ϵ stands for the probability of teleporting. Intuitively, we expect Paper 1 to be the least important paper, since it does not receive any citations. Papers 2 and 3 are more important than paper 1, since they receive one citation each from Paper 1. Paper 4 is more important than Papers 2 and 3, since it receives one citation from Paper 2, which is more important than Paper 1. Papers 6 and 7 receive two citations. Their rankings therefore are higher than the rankings of Papers 3 and 4. Moreover, Paper 7 is cited by Paper 5, which receives three citations. This makes Paper 7 more important than Paper 6. These rankings are illustrated by Table 3. The table shows these rankings to be independent of the teleportation method used.

TM$_1$: ϵ	P_1	P_2	P_3	P_4	P_5	P_6	P_7
0.1	0.1270	0.1966	0.1966	0.2347	0.5627	0.3634	0.6347
0.3	0.2251	0.2915	0.2915	0.3111	0.5218	0.4029	0.4980
0.5	0.2682	0.3236	0.3236	0.3351	0.4879	0.4043	0.4526
0.7	0.2924	0.3391	0.3391	0.3465	0.4660	0.4019	0.4310
0.9	0.3079	0.3480	0.3480	0.3532	0.4512	0.3992	0.4143

TM$_2$: ϵ	P_1	P_2	P_3	P_4	P_5	P_6	P_7
0.1	0.1356	0.2062	0.2062	0.2430	0.5619	0.3695	0.6207
0.3	0.2559	0.3150	0.3150	0.3287	0.4983	0.4047	0.4647
0.5	0.3137	0.3511	0.3511	0.3556	0.4455	0.3980	0.4143
0.7	0.3475	0.3671	0.3671	0.3682	0.4109	0.3889	0.3926
0.9	0.3696	0.3753	0.3753	0.3754	0.3871	0.3812	0.3815

TM$_3$: ϵ	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8
0.1	0.0254	0.0561	0.0561	0.0934	0.4571	0.2066	0.8301	0.2091
0.3	0.0835	0.1406	0.1406	0.1795	0.5082	0.3020	0.6367	0.4078
0.5	0.1247	0.1863	0.1863	0.2167	0.4837	0.3238	0.5236	0.5051
0.7	0.1523	0.2117	0.2117	0.2348	0.4535	0.3263	0.4562	0.5583
0.9	0.1715	0.2268	0.2268	0.2446	0.4279	0.3235	0.4138	0.5907

Table 3: Ranking of papers by using three teleportation methods and different values of ϵ. P_8 denotes the dummy paper.

Table 3 shows the relative importance of Papers 5 and 7 to depend on the value of the teleportation parameter ϵ. If $\epsilon > 0$ is small, the reader generally looks for a new paper by following the citation path, while if ϵ is large, the reader frequently ignores the citations and selects a random paper. For all of the three teleportation methods, Table 3 shows that as $\epsilon \to 0$, we obtain the ranking $P_7 > P_5 > P_6 > P_4 > P_3 = P_2 > P_1$. In particular, Paper 7 is more important than Paper 5, even though Paper 5 has three citations, while Paper 7 only has 2. When $\epsilon \to 1$, Table 3 shows the ranking $P_5 > P_7 > P_6 > P_4 > P_3 = P_2 > P_1$. Thus, when ϵ is large, and therefore the probability of teleporting is large, the relative importance of papers with many citations increases. □

The importance of a paper is determined by how many papers it is cited by, and in the BDR and our models also by the importance of the citing papers, but not by how many times a paper is cited in each one of the citing papers. However, when Paper 1 cites Paper 2 several times, the importance of Paper 2 for the development of Paper 1 is likely to be more important than if Paper 1 cites Paper 2 only once. For instance, the present paper cites the papers [2, 3] many times, because they are important for the development of the present paper. It therefore may be meaningful to equip the adjacency matrix with weights that grow with the number of citations from one paper to another.

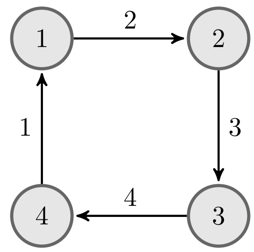

Figure 5: Directed network of Example 4.3.

EXAMPLE 4.3. Consider four papers $P_1, \ldots, P_4$, which cite each other several times. Specifically, P_i cites P_{i+1} $i+1$ times for $i = 1, 2, 3$, and P_4 cites P_1 once. The network is displayed by Figure 5, with the weights shown near each edge, and the corresponding adjacency matrix is given by

$$H = \begin{bmatrix} 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 4 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

It has the left Perron vector

$$x^T = \begin{bmatrix} 0.3489 & 0.3153 & 0.4273 & 0.7722 \end{bmatrix},$$

which shows Paper 4 to be most important. $\square$

We are interested in investigating how the left Perron vector changes if we add a new link from node o_r to node o_s. Similar results for row-stochastic matrices are shown by Bini et al. [2, 3]. We first consider directed networks, and subsequently will discuss undirected ones.

EXAMPLE 4.4. Regard the unweighted network associated with the network of Figure 5, i.e., all weights are set to one. The left Perron vector for the network is given by

$$x^T = \begin{bmatrix} 0.5 & 0.5 & 0.5 & 0.5 \end{bmatrix}.$$

Adding a citation (link) from Paper 2 to Paper 4 gives a new network with Perron vector

$$x^T = \begin{bmatrix} 0.5262 & 0.4311 & 0.3531 & 0.6424 \end{bmatrix}.$$

The importance of Paper 4 increases as expected. $\square$

The behavior of the left Perron vector in the above example is a consequence of the perturbation results formulated as Theorem 1 below. Theorem 1 is shown by Bini et al. [2, 3] in the special case when $\rho = \hat{\rho} = 1$.

