Manifold Learning for Rank Aggregation

Shangsong Liang
KAUST
Thuwal, Saudi Arabia
shangsong.liang@kaust.edu.sa

Ilya Markov
University of Amsterdam
Amsterdam, The Netherlands
i.markov@uva.nl

Zhaochun Ren
JD AI
Beijing, China
renzhaochun@jd.com

Maarten de Rijke
University of Amsterdam
Amsterdam, The Netherlands
derijke@uva.nl

ABSTRACT

We address the task of fusing ranked lists of documents that are retrieved in response to a query. Past work on this task of rank aggregation often assumes that documents in the lists being fused are independent and that only the documents that are ranked high in many lists are likely to be relevant to a given topic. We propose manifold learning aggregation approaches, ManX and v-ManX, that build on the cluster hypothesis and exploit inter-document similarity information. ManX regularizes document fusion scores, so that documents that appear to be similar within a manifold, receive similar scores, whereas v-ManX first generates virtual adversarial documents and then regularizes the fusion scores of both original and virtual adversarial documents. Since aggregation methods built on the cluster hypothesis are computationally expensive, we adopt an optimization method that uses the top-k documents as anchors and considerably reduces the computational complexity of manifold-based methods, resulting in two efficient aggregation approaches, a-ManX and a-v-ManX. We assess the proposed approaches experimentally and show that they significantly outperform the state-of-the-art aggregation approaches, while a-ManX and a-v-ManX run faster than ManX, v-ManX, respectively.

CCS CONCEPTS

• **Information systems → Rank aggregation**; • **Computing methodologies → Dimensionality reduction and manifold learning**;

KEYWORDS

Ad hoc retrieval; rank aggregation; manifold learning

1 INTRODUCTION

Rank aggregation, also known as data fusion [17, 18, 30], is an important technique in information retrieval. Rank aggregation combines multiple ranked lists of documents retrieved from a corpus in response to a query by multiple retrieval algorithms. The ranked lists can be produced by any retrieval approach, using different ranking functions and multiple query and/or document representations [4, 17]. The combination of several retrieval approaches is assumed to improve retrieval effectiveness of the final fused ranked list of documents.

Past work on data fusion mostly assumes that documents in the lists being combined are independent and that only documents that are ranked high in many lists, are likely to be relevant to a given query [17, 30, 31]. However, the cluster hypothesis states that *documents in the same intrinsic structure, i.e., cluster or manifold, are likely to have a similar degree of relevance to the same information need underlying a given query* [17, 38]. This idea has been successfully applied to many ranking problems in information retrieval and data mining [4, 7, 36, 39]. In data fusion the cluster hypothesis has only been used to a limited extent [16] and with a negative impact on efficiency.

We propose a novel **Man**ifold-based data fusion approach, ManX, which (1) builds on a generic data fusion method **X**, and (2) lets similar documents provide support to each other by using inter-document similarities within a global manifold of documents being fused. Our manifold-based data fusion technique, ManX, is computationally demanding at two stages: in graph construction and in fusion score regularization. Therefore, we adopt an efficient design of the adjacency matrix for graph construction [24, 36], which supports document fusion for large datasets. Using this adjacency matrix, we propose a more efficient version of ManX, a-ManX, where the top-k documents from the ranking produced by an underlying fusion method X are assumed to be relevant and are used as **a**nchors to represent fusion scores of other documents.

Many machine learning models are vulnerable to adversarial data [33]. To further improve the performance of rank aggregation, we propose a virtual adversarial manifold learning algorithm, v-ManX, and an efficient version that utilizes anchor documents, a-v-ManX. Our proposed virtual adversarial manifold learning algorithms first generate a virtual adversarial document for each original document, then regularize the model so that given a document, the models will produce the same output distribution as they produce on an adversarial perturbation of that document. They improve the robustness to virtual adversarial documents and the generalization performance for original documents, thus enhancing the performance of manifold learning for rank aggregation.

To evaluate the effectiveness and efficiency of ManX and a-ManX, we conduct experiments using retrieval runs (at the Text REtrieval

WWW 2018, April 23–27, 2018, Lyon, France

ACM ISBN 978-1-4503-5639-8/18/04.
https://doi.org/10.1145/3178876.3186085

Conference (TREC),[1] a ranked list of documents is also called a run) submitted to TREC-3 [12], TREC-10 [13] and TREC-12 [34]. Experimental results confirm the theoretical findings, the effectiveness and efficiency of the proposed methods.

Our contributions in this paper are:

(1) We propose novel rank aggregation approaches, ManX, a-ManX, v-ManX and a-v-ManX, that exploit the manifold structure of documents being fused.

(2) We propose two virtual adversarial learning algorithms, v-ManX and a-v-ManX for rank aggregation.

(3) We propose a new virtual adversarial construction algorithm in our v-ManX and a-v-ManX algorithms.

(4) We propose an efficient design of the adjacency matrix and the anchor-based method to reduce the computational complexity of a-ManX and a-v-ManX.

(5) Through extensive experiments we show that the proposed ManX method outperforms the state-of-the-art data fusion techniques in terms of effectiveness, while a-ManX also considerably improves data fusion efficiency.

The remainder of the paper is organized as: § 2 reviews existing data fusion techniques and manifold-based algorithms. § 3 lists preliminaries. § 4 presents our manifold-based data fusion methods, ManX and a-ManX. § 5 describes the experimental setup, while § 6 discusses our experimental results. We conclude in § 7.

2 RELATED WORK

Three types of research relate to our work: rank aggregation, manifold-based algorithms and adversarial learning algorithms.

2.1 Rank aggregation

A large number of rank aggregation algorithms have been proposed. Well-known examples include the CombSUM data fusion family [30], Borda data fusion [1], supervised rank aggregation [27], λ-Merge [31], cluster-based data fusion [16], fusion for diversification [18], and, more recently, an aggregation algorithm that learns joint models on both lists and object features [2] and rule-based aggregation [3].

The state-of-the-art fusion method that we use for comparison is the cluster-based approach, ClustFuse, proposed in [16]. ClustFuse uses a combination of a fusion method, like CombSUM, and cluster-based retrieval, thus building on the cluster hypothesis. However, ClustFuse utilizes the nearest-neighbor clustering approach, which considers only the *local* similarity between documents [32, 38]. We argue that considering the *global* similarity within a document manifold will allow us to use the full power of the cluster hypothesis, which will further improve the performance of data fusion. To validate this intuition, we propose a number of manifold-based aggregation methods. To the best of our knowledge, ours is the first attempt to utilize manifolds in rank aggregation.

2.2 Manifold-based algorithms

Many manifold-based algorithms have been proposed, for a range of problems in applications. Recent ones include neural networks-based manifold learning [8]. In [38], manifold-based algorithms are used for document classification. In [10, 40], the authors use manifold-based algorithms for recognizing handwritten digits. In [26] manifold-based algorithms are used for video prediction, in [35] for detecting collective motion, and in [8] for face recognition.

We propose to use manifolds to regularize document scores in data fusion. Our manifold algorithms differ from previous ones [7, 24, 36, 39] by introducing virtual perturbation to documents, which allows us to significantly improve the performance. Since the computation of regularized scores is expensive we also propose an efficient version of manifold-based data fusion that uses the top-k documents as anchors. To the best of our knowledge, we are the first to utilize top-k ranked documents for efficient manifold learning.

2.3 Virtual adversarial learning algorithms

Adversarial learning is the process of training a model to correctly label both unmodified data and adversarial data [9, 11, 33]. It improves not only robustness to adversarial data, but also generalization performance for original data.

Virtual adversarial learning [28, 29] extends the idea of adversarial learning to the semi-supervised regime and unlabeled data. This is done by regularizing the model so that given an example, the model will produce the same output distribution as it produces on an adversarial perturbation of that example. One key to the success of virtual adversarial learning is the way proper virtual adversarial data is generated. Miyato et al. [28, 29] resort to an iteration method and finite difference method to approximately generate *local* virtual adversarial data, where "local" indicates that each virtual adversarial example is generated by considering its own original example only but not other data. Unlike previous adversarial learning algorithms where each adversarial example of the data is generated locally, our virtual adversarial algorithms generate each virtual adversarial document *globally* by considering not only the original itself but all the documents for adversarial perturbation construction. In addition, our two virtual adversarial manifold learning algorithms are unsupervised, compared to any of the existing adversarial learning algorithms that are either supervised or semi-supervised [9, 28, 29]. See [11] for a more thorough review of adversarial learning methods. To the best of our knowledge, we are the first to globally generate virtual adversarial data in adversarial learning, and the first to utilize virtual adversarial perturbation with manifolds for rank aggregation.

3 PRELIMINARIES

We detail the task we address and recall standard fusion algorithms that most state-of-the-art fusion methods, including ours, build on.

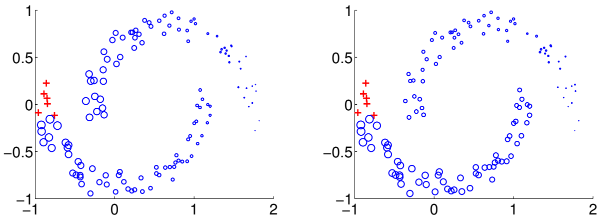
3.1 Problem formulation

We begin by defining the data fusion task that we address. The task is: given a query q and a set of ranked lists of documents $\mathcal{L}_1, \ldots, \mathcal{L}_m$, produced in response to a query q by m different retrieval systems, combine documents contained in the given lists by a data fusion method X into a single ranked list $\mathcal{L}_f$. The aggregation algorithm X is essentially a function f_{X} that satisfies:

$$q, \mathcal{L}_1, \mathcal{L}_2, \ldots, \mathcal{L}_m \xrightarrow{f_{\text{X}}} \mathcal{L}_f.$$

The goal of the task is to improve the performance of the final fused list $\mathcal{L}_f$ over the expected performance of the input lists.

[1] http://trec.nist.gov.



(a)

(b)

Figure 1: Illustration of manifold-based fusion model. (a) Ranking generated by a nearest-neighbor-based technique. (b) Ideal ranking generated by a manifold-base technique.

3.2 Standard data fusion

The CombSUM family is a set of simple unsupervised standard methods for data fusion [30]. They assume that documents that are ranked high in many input result lists should also be ranked high in the final fused list. The methods of the CombSUM family are among the best performing unsupervised data fusion techniques, reaching performance levels of sophisticated supervised approaches [17, 18].

Given a document d and a query q, CombSUM [17, 30] calculates the fusion score $f_{\mathrm{X}}(d;q)$ based on the retrieval score (or rank if no retrieval score is available) of d in given lists $\mathcal{L}_1, \ldots, \mathcal{L}_m$. CombSUM sums list-specific document scores:

$$f_{\mathrm{CombSUM}}(d;q) := \sum_{\mathcal{L}_i} \mathrm{score}_{\mathcal{L}_i}(d), \tag{1}$$

where $\mathrm{score}_{\mathcal{L}_i}(d)$ is the retrieval score of d produced by list $\mathcal{L}_i$ and $\mathrm{score}_{\mathcal{L}_i}(d) = 0$ if $d \notin \mathcal{L}_i$. In addition to summing scores, CombMNZ rewards documents that are ranked high in many lists:

$$f_{\mathrm{CombMNZ}}(d;q) := |\{\mathcal{L}_i : d \in \mathcal{L}_i\}| \cdot f_{\mathrm{CombSUM}}(d;q),$$

where $|\{\mathcal{L}_i : d \in \mathcal{L}_i\}|$ is the number of input lists in which document d appears.