LEMMA 1. *([9, 11]) Let the matrices $A = [a_{ij}] \in \mathbb{R}^{n\times n}$ and $\hat{A} = [\hat{a}_{ij}] \in \mathbb{R}^{n\times n}$ be nonnegative and irreducible, with spectral radii ρ and $\hat{\rho}$, respectively, and associated right Perron vectors $x = [x_1, \ldots, x_n]^T$ and $\hat{x} = [\hat{x}_1, \ldots, \hat{x}_n]^T$. Thus, $Ax = \rho x$ and $\hat{A}\hat{x} = \hat{\rho}\hat{x}$. Define the index sets $\Gamma_0 = \{1 \leqslant i \leqslant n : \hat{a}_{ij} = a_{ij}, \forall\, 1 \leqslant j \leqslant n\}$ and $\Gamma = \{1 \leqslant i \leqslant n : i \notin \Gamma_0\}$. Hence, Γ_0 contains the indices of the rows of $\hat{A}$ that are the same as the rows of A, while Γ contains the remaining indices. We assume that both sets Γ_0 and Γ are non-empty. Then the following inequalities hold:*

(a) *If $\hat{\rho} > \rho$, then $\frac{\hat{x}_k}{x_k} \leqslant \frac{\rho}{\hat{\rho}} \max\limits_{j\in\Gamma} \left(\frac{\hat{x}_j}{x_j}\right)$ for all k in Γ_0.*

(*b*) *If* $\hat{\rho} < \rho$, *then* $\frac{\hat{x}_k}{x_k} \geqslant \frac{\rho}{\hat{\rho}} \min_{j\in\Gamma}\left(\frac{\hat{x}_j}{x_j}\right)$ *for all* k *in* Γ_0.

(*c*) *If* $\hat{\rho} = \rho$, *then both inequalities above hold for all* k.

Let $A = [a_{ij}], B = [b_{ij}] \in \mathbb{R}^{n\times n}$. We write $A \geqslant 0$ if all $a_{ij} \geqslant 0$, and $A \leqslant B$ is $B - A \geqslant 0$.

LEMMA 2. *([15, Corollary 8.1.19]) Let the matrices* $A, B \in \mathbb{R}^{n\times n}$ *have spectral radii* $\rho(A)$ *and* $\rho(B)$, *respectively. If* $0 \leqslant A \leqslant B$, *then* $\rho(A) \leqslant \rho(B)$.

THEOREM 1. *Let the adjacency matrix* $H = [h_{ij}] \in \mathbb{R}^{n\times n}$ *be irreducible and assume that* $h_{rs} = 0$. *Let the entries of the adjacency matrix* $\hat{H} = [\hat{h}_{ij}] \in \mathbb{R}^{n\times n}$ *agree with the entries of* H, *except for* $\hat{h}_{rs} = 1$. *Assume that* H *and* $\hat{H}$ *have spectral radii* ρ *and* $\hat{\rho}$, *and associated left unit Perron vectors* $x = [x_1, \ldots, x_n]^T$ *and* $\hat{x} = [\hat{x}_1, \ldots, \hat{x}_n]^T$, *respectively. Thus,* $\rho x^T = x^T H$ *and* $\hat{\rho}\hat{x}^T = \hat{x}^T \hat{H}$. *Then*

$$\frac{\hat{x}_j}{x_j} \leqslant \frac{\rho}{\hat{\rho}}\frac{\hat{x}_s}{x_s} \leqslant \frac{\hat{x}_s}{x_s}, \qquad j \neq s, \tag{2}$$

and

$$\frac{\hat{x}_s}{x_s} \geqslant 1. \tag{3}$$

Proof. Let $e_j = [0, \ldots, 0, 1, 0, \ldots, 0]^T \in \mathbb{R}^n$ denote the j^{th} axis vector, and note that $\hat{H} = H + e_r e_s^T$. Thus, $\hat{H} \geqslant H$, and Lemma 2 yields $\hat{\rho} \geqslant \rho$. Let $A = H^T$ and $B = \hat{H}^T$. Then the matrices A and B differ in row s only, and $Ax = \rho x$ and $B\hat{x} = \hat{\rho}\hat{x}$. Lemma 1(a) now gives (2).

Assume that $\hat{x}_s < x_s$. It then follows from (2) that $\|\hat{x}\|_2 < \|x\|_2$, which contradicts that both $\hat{x}$ and x are unit vectors. Here and below $\|\cdot\|_2$ denotes the Euclidean vector norm. ■

The above theorem shows that if we add an extra link from object o_r to object o_s, then the rank of object o_s increases more than the rank of the other objects. This also holds for k-class models, for $k \geq 2$, because it is a rank-one perturbation. However, Theorem 1 does not carry over to the situation when more than one link is added to the network. This is illustrated by the following example.

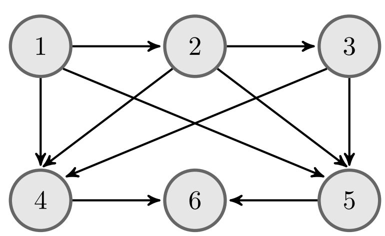

Figure 6: Directed network of Example 4.5.

EXAMPLE 4.5. Consider a network defined by 6 papers with citations according to Figure 6. Paper 6 is a sink. We used teleportation method 1 with $\epsilon = 10^{-5}$. This gives the left Perron vector

$$\hat{x}^T = \begin{bmatrix} 0.00009 & 0.00082 & 0.00673 & 0.06177 & 0.06177 & 0.99615 \end{bmatrix}.$$

Now, modify the network so that Paper 1 cites Papers 3 and 6. Thus, we introduce links from Paper 1 to Paper 3 and to Paper 6. The left Perron vector for the modified network is given by

$$\hat{x}^T_{new} = \begin{bmatrix} 0.000089 & 0.000795 & 0.007075 & 0.062975 & 0.062975 & 0.996001 \end{bmatrix}.$$

Element-wise division of the vector entries gives

$$\frac{\hat{x}_{new}^T}{\hat{x}^T} = \begin{bmatrix} 0.9822 & 0.9647 & 1.0506 & 1.0196 & 1.0196 & 0.9998 \end{bmatrix}.$$

Since both Papers 3 and 6 receive a new citation, one might expect their relative importance to increase more than the relative importance of the other papers. However, the rank-value of Paper 6 deceases while the rank-value of Papers 4 and 5 increase. □

In Example 4.5 the rank of one of the papers (Paper 3), which receives new citations, increases in rank more than the other papers. The following result shows this behavior to hold in general. Corollary 1 is shown by Bini et al. [2, 3] in the special case when $\rho = \hat{\rho} = 1$.