4 PROPOSED METHODS

The cluster hypothesis states that similar documents should have a similar degree of relevance to a given information need. A common approach to exploring this hypothesis in its full power is to consider a document manifold and regularize scores based on global inter-document similarities within this manifold [7, 36, 39].

Consider the example in Fig. 1. Here, relevant documents are shown in the lower moon and non-relevant documents in the upper moon. The top-ranked relevant documents are indicated by red markers '+', while other documents are marked with blue circles 'o', where the size of a circle is proportional to the rank of the corresponding document (larger size denotes higher rank). As we see in Fig. 1(a), if we rank documents based on local similarities such as Euclidean distances, many non-relevant documents in the upper moon that are close to the red crosses will be ranked higher than relevant documents in the lower moon that are further away. However, as we see in Fig. 1(b), if we rank documents using global similarities within the document manifold, all relevant documents can be ranked higher than all of the non-relevant documents.

4.1 Manifold-based fusion

In this subsection, we propose our manifold-based data fusion algorithm, ManX, which integrates with a standard unsupervised data fusion method X such as CombSUM. Recall that we are given a set of ranked lists $\mathcal{L}_1, \ldots, \mathcal{L}_m$, returned in response to a query q by m retrieval systems. Our aim is to calculate a fusion score $f(d;q)$ for each document $d \in C_{\mathcal{L}}$, where $C_{\mathcal{L}} := \bigcup_{i=1}^{m} \mathcal{L}_i$ is a set of documents appearing in the input result lists to be fused, and then rank these documents by their fusion scores to form a single fused result list $\mathcal{L}_f$.

Our first goal is to consider inter-document similarities of all documents in $C_{\mathcal{L}}$ for regularizing fusion scores $\mathbf{f}_{\mathrm{X}} = [f_{\mathrm{X}}(d_1;q), \ldots, f_{\mathrm{X}}(d_n;q)]$, produced by an unsupervised data fusion method X, like CombSUM (see (1)). Here, we let n denote the number of documents appearing in the input lists being fused, i.e., $n = |C_{\mathcal{L}}|$. We define a adjacency matrix:

$$\mathbf{W} \in \mathbb{R}^{n \times n} \tag{2}$$

of inter-document similarities, where $W_{ij} = \mathrm{sim}(d_i, d_j)$ for all pairs of documents in $C_{\mathcal{L}}$ for $i \neq j$ and $W_{ii} = 0$ (required by all manifold models [7, 24, 39]). We compute $\mathrm{sim}(d_i, d_j)$ as:

$$W_{ij} = \mathrm{sim}(d_i, d_j) = \exp\left\{-\frac{1}{2}\left(\mathrm{KL}(\mathbf{d}_i \| \mathbf{d}_j) + \mathrm{KL}(\mathbf{d}_j \| \mathbf{d}_i)\right)\right\}, \tag{3}$$

where $\mathbf{d}_i = [\theta_{d_{i1}}, \theta_{d_{i2}}, \ldots, \theta_{d_{is}}]$ is a vector representation for document d_i with $\theta_{d_{ij}}$ being the j-th word v_j's probability in d computed by an unsupervised language model [4], s is the size of the vocabulary, and $\mathrm{KL}(\cdot\|\cdot)$ is the Kullback-Leibler divergence. We obtain $\theta_{d_{ij}}$, the element in the vector $\mathbf{d}_i$ of document d, by an unsupervised language model with Dirichlet smoothing as:

$$\theta_{d_{ij}} = \frac{c(v_j;d) + \delta \cdot p(v_j \mid C)}{\sum_v c(v;d) + \delta}, \tag{4}$$

where $c(v;d)$ is the total number of times the word v appearing in document d, $p(v \mid C)$ is the probability of the word v appearing in the whole corpus, and δ is the smoothing parameter that is set to the average length of the documents in the corpus [37]. Then, according to [25, 38], we can compute regularized scores $\mathbf{f}_{\mathrm{ManX}}$ of our manifold-based fusion method, ManX, by minimizing the following objective function:

$$\begin{aligned}
\mathbf{f}^*_{\mathrm{ManX}} &= \mathop{\arg\min}_{\mathbf{f}_{\mathrm{ManX}}} Q(\mathbf{f}_{\mathrm{ManX}}) \\
&= \mathop{\arg\min}_{\mathbf{f}_{\mathrm{ManX}}} \frac{1}{2} \sum_{i,j=1}^{n} W_{ij} \left\| \frac{\mathbf{f}_{\mathrm{ManX}\,i}}{\sqrt{D_{ii}}} - \frac{\mathbf{f}_{\mathrm{ManX}\,j}}{\sqrt{D_{jj}}} \right\|^2 \\
&\quad + \frac{1}{2}\mu \sum_{i=1}^{n} \|\mathbf{f}_{\mathrm{ManX}} - \mathbf{f}_{\mathrm{X}}\|^2,
\end{aligned} \tag{5}$$

where μ is a regularization parameter, D_{ii} is an element in the diagonal matrix $\mathbf{D} = \mathrm{diag}(D_{11}, \ldots, D_{nn})$ defined as $D_{ii} = \sum_{j=1}^{n} W_{ij}$ (note that W_{ij} is an element in $\mathbf{W}$ computed by (3)), and $\|\cdot\|$ is the 2-norm. The first component in the middle line in (5) smoothes the fusion score vector $\mathbf{f}_{\mathrm{ManX}}$ by assigning similar scores to similar documents. The second component $||\mathbf{f}_{\mathrm{ManX}} - \mathbf{f}_{\mathrm{X}}||^2$ forces the ManX fusion scores to be close to the original scores $\mathbf{f}_{\mathrm{X}}$ obtained by an unsupervised fusion method such as CombSUM. The amount of regularization is controlled by the parameter μ. The final fused list $\mathcal{L}_f$ is constructed by ranking documents $d \in C_{\mathcal{L}}$ according to their regularized fusion scores $f^*_{\text{ManX}}(d;q)$. The solution of the optimization problem (5) can be found either iteratively or in closed form [38]. The iterative solution is the following:

$$\mathbf{f}_{\text{ManX}}(t+1) = \alpha \mathbf{S} \mathbf{f}_{\text{ManX}}(t) + (1-\alpha)\mathbf{f}_{\text{X}}, \tag{6}$$

where $\mathbf{f}_{\text{ManX}}(t)$ is the vector of regularized fusion scores at iteration t, $\alpha = 1/(1+\mu)$ and $\mathbf{S} = \mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2}$. This can be rewritten as:

$$\mathbf{f}_{\text{ManX}}(t+1) = (\alpha \mathbf{S})^t \mathbf{f}_{\text{X}} + (1-\alpha)\textstyle\sum_{i=0}^{t} (\alpha \mathbf{S})^i \mathbf{f}_{\text{X}}. \tag{7}$$

The time complexity of this iterative process is equal to the complexity of matrix multiplication, i.e., $O(n^3)$. The closed form solution of (5) can be written as follows:

$$\mathbf{f}^*_{\text{ManX}} = (1-\alpha)(\mathbf{I}_n - \alpha \mathbf{S})^{-1}\mathbf{f}_{\text{X}}, \tag{8}$$

where $\mathbf{I}_n$ is an $n \times n$ identity matrix. This means that in order to calculate regularized fusion scores in closed form, one needs to inverse the matrix $\mathbf{I} - \alpha\mathbf{S}$, which also requires $O(n^3)$ time. Thus, both the iterative and closed form approaches to computing regularized scores $\mathbf{f}_{\text{ManX}}$ have cubic complexity. To make our ManX technique applicable in practice, we develop an efficient version below.

4.2 Efficient manifold-based fusion

We propose *a-ManX*, a revised ManX aggregation method that utilizes *anchors* for efficient improvement, to reduce the complexity of ManX. It differs from previous manifold-based algorithms [24, 36] in the way we design the adjacency matrix used in manifolds, the anchor definition and how we reduce the computational time. We first discuss the way anchor-documents can be chosen and used to represent all documents. We then show how the optimization problem (5) and its optimal solution (8) should be adjusted to anchors.

4.2.1 Defining anchors. In a variety of real world information retrieval applications, including web search, users mainly pay attention to the top-k documents and ignore documents that are ranked low [6, 14, 36]. Following this idea, a-ManX assumes the top-k documents ($k \ll n$) produced by a basic unsupervised data fusion method X to be relevant and considers them as anchors. We denote the (unknown) regularized fusion scores of these documents as $\mathbf{a}_{\text{a-ManX}} = [f_{\text{a-ManX}}(a_1;q), \ldots, f_{\text{a-ManX}}(a_k;q)]$.

Then, we represent regularized fusion score $f_{\text{a-ManX}}(d_i;q)$ for document d_i as a linear combination of scores in $\mathbf{a}_{\text{a-ManX}}$:

$$f_{\text{a-ManX}}(d_i;q) = \textstyle\sum_{j=1}^{k} Z_{ij} f_{\text{a-ManX}}(a_j;q), \tag{9}$$

where k is the total number of top-k documents acting as anchors, and Z_{ij} are the weights discussed below. In matrix form for all documents this can be written as follows:

$$\mathbf{f}_{\text{a-ManX}} = \mathbf{Z}\mathbf{a}_{\text{a-ManX}}, \tag{10}$$

where a good design principle for the weight matrix $\mathbf{Z}$ is to have $\sum_{j=1}^{k} Z_{ij} = 1$ and $Z_{ij} \geq 0$. Therefore, we define Z_{ij} as:

$$Z_{ij} = \frac{\text{sim}(d_i, a_j)}{\sum_{l=1}^{k} \text{sim}(d_i, a_l)}. \tag{11}$$

Hence, the more similar document d_i and anchor a_j are, the higher the weight Z_{ij}. Thus, documents similar to anchors will have higher regularized scores, which is a desired property as we assume anchor documents to be relevant.

We need to redefine the similarity matrix $\mathbf{W}$ and propose a new design based on the anchors for graph construction:

$$\mathbf{W} = \mathbf{Z}\mathbf{Z}^\top. \tag{12}$$

According to this definition, two documents d_i and d_j have positive similarity $W_{ij} > 0$ if they share at least one anchor-document d_l: $Z_{il} \neq 0$ & $Z_{jl} \neq 0$. The more anchors are shared, the more similar the documents are. Compared to the original adjacency matrix defined in (2), where an $n \times n$ matrix $\mathbf{W}$ needs to be kept in memory, the adjacency matrix $\mathbf{W}$ in (12) is scalable for ranking large datasets, as it only needs to save the $n \times k$ matrix $\mathbf{Z}$.