COROLLARY 1. *Let the adjacency matrix $H = [h_{ij}] \in \mathbb{R}^{n\times n}$ be irreducible and such that $h_{rs_1} = \ldots = h_{rs_\ell} = 0$. Assume that the entries of the adjacency matrix $\hat{H} = [\hat{h}_{ij}] \in \mathbb{R}^{n\times n}$ agree with the entries of H, except for $\hat{h}_{rs_1} = \ldots = \hat{h}_{rs_\ell} = 1$. Define the index sets $\Gamma = \{s_1, s_2, \ldots, s_\ell\}$ and $\Gamma_0 = \{1 \leqslant i \leqslant n : \ i \not\in \Gamma_0\}$. Let H and $\hat{H}$ have spectral radii ρ and $\hat{\rho}$, and associated left unit Perron vectors $x = [x_1, \ldots, x_n]^T$ and $\hat{x} = [\hat{x}_1, \ldots, \hat{x}_n]^T$, respectively. Then*

$$\frac{\hat{x}_i}{x_i} \leqslant \max_{j\in\Gamma} \frac{\hat{x}_j}{x_j} \qquad \forall\ i \in \Gamma_0 \tag{4}$$

and

$$\max_{j\in\Gamma} \frac{\hat{x}_j}{x_j} \geqslant 1. \tag{5}$$

Proof. The proof is similar to that of Theorem 1. Let $A = H^T$ and $\hat{A} = \hat{H}^T$. The matrices A and $\hat{A}$ differ in rows with indices in the set Γ. By Lemma 2, we have $\hat{\rho} \geqslant \rho$, and Lemma 1(a) yields

$$\frac{\hat{x}_i}{x_i} \leqslant \frac{\rho}{\hat{\rho}} \max_{j\in\Gamma} \frac{\hat{x}_j}{x_j} \qquad \forall\ i \in \Gamma_0,$$

which shows (4).

Assume that inequality (5) is violated. Then $0 < \hat{x}_j < x_j$ for $1 \leqslant j \leqslant n$. Therefore, x and $\hat{x}$ cannot both be unit vectors. This shows (5). ■

We remark that Theorem 1 and Corollary 1 carry over to k-class models, for $k \geq 2$, to be defined in Sections 5 and 6; the adjacency matrices for these models for $k = 2$ and $k = 3$ are shown in the right-hand side columns of Tables 1 and 2. The theorem and corollary do not hold for the BDR three-class model introduced in Table 2, because adding a new citation from a paper to another paper will modify the matrix FHF^T and, therefore, result in a rank-two perturbation. Corollary 1 also is useful when studying undirected networks, for which the adjacency matrix $H = [h_{ij}]$ is symmetric. Let $h_{ij} = 0$ with $i \neq j$. When setting $h_{ij} = 1$, we also set $h_{ji} = 1$ to preserve symmetry.

A converse to Theorem 1 can be shown, which addresses the situation when an edge is removed from a network. Consider the removal of an edge that points from node o_r to node o_s. We would expect this removal to decrease the rank of node o_s more than the rank of the other nodes; see Example 4.6 and Theorem 2 below. The new graph obtained when the edge has been removed has to be connected. A converse of Corollary 1 also can be shown. It addresses the situation when several edges that point from node o_r to nodes $o_{s_1}, \ldots, o_{s_\ell}$ are removed. This is discussed by Corollary 2 below.

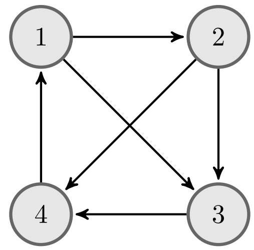

Figure 7: Directed network of Example 4.6.

EXAMPLE 4.6. Consider a network of four papers $P_1, \dots, P_4$ connected as shown by Figure 7. The unit left Perron vector for the network is given by

$$x^T = \begin{bmatrix} 0.4484 & 0.3213 & 0.5516 & 0.6526 \end{bmatrix}.$$

Removing a citation (link) from Paper 2 to Paper 4 gives a new network with left Perron vector

$$x^T = \begin{bmatrix} 0.4311 & 0.3531 & 0.6424 & 0.5262 \end{bmatrix},$$

The importance of Paper 4 decreases as expected. $\square$

THEOREM 2. *Let the adjacency matrix $H = [h_{ij}] \in \mathbb{R}^{n \times n}$ be irreducible and assume that $h_{rs} = 1$. Assume there exists at least two nonzero entries in the r^{th} row and the s^{th} column. Let the entries of the adjacency matrix $\hat{H} = [\hat{h}_{ij}] \in \mathbb{R}^{n \times n}$ agree with the entries of H, except for $\hat{h}_{rs} = 0$. Assume that H and $\hat{H}$ have spectral radii ρ and $\hat{\rho}$, and associated unit left Perron vectors $x = [x_1, \dots, x_n]^T$ and $\hat{x} = [\hat{x}_1, \dots, \hat{x}_n]^T$, respectively. Thus, $\rho x^T = x^T H$ and $\hat{\rho}\hat{x}^T = \hat{x}^T \hat{H}$. Then*

$$\frac{\hat{x}_j}{x_j} \geqslant \frac{\rho}{\hat{\rho}} \frac{\hat{x}_s}{x_s} \geqslant \frac{\hat{x}_s}{x_s}, \qquad j \neq s, \tag{6}$$

and

$$\frac{\hat{x}_s}{x_s} \leqslant 1. \tag{7}$$

Proof. If either nodes r or s become a *source* or *sink* after edge removal, we can always use one of the three teleporting methods introduced in Section 4 to make $\hat{H}$ irreducible. Let $e_j = [0, \dots, 0, 1, 0, \dots, 0]^T \in \mathbb{R}^n$ denote the j^{th} axis vector, and note that $\hat{H} = H - e_r e_s^T$. Thus, $\hat{H} \leqslant H$, and Lemma 2 yields $\hat{\rho} \leqslant \rho$. Let $A = H^T$ and $B = \hat{H}^T$. Then the matrices A and B differ in row s only, and $Ax = \rho x$ and $B\hat{x} = \hat{\rho}\hat{x}$. Lemma 1(b) now gives (6).

Assume that $\hat{x}_s > x_s$. It then follows from (7) that $\|\hat{x}\|_2 > \|x\|_2$, which contradicts that both $\hat{x}$ and x are unit vectors. $\blacksquare$

COROLLARY 2. *Let the adjacency matrix $H = [h_{ij}] \in \mathbb{R}^{n \times n}$ be irreducible and such that $h_{rs_1} = \ldots = h_{rs_\ell} = 1$. Assume that there exists at least two nonzero entries in the r^{th} row and j^{th} columns for $j = 1, \dots, \ell$. Let the entries of the adjacency matrix $\hat{H} = [\hat{h}_{ij}] \in \mathbb{R}^{n \times n}$ agree with the entries of H, except for $\hat{h}_{rs_1} = \ldots = \hat{h}_{rs_\ell} = 0$. Define the index sets $\Gamma = \{s_1, s_2, \dots, s_\ell\}$ and $\Gamma_0 = \{1 \leqslant i \leqslant n : \ i \neq \Gamma_0\}$. Let H and $\hat{H}$ have spectral radii ρ and $\hat{\rho}$, and associated left unit Perron vectors $x = [x_1, \dots, x_n]^T$ and $\hat{x} = [\hat{x}_1, \dots, \hat{x}_n]^T$, respectively. Then*

$$\frac{\hat{x}_i}{x_i} \geqslant \min_{j \in \Gamma} \frac{\hat{x}_j}{x_j} \qquad \forall\, i \in \Gamma_0 \tag{8}$$

and

$$\min_{j \in \Gamma} \frac{\hat{x}_j}{x_j} \leqslant 1. \tag{9}$$

Proof. If either nodes r or s become a *source* or *sink* after edge removal, we can always use one of the three teleporting methods introduced in Section 4 to make $\hat{H}$ irreducible. The proof is similar to that of Theorem 2. Let $A = H^T$ and $\hat{A} = \hat{H}^T$. The matrices A and $\hat{A}$ differ in rows with indices in the set Γ. By Lemma 2, we have $\hat{\rho} \leqslant \rho$. and Lemma 1(b) yields

$$\frac{\hat{x}_i}{x_i} \geqslant \frac{\rho}{\hat{\rho}} \min_{j\in\Gamma} \frac{\hat{x}_j}{x_j} \qquad \forall\ i \in \Gamma_0,$$

which shows (8).

Assume that inequality (9) is violated. Then $0 < x_j < \hat{x}_j$ for $1 \leqslant j \leqslant n$. Therefore, x and $\hat{x}$ cannot both be unit vectors. This shows (9). ∎

5 Two-class model

We will use the following notation in this section:

- O_1, O_2: object classes. Objects in these classes may represent papers and authors in a bibliographic network, and users and movies in a network from the Douban web site [10].
- n_i: number of objects (nodes) in the i^{th} object class, $i \in \{1, 2\}$.
- $o_{i1}, o_{i2}, \ldots, o_{in_i}$: nodes in the class O_i, $i \in \{1, 2\}$.
- ω_{ij} for $1 \leqslant i, j \leqslant 2$: the weight $\omega_{ij} \geqslant 0$ determines the influence of objects in class i have on the importance of objects in class j. We set $\omega_{ij} = 0$ if objects in the class O_i have no influence on the importance of objects in the class O_j; otherwise we set $\omega_{ij} > 0$.

	O_1	O_2
O_1	ω_{11}	ω_{12}
O_2	ω_{21}	ω_{22}

Table 4: Weights ω_{ij} for links between objects in the classes O_1 and O_2.

Consider the object classes $O_1 = \{o_{1i}\}_{i=1}^{n_1}$ and $O_2 = \{o_{2i}\}_{i=1}^{n_2}$. Let $H_{ij} = [h_{rs}^{(ij)}] \in \mathbb{R}^{n_i \times n_j}$ be an adjacency matrix such that $h_{rs}^{(ij)} = 1$ when there is an edge from object (node) o_{ir} to object (node) o_{js}, and $h_{rs}^{(ij)} = 0$ otherwise, for $1 \leqslant i, j \leqslant 2$. The most general form of our two-class model has an adjacency block matrix of the form

$$S = \begin{bmatrix} \omega_{11} H_{11} & \omega_{12} H_{12} \\ \omega_{21} H_{21} & \omega_{22} H_{22} \end{bmatrix} \in \mathbb{R}^{(n_1+n_2)\times(n_1+n_2)}. \tag{10}$$

In all examples, we choose the weights $\omega_{11} = \omega_{12} = \omega_{21} = 1$ and $\omega_{22} = 0$. This gives adjacency block matrices of the form shown in the right-hand side of Table 1.

The relative importance of the $n_1 + n_2$ nodes is determined by computing the left Perron vector $x^T = [x_{O_1}^T\ x_{O_2}^T]$ for S. The relative size of the entries of the subvector $x_{O_1} \in \mathbb{R}^{n_1}$ yields the relative importance of the objects o_{1i} of O_1, while the relative importance of the objects o_{2i} of O_2 is proportional to the size of the entries of the subvector $x_{O_2}^T$. The relative importance of the objects depends on whether the matrices H_{ij} have been normalized. This is illustrated by examples below.