4.2.2 Efficient optimal solution. Instead of solving the optimization problem (5) for all regularized scores $\mathbf{f}_{\text{a-ManX}}$, we need to solve it only for $\mathbf{a}_{\text{a-ManX}}$:

$$\begin{aligned}
\mathbf{a}^*_{\text{a-ManX}} &= \operatorname*{arg\,min}_{\mathbf{a}_{\text{a-ManX}}} Q(\mathbf{a}_{\text{a-ManX}}) \\
&= \operatorname*{arg\,min}_{\mathbf{a}_{\text{a-ManX}}} \frac{1}{2} \sum_{i,j=1}^{n} W_{ij} \left\| \frac{\mathbf{Z}\mathbf{a}_{\text{a-ManX}\,i}}{\sqrt{D_{ii}}} - \frac{\mathbf{Z}\mathbf{a}_{\text{a-ManX}\,j}}{\sqrt{D_{jj}}} \right\|^2 \\
&\quad + \frac{1}{2}\mu \sum_{i=1}^{n} \|\mathbf{Z}\mathbf{a}_{\text{a-ManX}} - \mathbf{f}_{\text{X}}\|^2,
\end{aligned} \tag{13}$$

where $D_{ii} = \sum_{j=1}^{n} W_{ij}$ and $\mathbf{Z}\mathbf{a}_{\text{a-ManX}\,i}$ is the i-th element in $\mathbf{Z}\mathbf{a}_{\text{a-ManX}} = \mathbf{f}_{\text{a-ManX}}$, i.e., (10). Then, optimal regularized scores can be obtained as follows:

$$\mathbf{f}^*_{\text{a-ManX}} = \mathbf{Z}\mathbf{a}^*_{\text{a-ManX}}. \tag{14}$$

Then, the optimal regularized fusion scores for anchor-documents can be calculated in closed form as follows:

$$\mathbf{f}^*_{\text{a-ManX}} = \mathbf{Z}\mathbf{a}^*_{\text{a-ManX}} = (\mathbf{I}_n - \alpha\mathbf{S})^{-1}\mathbf{f}_{\text{X}}, \tag{15}$$

where $\mathbf{S} = \mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2} = \mathbf{D}^{-1/2}\mathbf{Z}\mathbf{Z}^\top\mathbf{D}^{-1/2}$. Note that here we drop the constant $(1-\alpha)$, because it does not affect the final ranking of documents (see (8)).

The matrix $\mathbf{I}_n - \alpha\mathbf{S}$ still has dimensions $n \times n$. However, if we set $\mathbf{P} = \mathbf{Z}^\top\mathbf{D}^{-\frac{1}{2}}$, then we can rewrite (15) as follows:

$$\mathbf{f}^*_{\text{a-ManX}} = (\mathbf{I}_n - \mathbf{P}^\top(\mathbf{P}\mathbf{P}^\top - \frac{1}{\alpha}\mathbf{I}_k)^{-1}\mathbf{P})\mathbf{f}_{\text{X}}, \tag{16}$$

where matrix $\mathbf{P}\mathbf{P}^\top - \frac{1}{\alpha}\mathbf{I}_k$ has dimensions $k \times k$ and, thus, requires only $O(k^3)$ rather than $O(n^3)$ to calculate the inverse. (15) and (16) are equivalent, as when we multiply matrix $\mathbf{I}_n - \alpha\mathbf{S}$ (not inverse) from (15) by the matrix from (16), we get the identity matrix $\mathbf{I}_n$. The proof that (15) and (16) are equivalent is included in Appendix A. In fact, in (16) for efficient computations, we do not need to save the newly designed adjacency matrix $\mathbf{W}$ in memory but only the matrix $\mathbf{Z}$. For the diagonal matrix $\mathbf{D}$ with $D_{ii} = \sum_{j=1}^{n} W_{ij}$ used in (16), we obtain $\mathbf{D}$ without using $\mathbf{W}$ as well, as $\mathbf{W} = \mathbf{Z}^\top\mathbf{Z}$ and $D_{ii} = \sum_{j=1}^{n} \mathbf{z}_i^\top \mathbf{z}_j$, where $\mathbf{z}_i$ is the i-th column vector of matrix $\mathbf{Z}$.

4.3 Virtual manifold learning

The manifold-based fusion model proposed in §4.1 lacks the ability to rank adversarial examples correctly. In this subsection, we propose a *virtual* manifold learning algorithm, v-ManX, that trains manifolds to correctly rank both unmodified (original) and virtual adversarial documents. It improves not only robustness to adversarial documents, but also generalization performance of original documents.

For each original document $d_i \in C_{\mathcal{L}}$, we find its virtual adversarial document d_i^{v}. Let's denote the vector representation of d_i^{v} as $\mathbf{d}_i^{\mathrm{v}}$ and let $\mathbf{d}_i^{\mathrm{v}} = \mathbf{d}_i + \mathbf{r}_i$, which is built based on original document d's vector representation $\mathbf{d}_i$ and its virtual adversarial perturbation $\mathbf{r}_i$. Here $\mathbf{r}_i$ is the virtual adversarial perturbation making to the original document d_i. We propose to obtain $\mathbf{r}_i$ as:

$$\begin{aligned}\mathbf{r}_i^* &= \mathop{\arg\min}_{\mathbf{r}_i, \|\mathbf{r}_i\|^2 \leq \epsilon} \sum_{j=1}^{n} \mathrm{sim}(\mathbf{d}_i + \mathbf{r}_i, \mathbf{d}_j) \\ &= \mathop{\arg\min}_{\mathbf{r}_i, \|\mathbf{r}_i\|^2 \leq \epsilon} \sum_{j=1}^{n} \exp\left\{-\frac{1}{2}\left(\mathrm{KL}(\mathbf{d}_i + \mathbf{r}_i \| \mathbf{d}_j) + \mathrm{KL}(\mathbf{d}_j \| \mathbf{d}_i + \mathbf{r}_i)\right)\right\},\end{aligned} \tag{17}$$

where $\epsilon > 0$ is a constant parameter that controls the amount of the perturbation. The motivation behind (17) is that the virtual adversarial document needs to be as far as possible from any of the original documents while making the perturbation such that it can significantly increase the loss incurred by our manifold learning model (see below), and thus improves the robustness of the model. Given a set of original documents $C_{\mathcal{L}} = \{d_j\}_{j=1}^{n}$ being fused, we can obtain the optimal $\mathbf{r}_i^*$ for each document d_i so as to obtain $\mathbf{d}_i^{\mathrm{v}}$ by *globally* considering all other documents as:

$$\mathbf{r}_i^* = \epsilon \times \overline{n\mathbf{d}_i - \sum_{j=1}^{n} \mathbf{d}_j}, \tag{18}$$

where $\overline{\mathbf{x}}$ denotes an operator acting on an arbitrary non-zero vector $\mathbf{x}$ that returns a unit vector in the direction of $\mathbf{x}$. The time complexity of computing $\mathbf{r}_i^*$ is $O(n)$ for each document, which is faster than any of the state-of-the-art algorithms [28, 29]: their time complexity is at least $O(n^2)$ and they approximately generate *local* virtual adversarial data, i.e., obtaining the perturbation by considering the example itself but not the others. See Appendix B for our derivation of getting $\mathbf{r}_i^*$.

After obtaining a virtual adversarial perturbation for each original document, in total we double the number of documents for manifold learning. Thus, we have more documents and, hence, more information to regularize fusion scores in our virtual manifold model. The document set in manifold learning becomes $\{d_1, \ldots, d_n, d_1^{\mathrm{v}}, \ldots, d_n^{\mathrm{v}}\}$. Let $\mathbf{f}_{\text{v-ManX}}$ be the regularized score for both original documents and virtual adversarial documents with $\mathbf{f}_{\text{v-ManX}\,i}$ being the regularized score shared by both d_i and its virtual adversarial document d_i^{v}. Note that the size of $\mathbf{f}_{\text{v-ManX}}$ is still n, as each "virtual and original document pair" needs to share the same regularized score. Compared to (2), the adjacency matrix (in $\mathbb{R}^{2n\times 2n}$) becomes:

$$\mathbf{W} = \begin{array}{c} \\ d_1 \\ \vdots \\ d_n \\ d_1^{\mathrm{v}} \\ \vdots \\ d_n^{\mathrm{v}} \end{array} \begin{array}{c} \begin{array}{cccccc} d_1 & \cdots & d_n & d_1^{\mathrm{v}} & \cdots & d_n^{\mathrm{v}} \end{array} \\ \begin{bmatrix} W_{11} & \cdots & W_{1n} & W_{1\,n+1} & \cdots & W_{1\,2n} \\ \vdots & \ddots & \vdots & \vdots & \cdots & \vdots \\ W_{n1} & \cdots & W_{nn} & W_{n\,n+1} & \cdots & W_{n\,2n} \\ W_{n+1\,1} & \cdots & W_{n+1\,n} & W_{n+1\,n+1} & \cdots & W_{n+1\,2n} \\ \vdots & \ddots & \vdots & \vdots & \cdots & \vdots \\ W_{2n\,1} & \cdots & W_{2n\,n} & W_{2n\,n+1} & \cdots & W_{2n\,2n} \end{bmatrix} \end{array}. \tag{19}$$

To simplify the discussion, we write $\mathbf{W} = \begin{bmatrix} \mathbf{W}_{11} & \mathbf{W}_{12} \\ \mathbf{W}_{21} & \mathbf{W}_{22} \end{bmatrix}$, where $\mathbf{W}_{11}, \mathbf{W}_{12}, \mathbf{W}_{21}$ and $\mathbf{W}_{22}$ are the corresponding $n \times n$ sub-matrixes in $\mathbf{W}$. We write $\mathbf{D}$ for the diagonal matrix for $\mathbf{W}$: $\mathbf{D} = \begin{bmatrix} \mathbf{D}_1 \\ \mathbf{D}_2 \end{bmatrix}$, where $\mathbf{D}_1 = \mathrm{diag}(D_{11}, D_{22}, \ldots, D_{nn})$ and $\mathbf{D}_2 = \mathrm{diag}(D_{n+1\,n+1}, D_{n+2\,n+2}, \ldots, D_{2n\,2n})$. In our proposed virtual adversarial manifold learning based aggregation method, v-ManX, we obtain the optimal $\mathbf{f}^*_{\text{v-ManX}}$ for aggregation by minimizing the following objective function:

$$\mathbf{f}^*_{\text{v-ManX}} = \mathop{\arg\min}_{\mathbf{f}_{\text{v-ManX}}} Q(\mathbf{f}_{\text{v-ManX}}) = \tag{20}$$

$$\frac{1}{4}\sum_{i,j=1}^{2n} W_{ij} \left\| \frac{\mathbf{c}_{\text{v-ManX}\,i}}{\sqrt{D_{ii}}} - \frac{\mathbf{c}_{\text{v-ManX}\,j}}{\sqrt{D_{jj}}} \right\|^2 + \frac{\mu}{4}\sum_{i=1}^{2n} \|\mathbf{c}_{\text{v-ManX}\,i} - \mathbf{h}_{\mathrm{X}\,i}\|^2,$$