NM_1	No normalization is applied.
NM_2	Column normalize H_{21} with respect to 1-norm so that the importance of $o_{1i} \in O_1$ is independent of the number of nodes $o_{2i} \in O_2$. Replace the matrix H_{21} in (10) by the normalized matrix. With this normalization the influence of a user is not proportional to the number of movies he/she reviews.
NM_3	Column normalization of H_{21} as above, and analogous row normalization of H_{12}. Replace the matrices H_{21} and H_{12} by the corresponding normalized matrices in (10). With this normalization the influence of user is not proportional to the number of movies he/she reviews. Moreover, the total amount of importance that a user gives to the movies he/she reviews is independent of the number of movies reviewed.

Table 5: Normalization methods.

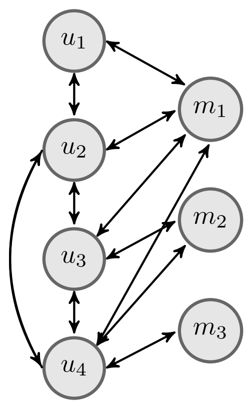

Figure 8: Directed network of Example 5.1.

EXAMPLE 5.1. Consider an example with 4 users $u_1, \ldots, u_4$ and 3 movies m_1, m_2, m_3. The relation between users are "friendships", which are modeled by undirected edges. An edge from u_i to m_j indicates that u_i reviewed m_j; the edge from m_j to u_i models that user u_i gets "importance" by reviewing movie m_j. Thus, all edges are undirected. We refer to movies for which our model gives relatively large components of the left Perron vector as "influential."

The importance of a user and movie depend on whether the columns of the matrix H_{12} or and the rows of matrix H_{21} are normalized. Three normalization methods are used in Table 6.

Method	u_1	u_2	u_3	u_4	m_1	m_2	m_3
$\mathbf{NM}_1$	0.2501	0.4532	0.4552	0.4840	0.4532	0.2592	0.1336
$\mathbf{NM}_2$	0.3445	0.5619	0.4333	0.4135	0.4355	0.1224	0.0476
$\mathbf{NM}_3$	0.3216	0.5087	0.4000	0.3794	0.5087	0.2463	0.1199

Table 6: Ranking of users and movies by using $\mathbf{NM}_1$, $\mathbf{NM}_2$, $\mathbf{NM}_3$

When using NM_1, user 4 is most influential, because he/she has two friends and writes reviews for three movies; movie 1 is the most influential movie because it receives 4 reviews. For this normalization method, the influence of a user is proportional to the number of friends he/she has and the number of reviews he/she writes. With NM_2, user 2 is most influential, because he/she has three friends; again movie 1 is most influential. For $\mathbf{NM}_2$, the influence of a user is proportional to the number of friends, but independent of the number of movies he/she reviews. When using $\mathbf{NM}_3$, the total amount of influence that a user gives to each movie is the same. A user with many friends is likely to be most influential. $\square$

6 k-class model

This section uses the following notation:

- $O_1, \ldots, O_k$: object classes. Objects in these classes may represent papers, authors, and journals in a bibliographic network, and users, movies, and groups in a network from [10].
- n_i: number of nodes in the i^{th} object class, $1 \leqslant i \leqslant k$.
- $o_{i1}, o_{i2}, \ldots, o_{in_i}$: objects (nodes) in the i^{th} object class, $1 \leqslant i \leqslant k$.
- ω_{ij} for $1 \leqslant i, j \leqslant k$: the weight $\omega_{ij} \geqslant 0$ determines the influence of objects in the class O_i have on the importance of the objects in the class O_j.

Regard the object classes $O_j = \{o_{ji}\}_{i=1}^{n_j}$, $1 \leqslant j \leqslant k$, and let $H_{ij} = [h_{rs}^{(ij)}] \in \mathbb{R}^{n_1 \times n_j}$ be an adjacency matrix such that $h_{rs}^{(ij)} = 1$ when there is an edge from object o_{ir} to object o_{js}, and $h_{rs}^{(ij)} = 0$ otherwise. The most general form of our k-class model has an adjacency block matrix of the form

$$S = \begin{bmatrix} \omega_{11} H_{11} & \omega_{12} H_{12} & \ldots & \omega_{1k} H_{1k} \\ \omega_{21} H_{21} & \omega_{22} H_{22} & \ldots & \omega_{2k} H_{2k} \\ \ldots & \ldots & \ldots & \ldots \\ \omega_{k1} H_{k1} & \omega_{k2} H_{k2} & \ldots & \omega_{kk} H_{kk} \end{bmatrix} \in \mathbb{R}^{(n_1 + \ldots + n_k) \times (n_1 + \ldots + n_k)}.$$

In all our examples, we use the weights $\omega_{ij} = 1$ if $i = 1$ or $j = 1$, and $\omega_{ij} = 0$ otherwise. When $k = 3$ this yields an adjacency matrix of the form depicted in the right-hand side of Table 2. To determine the importance of the objects o_{ij}, we compute the left unit Perron vector x for S. Dividing this vector into subvectors $x^T = [x_{O_1}^T \quad x_{O_2}^T, \ldots, x_{O_k}^T]$ such that $x_{O_j} \in \mathbb{R}^{n_j}$, $1 \leqslant j \leqslant k$, makes it possible to determine the importance of the objects in each class; the importance of the objects in O_j is determined by the relative size of the entries of the subvector x_{O_j}.

7 Numerical experiments

To show that our model is practical, we applied it to networks from [10] and bibliographic networks defined by real data. Example 7.1 discusses the ranking of objects of a four-class network from [10] with object classes users, movies, groups, and locations. The data for this network can be downloaded from the HIN Resource Dataset [10, 25]. Example 7.2 is concerned with a two-class bibliographic network defined by real data downloaded from the CiteSeer website.

EXAMPLE 7.1. The data set from the Douban web site [10] includes 3022 users, 6971 movies, 2269 groups, and 244 locations. Thus, this is a four-class model. Intuitively, we expect the user with most friends, the movie with most reviews, the group with most users, and the location with most users to be the most influential ones in each object class. The relative importance of the 12506 nodes is determined by computing the left Perron vector $x^T = [x_{\mathrm{U}}^T \quad x_{\mathrm{M}}^T \quad x_{\mathrm{G}}^T \quad x_{\mathrm{L}}^T]$. The relative importance of these objects depends on the normalization of the matrices H_{ij}. In particular, the normalization methods for two-class models discussed in Section 5 have to be generalized to be applicable to the four-class model of this example. We let NM_i^3 denote normalization methods for the group objects; see Table 7. For location objects, in the fourth class, we introduce an adjacency matrix $H_{41} = [h_{ij}^{41}]$, such that $h_{ij}^{41} = 1$ if user j is at location i, and $h_{ij}^{41} = 0$ otherwise. No normalization has to be applied to the matrix H_{41}, because it is impossible for user i to appear at more than one location simultaneously, that is, the importance of a user is not proportional to the number of locations. Furthermore, no normalization has to be applied to the matrix H_{14}, because we assume that the proportion of importance that each user gives to a location is the same. We denote the normalization method for locations by NM_1^4, i.e., no normalization is applied, for the object location in this paper; see Table 7.

NM_1^3	No normalization is applied.
NM_2^3	Column normalize H_{31} with respect to the 1-norm so that the importance of $o_{1i} \in O_1$ is independent of the number of nodes $o_{3i} \in O_3$. Replace the matrix H_{31} by the normalized matrix. With this normalization the influence of a user is not proportional to the number of groups he/she belongs to.
NM_3^3	Column normalization of H_{31} as above, and analogous row normalization of H_{13}. Replace the matrices H_{31} and H_{13} by the corresponding normalized matrices. With this normalization the influence of a user is not proportional to the number of groups he/she belongs to. Moreover, the total amount of importance that a user gives to the groups is independent of the number of groups he/she belongs to.
NM_1^4	No normalization is applied.

Table 7: Updated Normalization methods for the third object group and fourth object location

User ID	num. of friends	User ID	num. of movies	User ID	num. of groups
2070	36	2756	183	372	18
1116	25	546	181	636	17
1037	14	38	171	2444	17
2483	13	851	171	1307	13
1011	10	1088	171	2781	13

Table 8: *Columns 1, 3, 5 list the top five users in decreasing order based on the number of friends the user has, the number of movies the user reviews, and number of groups the user belongs to, respectively.*

Group ID	num. of users	Movie ID	num. of reviews	Location ID	num. of users
1867	61	1728	403	213	410
220	26	3649	381	93	292
774	26	223	378	142	108
1248	26	2135	375	68	97
1607	25	6691	375	54	96

Table 9: *Columns 1, 3, 5 list the top five objects among groups, movies, and locations in decreasing order based on the number of users the group has, the number of reviews the movie receives, and number of users the location has, respectively.*

The importance of a user, movie, and group depend on whether the columns of the matrices H_{12}, H_{13} and the rows of the matrices H_{21}, H_{31} are normalized. Three normalization methods are described in Table 10.

Mtds	$\mathrm{NM}_1 + \mathrm{NM}_1^3 + \mathrm{NM}_1^4$				$\mathrm{NM}_2 + \mathrm{NM}_2^3 + \mathrm{NM}_1^4$				$\mathrm{NM}_3 + \mathrm{NM}_3^3 + \mathrm{NM}_1^4$			
Rank	User	Movie	Group	Loca.	User	Movie	Group	Loca.	User	Movie	Group	Loca.
1	2756	1728	1867	213	2070	1728	1867	213	2070	1728	774	213
2	47	2135	1607	93	1011	3024	774	93	2427	3649	1679	93
3	1242	223	1248	142	2031	3649	374	142	2031	3024	1073	68
4	913	6370	774	68	2483	2135	1381	68	1011	2135	1381	142
5	779	3649	1641	54	2427	6370	2065	54	2483	637	1777	183

Table 10: Ranking of users, movies, groups, and locations.

When using the normalizations $\mathrm{NM}_1 + \mathrm{NM}_1^3 + \mathrm{NM}_1^4$, the influence of a user is proportional to the number of friends, movies he/she reviews, and groups he/she belongs to. The influence of a movie is determined by the number of reviews it received, the influence of a group is determined by the number of members it has, and the influence of the location is determined by the number of users at the location. Tables 8, 9, and 10 show that user 2756 reviewed 183 movies, user 2070 has 36 friends, and user 372 belongs to 18 groups. Moreover, user 2756 is most influential, because 183 is larger than 36 and 18. Movie 1728 is most the influential, because it received 403 reviews, group 1867 is the most influential, since it has 61 members, and location 213 is most the important one, because it has 410 users. With the normalization $\mathrm{NM}_2 + \mathrm{NM}_2^3 + \mathrm{NM}_1^4$, the influence of a user is proportional to the number of friends, but independently of the number of movies he/she reviewed, and the number of groups he/she joins. The influence of a movie, a group, and the location is determined by the number of reviews it received, and the number of users the group or the location has. From Table 10, we can see that user 2070 now is most influential, since he/she has 36 friends. Movie 1728, group 1867, and location 213 are still the most influential ones. When instead using the normalization $\mathrm{NM}_3 + \mathrm{NM}_3^3 + \mathrm{NM}_1^4$, the influence of a user is primarily based on the number of friends he/she has, and the total amount of influence that a user gives to each movie and group is the same, that is, the influence of movie i and group j is based on the influence of the user who reviews movie j and joins group j, rather than the number of reviews movie i received and the number of members group j has. The influence of location is based on the number of users it has and on the influence of these users. Table 10 shows user 2070 and movie 1728 to still be the most influential ones, but group 774 becomes the most influential group, because it receives the influence from

26 users and some of the users join the most influential group and review the most influential movies. We conclude that the choice of normalization is important for the ranking of objects of multi-class networks. □

EXAMPLE 7.2. We consider a two-class model that uses the same data as Bini et al. [2, 3]. The data includes 716800 papers and 410930 authors. Since some of the papers are isolated and some of the papers do not have information about authors, we first removed isolated papers and papers without author information. We then have a data set with 358906 papers and 257650 authors.