where $\mathbf{c}_{\text{v-ManX}} = \mathbf{f}_{\text{v-ManX}} \oplus \mathbf{f}_{\text{v-ManX}}$ and $\mathbf{h}_{\mathrm{X}} = \mathbf{f}_{\mathrm{X}} \oplus \mathbf{f}_{\mathrm{X}}$ ($\mathbf{h}_{\mathrm{X}}$ can be obtained in an unsupervised way, as $\mathbf{f}_{\mathrm{X}}$ can be obtained by an unsupervised fusion method like CombSUM), and $\oplus$ is the concatenation operator for two vectors; $\mathbf{c}_{\text{v-ManX}\,i}$ and $\mathbf{h}_{\mathrm{X}\,i}$ are the i-th element in $\mathbf{c}_{\text{v-ManX}}$ and $\mathbf{h}_{\mathrm{X}}$, respectively. Unlike (5), which smoothes original documents only, (20) smoothes both original and virtual adversarial ones – between original documents themselves ($\mathbf{W}_{11}$), between virtual adversarial documents themselves ($\mathbf{W}_{22}$), between original and virtual adversarial documents ($\mathbf{W}_{12}$ and $\mathbf{W}_{21}$), by the first term on the right-hand side of (20), while still forcing the v-Manx fusion scores to be close to the original scores $\mathbf{f}_{\mathrm{X}}$ by the last term in (20). The closed form solution of (20) is $\mathbf{f}^*_{\text{v-ManX}} =$

$$(1-\alpha)\left(\mathbf{I} - \alpha\left(\frac{1}{2}\mathrm{S}_{11} + \frac{1}{2}\mathrm{S}_{12} + \frac{1}{2}\mathrm{S}_{21} + \frac{1}{2}\mathrm{S}_{22}\right)\right)^{-1} \mathbf{f}_{\mathrm{X}}, \tag{21}$$

where $\mathrm{S}_{11} = \mathbf{D}_1^{-1/2}\mathbf{W}_{11}\mathbf{D}_1^{-1/2}$, $\mathrm{S}_{12} = \mathbf{D}_1^{-1/2}\mathbf{W}_{12}\mathbf{D}_1^{-1/2}$, $\mathrm{S}_{21} = \mathbf{D}_2^{-1/2}\mathbf{W}_{21}\mathbf{D}_2^{-1/2}$, $\mathrm{S}_{22} = \mathbf{D}_2^{-1/2}\mathbf{W}_{22}\mathbf{D}_2^{-1/2}$, and again $\alpha = \frac{\mu}{1+\mu}$. The derivation of the closed form solution in (21) is included in Appendix C.

The iterative solution of (20) is $\mathbf{f}_{\text{v-ManX}}(t+1) =$

$$\alpha\left(\frac{1}{2}\mathrm{S}_{11} + \frac{1}{2}\mathrm{S}_{12} + \frac{1}{2}\mathrm{S}_{21} + \frac{1}{2}\mathrm{S}_{22}\right)\mathbf{f}_{\text{v-ManX}}(t) + (1-\alpha)\mathbf{f}_{\mathrm{X}}, \tag{22}$$

where $\mathbf{f}_{\text{v-ManX}}(t)$ is the vector of regularized fusion scores at iteration t. After $t \longrightarrow +\infty$ iterations, the optimal $\mathbf{f}^*_{\text{v-ManX}}$ can be set to be $\mathbf{f}_{\text{v-ManX}}(+\infty)$ and is the closed form solution, i.e., (21). See Appendix D for the proof that (22) is equivalent to (21) when $t \longrightarrow +\infty$. The time complexity of both (21) and (22) is $O(n^3)$.

4.4 Efficient virtual manifold learning

Similar to §4.2, we boost the efficiency of our virtual manifold learning algorithm proposed in §4.3 to arrive at a method called a-v-ManX, a revised v-ManX utilizing anchors for efficiency. As before, we denote the (unknown) regularized fusion scores of the top-k documents as $\mathbf{a}_{\text{a-v-ManX}} = [f_{\text{a-v-ManX}}(a_1; q), \ldots, f_{\text{a-v-ManX}}(a_k; q)]$. Then, we represent the regularized fusion score $f_{\text{a-v-ManX}}(d_i; q)$ for both document d_i and its virtual document d_i^{v} as a linear combination of scores in $\mathbf{a}_{\text{a-v-ManX}}$:

$$f_{\text{a-v-ManX}}(d_i; q) = \textstyle\sum_{j=1}^{k} Z_{ij} f_{\text{a-v-ManX}}(a_j; q), \tag{23}$$

where Z_{ij} is the weight computed by (11).

Instead of solving the optimization problem (20) for all regularized scores $\mathbf{f}_{\text{a-v-ManX}}$, we need to solve it only for $\mathbf{a}_{\text{a-v-ManX}}$:

$$
\begin{aligned}
\mathbf{a}^{*}_{\text{a-v-ManX}} &= \underset{\mathbf{a}_{\text{a-v-ManX}}}{\arg\min}\, Q(\mathbf{a}_{\text{a-v-ManX}}) \\
&= \underset{\mathbf{a}_{\text{a-v-ManX}}}{\arg\min}\, \frac{1}{4}\sum_{i,j=1}^{n} W_{ij} \left\| \frac{\mathbf{Z}_{11}\mathbf{a}_{\text{a-ManX}\,i}}{\sqrt{D_{ii}}} - \frac{\mathbf{Z}_{11}\mathbf{a}_{\text{a-ManX}\,j}}{\sqrt{D_{jj}}} \right\|^2 + \\
&\quad \frac{1}{4}\sum_{i=1}^{n}\sum_{j=n+1}^{2n} W_{ij} \left\| \frac{\mathbf{Z}_{12}\mathbf{a}_{\text{a-ManX}\,i}}{\sqrt{D_{ii}}} - \frac{\mathbf{Z}_{12}\mathbf{a}_{\text{a-ManX}\,j}}{\sqrt{D_{jj}}} \right\|^2 + \\
&\quad \frac{1}{4}\sum_{i=n+1}^{2n}\sum_{j=1}^{n} W_{ij} \left\| \frac{\mathbf{Z}_{21}\mathbf{a}_{\text{a-ManX}\,i}}{\sqrt{D_{ii}}} - \frac{\mathbf{Z}_{21}\mathbf{a}_{\text{a-ManX}\,j}}{\sqrt{D_{jj}}} \right\|^2 + \\
&\quad \frac{1}{4}\sum_{i=n+1}^{2n}\sum_{j=n+1}^{2n} W_{ij} \left\| \frac{\mathbf{Z}_{22}\mathbf{a}_{\text{a-ManX}\,i}}{\sqrt{D_{ii}}} - \frac{\mathbf{Z}_{22}\mathbf{a}_{\text{a-ManX}\,j}}{\sqrt{D_{jj}}} \right\|^2 + \\
&\quad \frac{\mu}{4}\sum_{i=1}^{2n} \|\mathbf{c}_{\text{a-v-ManX}\,i} - \mathbf{h}_{\text{X}\,i}\|^2 . \qquad (24)
\end{aligned}
$$

Here, $\mathbf{c}_{\text{a-v-ManX}} = \mathbf{f}_{\text{a-v-ManX}} \oplus \mathbf{f}_{\text{a-v-ManX}}$, $\mathbf{Z} = \begin{bmatrix} \mathbf{Z}_{11} & \mathbf{Z}_{12} \\ \mathbf{Z}_{21} & \mathbf{Z}_{22} \end{bmatrix}$ with each sub-matrix of size $n \times n$ and the element Z_{ij} computed by (11), and $\mathbf{W} = \begin{bmatrix} \mathbf{W}_{11} & \mathbf{W}_{12} \\ \mathbf{W}_{21} & \mathbf{W}_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{Z}_{11}\mathbf{Z}_{11}^{\top} & \mathbf{Z}_{12}\mathbf{Z}_{12}^{\top} \\ \mathbf{Z}_{21}\mathbf{Z}_{21}^{\top} & \mathbf{Z}_{22}\mathbf{Z}_{22}^{\top} \end{bmatrix}$.

Similar to (21), the closed form solution of (24) is $\mathbf{f}^{*}_{\text{a-v-ManX}} =$

$$
(1-\alpha)\left(\mathbf{I} - \alpha\left(\frac{1}{2}\mathbf{S}_{11} + \frac{1}{2}\mathbf{S}_{12} + \frac{1}{2}\mathbf{S}_{21} + \frac{1}{2}\mathbf{S}_{22}\right)\right)^{-1} \mathbf{f}_{\text{X}}, \qquad (25)
$$

where $\mathbf{S}_{11} = \mathbf{D}_1^{-1/2}\mathbf{Z}_{11}\mathbf{Z}_{11}^{\top}\mathbf{D}_1^{-1/2}$, $\mathbf{S}_{12} = \mathbf{D}_1^{-1/2}\mathbf{Z}_{12}\mathbf{Z}_{12}^{\top}\mathbf{D}_1^{-1/2}$, $\mathbf{S}_{21} = \mathbf{D}_2^{-1/2}\mathbf{Z}_{21}\mathbf{Z}_{21}^{\top}\mathbf{D}_2^{-1/2}$, $\mathbf{S}_{22} = \mathbf{D}_2^{-1/2}\mathbf{Z}_{22}\mathbf{Z}_{22}^{\top}\mathbf{D}_2^{-1/2}$. Again, if we set $\mathbf{P}_{11} = \mathbf{Z}_{11}^{\top}\mathbf{D}_1^{-\frac{1}{2}}$, $\mathbf{P}_{12} = \mathbf{Z}_{12}^{\top}\mathbf{D}_1^{-\frac{1}{2}}$, $\mathbf{P}_{21} = \mathbf{Z}_{21}^{\top}\mathbf{D}_2^{-\frac{1}{2}}$, and $\mathbf{P}_{22} = \mathbf{Z}_{22}^{\top}\mathbf{D}_2^{-\frac{1}{2}}$, the time complexity of (25) will reduce from $O(n^3)$ to $O(k^3)$ ($k \ll n$), which makes a-v-ManX more efficient than v-ManX.

5 EXPERIMENTAL SETUP

5.1 Research questions

The main research questions guiding the paper are:

RQ1 Do manifold-based fusion methods outperform state-of-the-art methods?

RQ2 Does adding virtual adversarial perturbation into manifold learning methods improve performance?

RQ3 Does our way of globally generating virtual adversarial perturbation outperform the local ways in rank aggregation?

RQ4 Do efficient manifold-based fusion methods, a-ManX and a-v-ManX, run faster? Does our way of generating virtual adversarial documents run faster than the state-of-the-art?

RQ5 What is the impact of the number of anchors used in efficient manifold-based methods?