For our one-class model, we obtain an adjacency matrix $H \in \mathbb{R}^{358906\times 358906}$ and applied TM_3 with weight $\epsilon = 0.1$. Table 11 shows the top five papers. Results for the BDR model are displayed in Table 12. The BDR model and our model give related, but different rankings.

Paper	Position	Number of Citations
Bryant–Boolean Function Manipulation	1	1640
Jacob, Karels–Congestion Avoidance and Control	4	1131
Kirkpatrick, Gelatt, Vecchi–Simulated Annealing	2	1344
Rivest, Shamir, Adleman–Digital Signatures	3	1219
Floyd, Jaconson–Detection Gateways	7	1024

Table 11: One-class model of Section 4 with TM_3 for $\epsilon = 0.1$. *The papers are shown in decreasing rank order. The first column shows the titles and authors of the papers, the second column displays the position of the papers ordered by decreasing number of citations received, and the last column shows the number of citations each paper received.*

Paper	Position	Number of Citations
Jacob, Karels–Congestion Avoidance and Control	4	1131
Diffie, Hellman–New Directions in Cryptography	31	556
Rivest, Shamir, Adleman–Digital Signatures	3	1219
Bryant–Boolean Function Manipulation	1	1640
Kirkpatrick, Gelatt, Vecchi–Simulated Annealing	2	1344

Table 12: BDR one-class model. *The papers are shown in decreasing rank order. The first column shows the titles and authors of the papers, the second column displays the position of the papers ordered by decreasing number of citation received, and the last column shows the number of citations each paper received.*

For the two-class model, we have the adjacency matrices $H \in \mathbb{R}^{358906\times 358906}$ and $K \in \mathbb{R}^{257650\times 358906}$. We used TM_3 with $\epsilon = 0.1$, and normalization NM_3. We introduced a *dummy author*, who writes the *dummy paper*. Table 13 shows the five highest ranked papers.

Author	Num. Pap.	Paper	Num. Cit.
Douglas C. Schmidt	329	Bryant–Boolean Function Manipulation	1640
Patrick C Hew	103	Jacob, Karels–Congestion Avoidance and Control	1131
Fachbereich Informatik	507	Kirkpatrick, Gelatt, Vecchi–Simulated Annealing	1344
Sally Floyd	91	Rivest, Shamir, Adleman–Digital Signatures	1219
Aniruddha Gokhale	58	Floyd, Jacob–Detection Gateways	1024

Table 13: Two-class model of Section 5 with TM_3 with weight $\epsilon = 0.1$. *The first column shows authors in decreasing rank order, and the second column shows the number of papers each author has written. The papers in the third column are identified by titles and authors, and displayed in decreasing rank order. The last column shows the number of citations each paper received.*

Rank	Author	num.pap.	Paper	num.cit.
1	Randal Bryant	83	Kirkpatrick, Gelatt, Vecchi–Simulated Annealing	1344
2	Sally Floyd	91	Bryant–Boolean Function Manipulation	1640
3	John K. Ousterhout	23	Rivest, Shamir, Adleman–Digital Signatures	1219
4	Dennis M. Ritchie	6	Canny–Computational Approach to Edge Detection	834
5	Timothy H. Harrison	17	Floyd, Jacobson–Detection Gateways	1024
9	Douglas C. Schmidt	329		

Table 14: BDR two-class model. *The second column shows authors in decreasing rank order, and the third column displays the number of papers each author has written. The papers in the fourth columns are identified by titles and authors, and shown in decreasing rank order. The last column displays the number of citations each paper received.*

The ordering determined by the one-class model of Section 4 and the BDR one-class model are quite similar. Papers that receive more citations have higher rank than those that receive fewer citations. The difference in the ranking determined by these models stems from the fact that in the BDR model the importance that paper i distributes to the papers it cites is scaled by the total number of papers in the reference list of paper i; in our model, the importance of a reference is independent of the number of papers in the list of references. We observe that the ranking of papers we obtained with the two-class model of Section 5 is consistent with the ranking determined by the one-class model of Section 4. The authors in the BDR two-class model receive a higher proportion of importance from other authors than the authors in our two-class model. This results in the difference of the author rankings of these models. We finally remark that if we adjust the weights ω_{ij}, then different rankings will result. $\square$

We conclude this section with some comments on the tensor-based approach by Ng et al. [21]. These authors study the probability that a "random surfer" reaches a node or uses an edge. Suitable normalizations give probability tensors for the nodes and edges. Zero "rows" or "columns" of the tensors are replaced by "rows" or "columns," respectively, in which each entry has the same value. This secures that the tensors are irreducible. The stationary probability distributions for the nodes and edges determine the importance of the nodes and edges, respectively. Other approaches to secure irreducibility, more in line with those used by Bini et al. [2, 3], could also be used, such as introducing dummy entities (e.g., dummy authors and dummy papers). A comparison of the tensor method by Ng et al. [21] with the 2-class model of the present paper when applied to a small example with four authors and six papers resulted in the same ranking, but with different "importance scores" for the authors and papers. We conclude that the method proposed by Ng et al. [21] provides an alternative to our approach, which is more closely related to the technique by Bini, Del Corso, and Romani [2, 3]. The relative advantages of these approaches requires further investigation and will depend on the networks and applications considered.

8 Conclusion

This paper proposes new models for the analysis of multi-class or heterogeneous networks, in particular for the ranking of nodes based on their importance or centrality. The use of block adjacency matrices provides a systematic approach to the analysis of multi-class networks, in which the influence of nodes in one class on the relative importance ranking of members of another class is carried by transitivity through the explicit relationships in the model, as expressed by the powers of the matrix. For example, it becomes unnecessary to explicitly include co-authorship relationships among authors when trying to rank articles by importance, as the co-authorship relationships arise naturally in the powers of the block adjacency matrix,

For simplicity, the perturbation results have been stated and proved for the one-class model, but they in fact hold for all k-call models for $k \in \mathbb{N}$. The results establish that the chosen notion of importance, based on the Perron eigenvector, satisfies a kind of local monotonicity, namely that when a single edge is added, the rank of the node that is receiving the influence cannot decrease.