RQ6 What is the effect of the number of lists to be fused on manifold-based methods compared to state-of-the-art methods?

5.2 Datasets

In order to evaluate the proposed manifold-based data fusion methods and answer the research questions stated in § 5.1, we need a set of queries and a number of ranked lists of documents for each query. To this end, We work with three publicly available large document datasets from the Text REtrieval Conference (TREC).[2] We use the titles of TREC topics as queries (150 queries in total) and the runs submitted by participants as ranked lists of documents to be fused. We focus on the ad hoc track at TREC-3 [12], web track at TREC-10 [13] and robust retrieval track at TREC-12 [34]. Table 1 summarizes the key statistics of the datasets. Participants produced 40, 97 and 78 runs for the tracks of TREC-3, TREC-10 and TREC-12, respectively. The precision at rank 20 (p@20) is the official evaluation metric of these three tracks. The precision at rank 20 of the submitted runs varies dramatically; see Table 1.

Table 1: Summary of the datasets used for our experiments.

Track dataset	#queries	#documents	#runs	p@20
TREC-3 ad hoc	50	741,856	40	0.062–0.674
TREC-10 web	50	1,692,096	97	0.001–0.473
TREC-12 robust retrieval	50	528,155	78	0.090–0.393

5.3 Baselines and evaluation measures

We include comparisons among the following algorithms:

Standard unsupervised fusion methods: CombSUM and CombMNZ [30].

Learning-to-rank method: λ-Merge [31] – a supervised, state-of-the-art learning-to-rank-based rank aggregation method.

Clustering-based fusion methods: ClustX [16] and a-ClustX that build on a standard fusion method, i.e., either X = CombSUM or X = CombMNZ. Here, ClustX is the original supervised clustering-based fusion method that creates clusters for each document, and a-ClustX is an efficient version of ClustX applying our efficiency framework from §4.2.

Virtual adversarial learning methods: v-LDSX [28, 29] (Local Distributional Smoothness), a-v-LDSX and our v-ManX (§4.3) and a-v-ManX (§4.4). The only difference between v-LDSX and our v-ManX (§4.3) is the way they generate the virtual adversarial perturbation. LDSX and a-v-LDSX use an iteration and finite difference method to generate local virtual adversarial perturbations.

Efficient fusion methods: a-ClustX, our a-ManX (§4.2) and a-v-ManX (§4.4). a-ClustX is an efficient version that utilizes our proposed efficient fusion framework in §4.2.

Manifold-based fusion methods: ManX (§4.1), a-ManX (§4.2), v-ManX (§4.3), a-v-ManX (§4.4), v-LDSX, and a-v-LDSX.

For convenience, we write *M*SUM for *M*CombSUM, and *M*MNZ for *M*CombSUM, respectively, where M = {Man, a-Man, v-Man, a-v-Man, Clust, a-Clust}. For instance, a-ManSUM is the abbreviation for a-ManCombSUM when M = a-Man. To measure performance, we use MAP [4], p@k (precision@k) and nDCG@k [15], k = {5, 10, 20}, all of which are the official metrics in these three TREC tracks.

5.4 Parameters and settings

For fusion methods λ-Merge, ManX, a-ManX, v-ManX, a-v-ManX, LDSX, a-LDSX, ClustX, and a-ClustX, we randomly split queries into three parts: 70% queries are used to train a fusion model, 20% queries are used to validate the model during training, and 10%

[2] Publicly available from http://trec.nist.gov.

Table 2: Performance of the methods and the best run on the TREC-3 dataset. The best performance per metric is given in bold. Statistically significant differences between ManX and ClustX, between ManX and a-ManX are marked in the upper and lower right hand corner of ManX's score, respectively. Statistically significant differences between any virtual adversarial learning method and ManX are marked in the upper right hand corner of the virtual adversarial learning method. Statistically significant differences between v-LDSX and a-v-LDSX, between v-ManX and a-v-ManX, are marked in the lower right hand corner of v-LDSX and v-ManX, respectively.

			p@			nDCG@		
		MAP	5	10	20	5	10	20
TREC-3	inq102	.1039	.7440	.7220	.6740	.7423	.7275	.6925
	CombSUM	.1073	.8040	.7620	.6960	.8009	.7736	.7245
	λ-Merge	.1081	.8052	.7701	.7120	.8031	.7765	.7321
	a-ClustSUM	.1093	.8075	.7741	.7257	.8102	.7815	.7421
	ClustSUM	.1199	.8200	.8020	.7430	.8136	.8038	.7638
	a-ManSUM	.1312	.8440	.8120	.7840	.8313	.8136	.7958
	ManSUM	.1317$^{\blacktriangle}$	.8480$^{\blacktriangle}$	.8260$^{\blacktriangle}_{\vartriangle}$	.7960$^{\blacktriangle}_{\vartriangle}$	.8336$^{\blacktriangle}$	.8238$^{\blacktriangle}$	.8054$^{\blacktriangle}$
	a-v-LDSSUM	.1421$^{\vartriangle}$	.8513	.8345	.8139$^{\vartriangle}$	.8423$^{\vartriangle}$	.8328$^{\vartriangle}$	.8247$^{\vartriangle}$
	v-LDSSUM	.1453$^{\blacktriangle}$	.8645$^{\blacktriangle}$	.8458$^{\blacktriangle}$	.8267$^{\blacktriangle}$	.8535$^{\blacktriangle}$	.8439$^{\blacktriangle}$	.8432$^{\blacktriangle}$
	a-v-ManSUM	.1470$^{\blacktriangle}$	.8683$^{\blacktriangle}_{\blacktriangle}$	.8537$^{\blacktriangle}_{\blacktriangle}$	.8364$^{\blacktriangle}_{\blacktriangle}$	.8621$^{\blacktriangle}_{\blacktriangle}$	.8543$^{\blacktriangle}_{\blacktriangle}$	.8532$^{\blacktriangle}_{\blacktriangle}$
	v-ManSUM	.1571$^{\blacktriangle}_{\vartriangle}$	.8857$^{\blacktriangle}_{\blacktriangle}$	.8639$^{\blacktriangle}_{\blacktriangle}$	.8437$^{\blacktriangle}_{\blacktriangle}$	.8756$^{\blacktriangle}_{\blacktriangle}$	.8673$^{\blacktriangle}_{\blacktriangle}$	**.8660**$^{\blacktriangle}_{\blacktriangle}$
	CombMNZ	.1065	.8080	.7700	.6970	.8021	.7781	.7254
	a-ClustMNZ	.1107	.8113	.7834	.7251	.8056	.7821	.7534
	ClustMNZ	.1236	.8240	.8040	.7630	.8127	.8031	.7766
	a-ManMNZ	.1305	.8310	.8200	.7910	.8120	.8173	.8001
	ManMNZ	.1324	.8334	.8240	.7930$^{\blacktriangle}$	.8148	.8183$^{\vartriangle}$	.8004$^{\blacktriangle}$
	a-v-LDSSUM	.1434$^{\vartriangle}$	.8538$^{\vartriangle}$	.8362$^{\vartriangle}$	.8140$^{\vartriangle}$	.8451$^{\vartriangle}$	.8334$^{\vartriangle}$	.8252$^{\vartriangle}$
	v-LDSSUM	.1454$^{\blacktriangle}$	.8656$^{\blacktriangle}$	.8463$^{\blacktriangle}$	.8279$^{\blacktriangle}$	.8552$^{\blacktriangle}$	.8447$^{\blacktriangle}$	.8442$^{\blacktriangle}$
	a-v-ManSUM	.1482$^{\blacktriangle}$	.8673$^{\blacktriangle}_{\vartriangle}$	.8548$^{\blacktriangle}$	.8375$^{\blacktriangle}_{\blacktriangle}$	.8642$^{\blacktriangle}_{\vartriangle}$	.8552$^{\blacktriangle}_{\blacktriangle}$	.8530$^{\blacktriangle}_{\blacktriangle}$
	v-ManSUM	**.1572**$^{\blacktriangle}_{\blacktriangle}$	**.8862**$^{\blacktriangle}_{\blacktriangle}$	**.8642**$^{\blacktriangle}_{\vartriangle}$	**.8448**$^{\blacktriangle}_{\blacktriangle}$	**.8767**$^{\blacktriangle}_{\blacktriangle}$	**.8679**$^{\blacktriangle}_{\blacktriangle}$	.8658$^{\blacktriangle}_{\blacktriangle}$

queries are used for testing the model. We train a fusion model by varying the values of its parameters and then choose the best values during validation. The training, validation, test splits are permuted until all queries were chosen once for the test set. Statistical significance of observed differences between two results is tested using a two-tailed paired t-test and is denoted using $^{\blacktriangle}$ (or $^{\blacktriangledown}$) for significant differences for $\alpha = .01$, or $^{\vartriangle}$ (and $^{\triangledown}$) for $\alpha = .05$.

6 RESULTS AND ANALYSIS

6.1 Effectiveness of proposed methods

RQ1: We compare the ranking performance of all of our manifold learning algorithms to that of the baselines. We use these algorithms to aggregate the top-5 best retrieval runs in each TREC dataset.

Tables 2, 3 and 4 show the results; we also show the performance of the best run in each TREC dataset, i.e., runs inq102, iit01m, and pricRBa2, respectively.[3]

There are several trends worth noting. (1) *Fusion vs. best single run*: All data fusion methods statistically significantly outperform the best single run, which underlines the value of data fusion for improving the performance of document ranking. (2) *Manifold-based fusion methods (ManX, v-LDSX, v-ManX and their efficient versions) vs. all other fusion methods*: Compared to other state-of-the-art fusion methods (CombSUM, CombMNZ, λ-Merge, ClustX), manifold-based methods are among the best performing fusion methods in terms of all metrics, and the performance differences are usually statistically significant. Thus, fusing documents via manifold algorithms can enhance the performance of data fusion. (3) *Manifold-based vs. clustering-based (ClustX and a-ClustX)*: Tables 2, 3 and 4 show that both manifold-based methods outperform cluster-based methods on all datasets and most improvements are statistically significant. Thus exploiting global inter-document similarities in manifolds helps to enhance performance. (4) *Efficient method vs. its original aggregation method, such as a-ManX vs. ManX*: The efficient method does not perform significantly worse than the original method, although it considers only the top-20 documents as anchors. Thus, our anchor-based approach maintains the effectiveness of manifold-based data fusion while considerably reducing its computational cost (see §6.3).

Table 3: Performance on the TREC-10 dataset. Notational conventions for statistical significant test results are as in Table 2.

			p@			nDCG@		
		MAP	5	10	20	5	10	20
TREC-10	iit01m	.2145	.6320	.5880	.4730	.5650	.5707	.5369
	CombSUM	.1988	.6480	.5620	.4650	.5870	.5679	.5409
	λ-Merge	.2052	.6480	.5731	.4721	.5824	.5731	.5435
	a-ClustSUM	.1902	.6483	.5738	.4842	.5810	.5764	.5543
	ClustSUM	.2351	.6500	.5980	.5370	.5821	.5945	.5955
	a-ManSUM	.2732	.6880	.6500	.6004	.6018	.6210	.6293
	ManSUM	.2734$^{\blacktriangle}$	.7040$^{\blacktriangle}$	.6580$^{\blacktriangle}$	.6020$^{\blacktriangle}$	.6293$^{\blacktriangle}_{\vartriangle}$	.6369$^{\blacktriangle}$	.6447$^{\blacktriangle}$
	a-v-LDSSUM	.2814$^{\vartriangle}$	.7124$^{\vartriangle}$	.6625$^{\vartriangle}$	.6247$^{\vartriangle}$	.6275$^{\vartriangle}$	.6347$^{\vartriangle}$	.6545$^{\vartriangle}$
	v-LDSSUM	.2924$^{\blacktriangle}$	.7232$^{\blacktriangle}$	.6741$^{\blacktriangle}$	.6345$^{\blacktriangle}$	.6357$^{\blacktriangle}$	.6472$^{\blacktriangle}$	.6759$^{\blacktriangle}$
	a-v-ManSUM	.2941$^{\blacktriangle}_{\blacktriangle}$	.7263$^{\blacktriangle}_{\blacktriangle}$	.6941$^{\blacktriangle}_{\blacktriangle}$	.6455$^{\blacktriangle}_{\blacktriangle}$	.6478$^{\blacktriangle}_{\blacktriangle}$	.6567$^{\blacktriangle}_{\blacktriangle}$	.6836$^{\blacktriangle}_{\blacktriangle}$
	v-ManSUM	**.3110**$^{\blacktriangle}_{\vartriangle}$	**.7451**$^{\blacktriangle}_{\blacktriangle}$	**.7040**$^{\blacktriangle}_{\blacktriangle}$	**.6537**$^{\blacktriangle}_{\blacktriangle}$	**.6548**$^{\blacktriangle}_{\vartriangle}$	**.6653**$^{\blacktriangle}_{\blacktriangle}$	**.6945**$^{\blacktriangle}_{\blacktriangle}$

Table 4: Performance on the TREC-12 dataset. Notational conventions for statistical significant test results are as in Table 2.

			p@			nDCG@		
		MAP	5	10	20	5	10	20
TREC-12	pircRBa2	.1849	.5280	.4880	.3930	.5004	.4892	.4580
	CombSUM	.1899	.5480	.4870	.3980	.5082	.4866	.4632
	λ-Merge	.1899	.5485	.4870	.4025	.5082	.4873	.4657
	a-ClustSUM	.1903	.5488	.4907	.4113	.5082	.4883	.4724
	ClustSUM	.2213	.5720	.5420	.4655	.5225	.5258	.5197
	a-ManSUM	.2445	.6102	.6000	.5045	.5452	.5603	.5507
	ManSUM	.2498$^{\blacktriangle}$	.6140$^{\blacktriangle}_{\vartriangle}$	.5910$^{\blacktriangle}$	.5085$^{\blacktriangle}$	.5527$^{\blacktriangle}$	.5637$^{\blacktriangle}$	.5611$^{\blacktriangle}$
	a-v-LDSSUM	.2571$^{\vartriangle}$	.6245$^{\vartriangle}$	.6037$^{\vartriangle}$	.5342$^{\vartriangle}$	.5589$^{\vartriangle}$	.5674$^{\vartriangle}$	.5678$^{\vartriangle}$
	v-LDSSUM	.2610$^{\blacktriangle}$	.6342$^{\blacktriangle}$	.6152$^{\blacktriangle}$	.5541$^{\blacktriangle}$	.5672$^{\blacktriangle}$	.5867$^{\blacktriangle}$	.5782$^{\blacktriangle}$
	a-v-ManSUM	.2654$^{\blacktriangle}_{\vartriangle}$	.6413$^{\blacktriangle}_{\blacktriangle}$	.6262$^{\blacktriangle}_{\blacktriangle}$	.5821$^{\blacktriangle}_{\blacktriangle}$	.5747$^{\blacktriangle}_{\blacktriangle}$	.5894$^{\blacktriangle}_{\blacktriangle}$	.5849$^{\blacktriangle}_{\blacktriangle}$
	v-ManSUM	**.2724**$^{\blacktriangle}_{\blacktriangle}$	**.6435**$^{\blacktriangle}_{\vartriangle}$	**.6347**$^{\blacktriangle}_{\blacktriangle}$	**.6054**$^{\blacktriangle}_{\blacktriangle}$	**.5841**$^{\blacktriangle}_{\blacktriangle}$	**.5973**$^{\blacktriangle}_{\vartriangle}$	**.6052**$^{\blacktriangle}_{\blacktriangle}$

[3] In Tables 3 and 4, we only report the results for CombSUM as a basic fusion method as the results for CombMNZ are similar.

Table 5: Running time (in sec.) comparisons among the methods.

	Number of runs				
	3	5	9	15	23
CombSUM	3.98e−4	8.07e−4	1.69e−3	2.86e−3	3.96e−3
a-ManSUM	1.46e−1	5.09e−1	1.79	4.48	11.17
a-v-ManSUM	2.30e−1	1.21	3.32	7.58	18.52
a-v-LDSSUM	2.55e−1	1.47	3.75	8.14	21.39
a-ClustSUM	6.83e−1	1.86	7.30	18.15	43.12
ManSUM	3.08	8.18	33.29	73.47	170.40
v-ManSUM	5.83	16.30	61.72	130.59	317.23
v-LDSSUM	6.21	17.52	65.42	134.23	325.38
ClustSUM	3.59	11.27	42.91	117.45	267.36

6.2 Contribution of virtual adversarial perturbation to manifold learning

RQ2: According to Tables 2, 3 and 4, all virtual adversarial learning methods statistically significantly outperform methods without virtual adversarial documents. E.g., v-LDSSUM outperforms all non-virtual adversarial learning methods, and v-ManSUM outperforms ManX. Thus, adding virtual adversarial perturbation into manifold learning methods improves the performance of aggregation.

RQ3: Tables 2, 3 and 4 show that our virtual adversarial manifold learning methods statistically significantly outperform other virtual adversarial manifold methods that use local information for generating virtual adversarial perturbation. E.g., v-ManSUM outperforms v-LDSSUM, and a-v-ManSUM outperforms a-v-LDSSUM. This highlights another merit of our virtual adversarial perturbation construction method: it globally generates virtual adversarial perturbation for each original document by considering not only the original document itself but all other documents to be fused, and thus makes a positive contribution to the performance of manifold learning. Due to space limitations we only discuss CombSUM as a basic data fusion method below; results for CombMNZ are similar.

6.3 Efficiency of proposed methods

RQ4: To show the efficiency of our proposed efficient manifold learning methods, we randomly choose $m = \{3, 5, 9, 15, 23\}$ runs from the ad hoc track of TREC-3. We combine these runs using the rank aggregation techniques and measure the running time. We repeat the experiment 20 times for each fusion method and report the average running time in seconds.

As can be seen in Table 5, all the efficient aggregation methods significantly run faster than the corresponding non-efficient methods. For instance, a-ManSUM runs faster than ManSUM. This demonstrates the merit of the proposed efficient manifold learning methods: they have lower computational costs and run faster, while still achieving a comparable performance to original non-efficient methods. Also, a-v-ManSUM runs faster than a-v-LDSSUM, and v-ManSUM runs faster than v-LDSSUM. The only differences between a-v-ManSUM and a-v-LDSSUM, and between v-ManSUM and v-LDSSUM, are the ways of obtaining the virtual documents. This demonstrates another merit of our way of generating virtual adversarial documents: our virtual adversarial document generation method runs faster than that of LDS method.

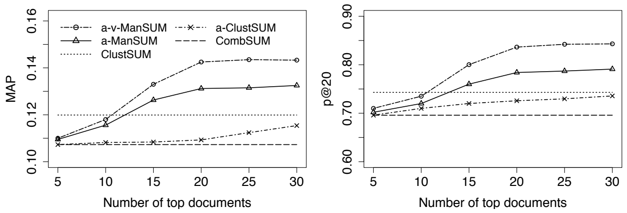


Figure 2: Performance on MAP and p@20 for varying number of anchor documents. The top-5 runs from TREC-3 dataset are fused.

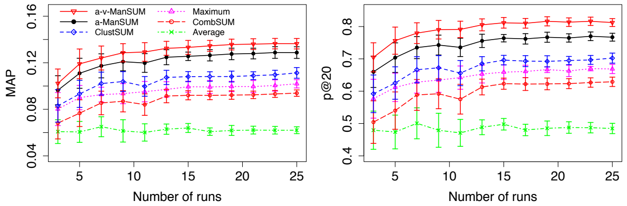


Figure 3: Performance on MAP and p@20 for varying number of anchor documents. The runs are sampled from the TREC-3 dataset.

6.4 Number of anchor documents

RQ5: Next, we examine the effect of the number of anchors on the performance of the proposed manifold learning algorithms that utilize anchors for efficiency. We use the efficient manifold methods, a-v-ManX, a-ManX and a-ClustX, as representatives, as the corresponding non-efficient methods perform better.

Fig. 2 depicts the MAP and p@20 performance for the aggregation methods when fusing the top-5 runs from the TREC-3 ad hoc track. The performance of all efficient methods increases with the number of anchor documents: with more anchor documents come more information to regularize scores of other documents. In contrast, the performance of a-ClustSUM also increases with the number of anchor documents, but cannot top that of ClustSUM.

a-ManSUM usually outperforms ClustSUM when the top-k ($k \geq 15$) documents are used as anchors. This shows that considering only a small number of the top-k documents can still improve data fusion performance in manifold-based approach. Still, a-v-ManSUM always significantly outperforms a-ManSUM. This, again, illustrates that adding virtual adversarial documents into efficient manifold learning improves the performance. Also, the performance of a-v-ManSUM and a-ManSUM seems to level off when more than 20 anchors are used. We conjecture that this is because the more top documents are considered, there is less room to make improvement.

6.5 Number of fused lists

RQ6: Finally, we explore the effect of the number of lists being aggregated on the performance. We use a-v-ManSUM and a-ManSUM as representatives, as their non-efficient versions work better. We randomly choose $m \in \{3, 5, \ldots, 25\}$ runs from the TREC-3 dataset and fuse them. For each m, we repeat the experiment 20 times and report the average performance and the standard deviations.

Fig. 3 shows the performance for varying m, using MAP and p@20. The plots also show the best-performing single run ("Maximum") and the average performance of input runs ("Average"). Data fusion performance usually increases for $m \leq 15$ and then stays almost flat, while the average performance of input runs fluctuates around the same value. This result is in line with other studies on data fusion, showing that the more individual lists are fused the better the performance. Fig. 3 also shows that a-v-ManSUM and a-ManSUM always outperform ClustSUM (here we use ClustSUM only, as ClustSUM outperforms a-ClustSUM). This confirms that utilizing manifolds does boost the performance of data fusion. Here, again, the manifold method, a-v-ManSUM, that utilizes virtual adversarial documents, always outperforms the manifold method, a-ManSUM, that does not integrate virtual adversarial documents. Clearly, most fusion methods beat the average performance and the best input run in most cases.