Weights and teleportation models affect the ranking in a transparent manner. For a one-class bibliographic network, the importance of a paper is determined by the number of papers it is cited by, and the importance of the citing papers. The importance of a paper is also determined by the number of times a paper is cited in each one of the citing papers if the bibliographic network is weighted. As the probability of teleporting increases, the relative importance of papers with many citations increases if it is a teleportation model. When authorships and publications are considered, the importance of a paper is also determined by the importance of its author and the importance of the journal where it is published, but independent of the number of authors it has.

In computational terms, using block-structured adjacency matrices can lead to matrices of very large dimensions. However, these matrices will in general be very sparse, so in many applications neither the computations nor the storage become too demanding. We note that large matrices do not have to be stored simultaneously in fast memory, only evaluations of matrix-vector products are required and this can be carried out in a piecemeal fashion.

Future work will include the use further information about the components of the network, including node and edge weights, as well as time dependence.

Acknowledgment

The authors would like to thank Francesco Romani for sharing the data set of the bibliographic network with us. They also would like to thank Michael Ng and the referees for comments. Work by the authors was supported in part by NSF grant DMS-1720259.

References

[1] A. Barrat, M. Barthelemy, and A. Vespignani, *Weighted evolving networks: Coupling topology and weight dynamics*, Physical Rev. Lett., 92 (2004), 228701.

[2] D. A. Bini, G. M. Del Corso, and F. Romani, *Evaluating scientific products by means of citation-based models: A first analysis and validation*, Electron. Trans. Numer. Anal., 33 (2008), pp. 1–16.

[3] D. A. Bini, G. M. Del Corso, and F. Romani, *A combined approach for evaluating papers, authors and scientific journals*, J. Comput. Appl. Math., 234 (2010), pp. 3104–3121.

[4] P. Bonacich, *Factoring and weighting approaches to status scores and clique identification.* J. Math. Sociol., 2 (1972), pp. 113–120.

[5] C. Brezinski and M. Redivo–Zaglia, *Rational extrapolation for the PageRank vector*, Math. Comp., 77 (2008), pp. 1585–1598.

[6] S. Brin and L. Page, *The anatomy of a large-scale hypertextual web search engine*, Comput. Networks ISDN Systems, 30 (1998), pp. 107–117.

[7] S. Cipolla, M. Redivo–Zaglia, and F. Tudisco, *Shifted and extrapolated power methods for tensor ℓ^p-eigenpairs*, Electron. Trans. Numer. Anal., 53 (2020), pp. 1–17.

[8] G. M. Del Corso and F. Romani, *Versatile weighting strategies for a citation-based research evaluation model*, Bull. Belg. Math. Soc. Simon Stevin, 16 (2009), pp. 723–743.

[9] E. Dietzenbacher, *Perturbation of matrices: A theorem on the Perron vector and its applications to input-output models*, J. Econom., 48 (1988), pp. 389–412.

[10] Douban Data. `http://shichuan.org/HIN_dataset.html`

[11] L. Elsner, C. R. Johnson, and M. Neumann, *On the effect of the perturbation of a nonnegative matrix on its Perron eigenvector*, Czechoslovak Math. J., 32 (1982), pp. 99–109.

[12] E. Estrada, *The Structure of Complex Networks: Theory and Applications*, Oxford University Press, Oxford, 2011.

[13] E. Garfield, *Citation analysis as a tool in journal evaluation*, Science, 178 (1972), pp. 471–479.

[14] C. Godsil and G. F. Royle, *Algebraic Graph Theory*, Springer, New York, 2013.

[15] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, 1985.

[16] A. N. Langville and C. D. Meyer, *Google's Pagerank and Beyond: The Science of Search Engine Rankings*, Princeton University Press, Princeton, 2006.

[17] J. Law and P. Lodge, *Science for Social Scientists*, Macmillan Press, London, 1984.

[18] X. Li, M. K. Ng, and Y. Ye, HAR: hub, authority and relevance scores in multi-relational data for query search, in Proceedings of the 2012 SIAM International Conference on Data Mining, eds. J. Ghosh, H. Liu,I Davidson, C. Domeniconi, and C. Kamath, SIAM, Philadelphia, 2012, pp. 141–152.

[19] M. E. J. Newman, *Analysis of weighted networks*, Phys. Rev. E 70, no. 5 (2004), 056131.

[20] M. E. J. Newman, *Networks: An Introduction*, Oxford University Press, Oxford, 2010.

[21] M. K. Ng, X. Li, and Y. Ye, MultiRank, Co-ranking for objects and relations in multi-relational data, in Proc. 17th ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, 2011, pp. 1217–1225.

[22] M. Ng, L. Qi, and G. Zhou, Finding the largest eigenvalue of a nonnegative tensor, SIAM J. Matrix Anal. Appl., 31 (2009), pp. 1090–1099.

[23] Y. Sun, Y. Yu, and J. Han, *Ranking-based clustering of heterogeneous information networks with star network schema*, Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Paris, France, 2009, pp. 797–806.

[24] S. H. Yook, H. Jeong, A.-L. Barabási, and Y. Tu, *Weighted evolving networks*, Phys. Rev. Lett., 86 (2001), pp. 5835–5838.

[25] J. Zheng, J. Liu, C. Shi, F. Zhuang, J. Li, and B. Wu, *Dual similarity regularization for recommendation*, In: J. Bailey, L. Khan, T. Washio, G. Dobbie, J. Huang, R. Wang (eds), Advances in Knowledge Discovery and Data Mining. PAKDD, Lecture Notes in Computer Science, vol. 9652. Springer, Cham, 2016, pp. 542–554.