7 CONCLUSIONS

We have studied the problem of rank aggregation. To enhance the performance of aggregation, we have proposed manifold-based aggregation methods. In our manifold learning methods, we let similar documents across the lists being fused provide support to each other by using inter-document similarities, and let documents in the same intrinsic structure enhance each other's relevance score by considering manifolds of the documents being fused. Since the manifold-based technique that we introduce first, ManX, suffers from high computational costs, we propose an efficient version that can handle large-scale datasets for data fusion, reduce the running time and achieve comparable performance. To improve performance and robustness, we propose a virtual adversarial manifold learning method where we generate virtual adversarial documents by adding perturbation to the original documents. We have conducted experiments on three datasets; the results show the effectiveness and efficiency of the proposed manifold learning methods and our way of generating virtual adversarial documents for manifolds.

There are many unexplored avenues. For instance, can we apply the proposed manifold learning to other information retrieval applications, such as clustering documents in streams [20–22] and diversifying search results [18, 19, 23]? Are there other virtual adversarial perturbation generation methods for manifold learning?

Acknowledgments

This research was supported by Ahold Delhaize, Amsterdam Data Science, the Bloomberg Research Grant program, the Criteo Faculty Research Award program, Elsevier, the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement nr 312827 (VOX-Pol), the Google Faculty Research Awards program, the Microsoft Research Ph.D. program, the Netherlands Institute for Sound and Vision, the Netherlands Organisation for Scientific Research (NWO) under project nrs 612.001.116, CI-14-25, 652.002.001, 612.001.551, 652.001.003, and Yandex. All content represents the opinion of the authors, which is not necessarily shared or endorsed by their respective employers and/or sponsors.

REFERENCES

[1] J. A. Aslam and M. Montague. Models for metasearch. In *SIGIR*, pages 276–284, New Orleans, Louisiana, USA, 2001. ACM.

[2] A. Bhowmik and J. Ghosh. Letor methods for unsupervised rank aggregation. In *WWW*, pages 1331–1340, 2017.

[3] I. Caragiannis, X. Chatzigeorgiou, G. A. Krimpas, and A. A. Voudouris. Optimizing positional scoring rules for rank aggregation. In *AAAI*, pages 430–436, 2017.

[4] W. B. Croft, D. Metzler, and T. Strohman. *Search engines: Information retrieval in practice*. Addison-Wesley Reading, 2015.

[5] I. Csiszar and J. Körner. *Information theory: coding theorems for discrete memoryless systems*. Cambridge University Press, 2011.

[6] J. S. Culpepper, M. Petri, and F. Scholer. Efficient in-memory top-k document retrieval. In *SIGIR*, pages 225–234, Portland, USA, 2012. ACM.

[7] F. Diaz. Regularizing ad hoc retrieval scores. In *CIKM*, pages 672–679, Bremen, Germany, 2005. ACM.

[8] Z. Dong, S. Jia, C. Zhang, M. Pei, and Y. Wu. Deep manifold learning of symmetric positive definite matrices with application to face recognition. In *AAAI*, pages 4009–4015, 2017.

[9] I. J. Goodfellow, J. Shlens, and C. Szegedy. Explaining and harnessing adversarial examples. In *ICLR*, 2015.

[10] A. Griffiths, H. C. Luckhurst, and P. Willett. Adaptive manifold learning. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 34(2):253–265, 2012.

[11] K. Grosse, N. Papernot, et al. Adversarial perturbations against deep neural networks for malware classification. *arXiv preprint arXiv:1606.04435*, 2016.

[12] D. Harman. Overview of the third text retrieval conference (TREC-3). In *TREC*, page 398, Gaithersburg, Maryland, USA, 1994. NIST.

[13] D. Hawking and N. Craswell. Overview of the trec-2001 web track. In *the TREC*, pages 1–8, Gaithersburg, Maryland, USA, 2002. NIST.

[14] K. Hofmann, S. Whiteson, and M. de Rijke. Fidelity, soundness, and efficiency of interleaved comparison methods. *ACM Trans. on Inf. Sys.*, pages 1–43, 2013.

[15] K. Järvelin and J. Kekäläinen. Cumulated gain-based evaluation of IR techniques. *ACM Trans. Inf. Syst.*, 20(4):422–446, 2002.

[16] A. Khudyak-Kozorovitsky and O. Kurland. Cluster-based fusion of retrieved lists. In *SIGIR*, pages 893–902, Beijing, China, 2011. ACM.

[17] S. Liang and M. de Rijke. Burst-aware data fusion for microblog search. *Information Processing & Management*, pages 89–113, 2015.

[18] S. Liang, Z. Ren, and M. de Rijke. Fusion helps diversification. In *SIGIR*, pages 303–312, Gold Coast, Australia, 2014. ACM.

[19] S. Liang, F. Cai, Z. Ren, and M. de Rijke. Efficient structured learning for personalized diversification. *IEEE Transactions on Knowledge and Data Engineering*, 28 (11):2958–2973, 2016.

[20] S. Liang, E. Yilmaz, and E. Kanoulas. Dynamic clustering of streaming short documents. In *KDD*, pages 995–1004, 2016.

[21] S. Liang, Z. Ren, E. Yilmaz, and E. Kanoulas. Collaborative user clustering for short text streams. In *AAAI*, pages 3504–3510, 2017.

[22] S. Liang, Z. Ren, Y. Zhao, J. Ma, E. Yilmaz, and M. de Rijke. Inferring dynamic user interests in streams of short texts for user clustering. *ACM Trans. Inf. Syst.*, 36(1):10:1–10:37, 2017.

[23] S. Liang, E. Yilmaz, H. Shen, M. de Rijke, and W. B. Croft. Search result diversification in short text streams. *ACM Trans. Inf. Syst.*, 36(1):8:1–8:35, 2017.

[24] W. Liu, J. He, and S.-F. Chang. Large graph construction for scalable semi-supervised learning. In *ICML*, pages 679–686, Haifa, Israel, 2010. Omnipress.

[25] W. Liu, J. Wang, and S.-F. Chang. Robust and scalable graph-based semisupervised learning. *Proceedings of the IEEE*, 100(9):2624–2638, 2012.

[26] Y. Liu, Z. Gu, Y.-m. Cheung, and K. A. Hua. Multi-view manifold learning for media interestingness prediction. In *ICMR*, pages 308–314, 2017.

[27] Y.-T. Liu, T.-Y. Liu, et al. Supervised rank aggregation. In *WWW*, pages 481–489, Banff, Alberta, Canada, 2007. ACM.

[28] T. Miyato, S.-i. Maeda, M. Koyama, K. Nakae, and S. Ishii. Distributional smoothing with virtual adversarial training. In *ICLR*, 2016.

[29] T. Miyato, A. M. Dai, and I. Goodfellow. Adversarial training methods for semi-supervised text classification. In *ICLR*, 2017.

[30] J. A. Shaw and E. A. Fox. Combination of multiple searches. In *TREC*, pages 243–252, Gaithersburg, Maryland, USA, 1994. NIST.

[31] D. Sheldon, M. Shokouhi, M. Szummer, and N. Craswell. LambdaMerge: merging the results of query reformulations. In *WSDM*, pages 795–804. ACM, 2011.

[32] M. D. Smucker and J. Allan. A new measure of the cluster hypothesis. In *ICTIR*, pages 281–288. Springer-Verlag, 2009.

[33] C. Szegedy, W. Zaremba, et al. Intriguing properties of neural networks. In *ICLR*, 2014.

[34] E. M. Voorhees. Overview of the TREC 2005 robust retrieval track. In *the TREC*, pages 1–9, Gaithersburg, Maryland, USA, 2005. NIST.

[35] Q. Wang, M. Chen, and X. Li. Quantifying and detecting collective motion by manifold learning. In *AAAI*, pages 4292–4298, 2017.

[36] B. Xu, J. Bu, et al. Efficient manifold ranking for image retrieval. In *SIGIR*, pages 525–534. ACM, 2011.

[37] C. Zhai and J. D. Lafferty. A study of smoothing methods for language models applied to ad hoc information retrieval. In *SIGIR*, 2001.

[38] D. Zhou, O. Bousquet, et al. Learning with local and global consistency. In *NIPS*, pages 321–328. MIT Press, 2004.

[39] D. Zhou, J. Weston, et al. Ranking on data manifolds. In *NIPS*, pages 169–176, 2004.

[40] X. Zhu, Z. Ghahramani, and J. Lafferty. Semi-supervised learning using gaussian fields and harmonic functions. In *ICML*, 2003.

A PROOF OF THE EQUALITY

All notations below are defined in the body of the paper. Analogous proof is found in [36]. We multiply matrix $\mathbf{I}_n - \alpha\mathbf{S}$ (not inverse) from (15) by the matrix from (16) and should get the identity matrix $\mathbf{I}_n$ if they are equivalent, as a result:

$$
\begin{aligned}
&(\mathbf{I}_n - \alpha\mathbf{S}) \times (\mathbf{I}_n - \mathbf{P}^\top(\mathbf{P}\mathbf{P}^\top - \frac{1}{\alpha}\mathbf{I}_k)^{-1}\mathbf{P}) \\
&= (\mathbf{I}_n - \alpha\mathbf{P}^\top\mathbf{P}) \times (\mathbf{I}_n - \mathbf{P}^\top(\mathbf{P}\mathbf{P}^\top - \frac{1}{\alpha}\mathbf{I}_k)^{-1}\mathbf{P}) \\
&= \mathbf{I}_n - \mathbf{P}^\top(\mathbf{P}\mathbf{P}^\top - \frac{1}{\alpha}\mathbf{I}_k)^{-1}\mathbf{P} - \alpha\mathbf{P}^\top\mathbf{P} + \alpha\mathbf{P}^\top\mathbf{P}\mathbf{P}^\top(\mathbf{P}\mathbf{P}^\top - \frac{1}{\alpha})^{-1}\mathbf{P} \\
&= \mathbf{I}_n - (\mathbf{P}^\top - \alpha\mathbf{P}^\top\mathbf{P}\mathbf{P}^\top)(\mathbf{P}\mathbf{P}^\top - \frac{1}{\alpha}\mathbf{I}_k)^{-1}\mathbf{P} - \alpha\mathbf{P}^\top\mathbf{P} \\
&= \mathbf{I}_n - \alpha\mathbf{P}^\top(\frac{1}{\alpha}\mathbf{I}_k - \mathbf{P}\mathbf{P}^\top)(\mathbf{P}\mathbf{P}^\top - \frac{1}{\alpha}\mathbf{I}_k)^{-1}\mathbf{P} - \alpha\mathbf{P}^\top\mathbf{P} \\
&= \mathbf{I}_n + \alpha\mathbf{P}^\top\mathbf{P} - \alpha\mathbf{P}^\top\mathbf{P} = \mathbf{I}_n
\end{aligned}
$$

B DERIVATION OF THE PERTURBATION $\mathbf{r}_i^*$

The derivation of the optimal perturbation $\mathbf{r}_i$ is as follows:

$$
\begin{aligned}
\mathbf{r}_i^* &= \underset{\mathbf{r}_i, \|\mathbf{r}_i\|^2 \le \epsilon}{\arg\min} \sum_{j=1}^{n} \mathrm{sim}(\mathbf{d}_i + \mathbf{r}_i, \mathbf{d}_j) \\
&= \underset{\mathbf{r}_i, \|\mathbf{r}_i\|^2 \le \epsilon}{\arg\min} \sum_{j=1}^{n} \exp\left\{-\frac{1}{2}\left(\mathrm{KL}(\mathbf{d}_i + \mathbf{r}_i \| \mathbf{d}_j) + \mathrm{KL}(\mathbf{d}_j \| \mathbf{d}_i + \mathbf{r}_i)\right)\right\}.
\end{aligned}
$$

As $\sum_{j=1}^{n} \exp\{-x_j\} \ge n \exp\{-\sum_{j=1}^{n} x_j\} \ge \exp\{-\sum_{j=1}^{n} x_j\}$ when $x_j \ge 0$ and $\mathrm{KL}(\cdot\|\cdot) \ge 0$, the above becomes:

$$
\begin{aligned}
&= \underset{\mathbf{r}_i, \|\mathbf{r}_i\|^2 \le \epsilon}{\arg\min} \exp\left\{-\frac{1}{2}\sum_{j=1}^{n}\left(\mathrm{KL}(\mathbf{d}_i + \mathbf{r}_i \| \mathbf{d}_j) + \mathrm{KL}(\mathbf{d}_j \| \mathbf{d}_i + \mathbf{r}_i)\right)\right\} \\
&= \underset{\mathbf{r}_i, \|\mathbf{r}_i\|^2 \le \epsilon}{\arg\max} \exp\left\{\sum_{j=1}^{n}\left(\mathrm{KL}(\mathbf{d}_i + \mathbf{r}_i \| \mathbf{d}_j) + \mathrm{KL}(\mathbf{d}_j \| \mathbf{d}_i + \mathbf{r}_i)\right)\right\}.
\end{aligned}
$$

According to Pinsker's inequality [5] $\mathrm{KL}(\mathbf{x}\|\mathbf{y}) \ge \frac{1}{2\ln 2}\|\mathbf{x} - \mathbf{y}\|_1^2 = \frac{1}{2\ln 2}\|\mathbf{y} - \mathbf{x}\|_1^2$, where $\mathbf{x}$ and $\mathbf{y}$ are two distributions and $\|\cdot\|_1$ is the 1-norm, the above becomes:

$$
\begin{aligned}
&= \underset{\mathbf{r}_i, \|\mathbf{r}_i\|^2 \le \epsilon}{\arg\max} \exp\left\{\sum_{j=1}^{n} \frac{1}{2\ln 2}\|\mathbf{d}_i + \mathbf{r}_i - \mathbf{d}_j\|_1^2 + \frac{1}{2\ln 2}\|\mathbf{d}_j - (\mathbf{r}_i + \mathbf{d}_i)\|_1^2\right\} \\
&= \underset{\mathbf{r}_i, \|\mathbf{r}_i\|^2 \le \epsilon}{\arg\max} \exp\left\{\sum_{j=1}^{n} \frac{2}{2\ln 2}\|\mathbf{d}_i + \mathbf{r}_i - \mathbf{d}_j\|_1^2\right\} \\
&= \underset{\mathbf{r}_i, \|\mathbf{r}_i\|^2 \le \epsilon}{\arg\max} \exp\left\{\sum_{j=1}^{n} \|\mathbf{d}_i + \mathbf{r}_i - \mathbf{d}_j\|_1^2\right\} \\
&= \arg_{\mathbf{r}_i} \exp\left\{\max_{\|\mathbf{r}_i\|^2 \le \epsilon} \sum_{j=1}^{n} \|\mathbf{d}_i + \mathbf{r}_i - \mathbf{d}_j\|_1^2\right\}.
\end{aligned}
\tag{26}
$$

In the above equation (26), $\max_{\|\mathbf{r}_i\|^2 \le \epsilon} \sum_{j=1}^{n} \|\mathbf{d}_i + \mathbf{r}_i - \mathbf{d}_j\|_1^2$ is derived as:

$$
\begin{aligned}
&\max_{\|\mathbf{r}_i\|^2 \le \epsilon} \sum_{j=1}^{n} \|\mathbf{d}_i + \mathbf{r}_i - \mathbf{d}_j\|_1^2 = \max_{\|\mathbf{r}_i\|^2 \le \epsilon} \sum_{j=1}^{n} \|\mathbf{r}_i\|_1^2 + 2\|\mathbf{r}_i(\mathbf{d}_i - \mathbf{d}_j)\|_1 + \|\mathbf{d}_i - \mathbf{d}_j\|_1^2 \\
&= \max_{\|\mathbf{r}_i\|^2 \le \epsilon} \left\{n\|\mathbf{r}_i\|^2 + 2\|\mathbf{r}_i\|_1 \sum_{j=1}^{n} \|\mathbf{d}_i - \mathbf{d}_j\| + \sum_{j=1}^{n} \|\mathbf{d}_i - \mathbf{d}_j\|^2\right\} \\
&\le n\epsilon + \sum_{j=1}^{n} \|\mathbf{d}_i - \mathbf{d}_j\|_1^2 + \max_{\|\mathbf{r}_i\|^2 \le \epsilon} 2\|\mathbf{r}_i\|_1 \sum_{j=1}^{n} \|\mathbf{d}_i - \mathbf{d}_j\|_1 \\
&\le n\epsilon + \sum_{j=1}^{n} \|\mathbf{d}_i - \mathbf{d}_j\|_1^2 + \left(\sum_{j=1}^{n} \|\mathbf{d}_i - \mathbf{d}_j\|_1\right)^2 + \max_{\|\mathbf{r}_i\|^2 \le \epsilon} \|\mathbf{r}_i\|_1^2 \\
&\le n\epsilon + \sum_{j=1}^{n} \|\mathbf{d}_i - \mathbf{d}_j\|_1^2 + \left(\sum_{j=1}^{n} \|\mathbf{d}_i - \mathbf{d}_j\|_1\right)^2 + \epsilon = \text{constant}.
\end{aligned}
\tag{27}
$$

If we let $\mathbf{r}_i = \epsilon \times \overline{\sum_{j=1}^{n}(\mathbf{d}_i - \mathbf{d}_j)} = \epsilon \times \overline{n\mathbf{d}_i - \sum_{j=1}^{n} \mathbf{d}_j}$, $\max_{\|\mathbf{r}_i\|^2 \le \epsilon} \sum_{j=1}^{n} \|\mathbf{d}_i + \mathbf{r}_i - \mathbf{d}_j\|_1^2$, i.e., (27), reaches its maximum. Thus, by combining (26) and (27), we have the final virtual adversarial perturbation $\mathbf{r}_i^* = \epsilon \times \overline{n\mathbf{d}_i - \sum_{j=1}^{n} \mathbf{d}_j}$.

C DERIVATION OF THE CLOSED FORM $\mathbf{f}^*_{\text{v-manx}}$

Writing Δ for $\|\frac{\mathbf{f}_{\text{v-ManX}\,i}}{\sqrt{D_{ii}}} - \frac{\mathbf{f}_{\text{v-ManX}\,j}}{\sqrt{D_{jj}}}\|^2$, the cost function in (20) is:

$$
\begin{aligned}
Q(\mathbf{f}_{\text{v-ManX}}) &= \frac{1}{4}\sum_{i,j=1}^{n} W_{ij}\Delta + \frac{1}{4}\sum_{i=1}^{n}\sum_{j=n+1}^{2n} W_{ij}\Delta \\
&+ \frac{1}{4}\sum_{i,j=n+1}^{2n} W_{ij}\Delta + \frac{1}{4}\sum_{i=n+1}^{2n}\sum_{j=1}^{n} W_{ij}\Delta + \frac{\mu}{2}\sum_{i=1}^{n}\|\mathbf{f}_{\text{v-ManX}\,i} - \mathbf{f}_{\text{X}\,i}\|^2.
\end{aligned}
$$

We differentiate $Q(\mathbf{f}_{\text{v-ManX}})$ with respect to $\mathbf{f}_{\text{v-ManX}}$ and have:

$$
\begin{aligned}
\frac{\partial Q(\mathbf{f}_{\text{v-ManX}})}{\partial \mathbf{f}_{\text{v-ManX}}} &= \mathbf{f}_{\text{v-ManX}} - \frac{1}{2}\mathbf{D}_1^{-1/2}\mathbf{W}_{11}\mathbf{D}_1^{-1/2}\mathbf{f}_{\text{v-ManX}} - \frac{1}{2}\mathbf{D}_1^{-1/2}\mathbf{W}_{12}\mathbf{D}_1^{-1/2}\mathbf{f}_{\text{v-ManX}} \\
&- \frac{1}{2}\mathbf{D}_2^{-1/2}\mathbf{W}_{21}\mathbf{D}_2^{-1/2}\mathbf{f}_{\text{v-ManX}} - \frac{1}{2}\mathbf{D}_2^{-1/2}\mathbf{W}_{22}\mathbf{D}_2^{-1/2}\mathbf{f}_{\text{v-ManX}} + \mu(\mathbf{f}_{\text{v-ManX}} - \mathbf{f}_{\text{X}}).
\end{aligned}
$$

To obtain the closed form solution $\mathbf{f}^*_{\text{v-ManX}}$, we set $\frac{\partial Q(\mathbf{f}_{\text{v-ManX}})}{\partial \mathbf{f}_{\text{v-ManX}}} = 0$ and have:

$$
\begin{aligned}
&(1+\mu)\mathbf{f}^*_{\text{v-ManX}} - \Big(\frac{1}{2}\mathbf{D}_1^{-1/2}\mathbf{W}_{11}\mathbf{D}_1^{-1/2} + \frac{1}{2}\mathbf{D}_1^{-1/2}\mathbf{W}_{12}\mathbf{D}_1^{-1/2} + \\
&\frac{1}{2}\mathbf{D}_2^{-1/2}\mathbf{W}_{21}\mathbf{D}_2^{-1/2} + \frac{1}{2}\mathbf{D}_2^{-1/2}\mathbf{W}_{22}\mathbf{D}_2^{-1/2}\Big)\mathbf{f}^*_{\text{v-ManX}} - \mu\mathbf{f}_{\text{X}} = 0,
\end{aligned}
$$

which results in the following closed form solution:

$$
\mathbf{f}^*_{\text{v-ManX}} = (1-\alpha)\left(\mathbf{I} - \alpha(\frac{1}{2}\mathrm{S}_{11} + \frac{1}{2}\mathrm{S}_{12} + \frac{1}{2}\mathrm{S}_{21} + \frac{1}{2}\mathrm{S}_{22})\right)^{-1}\mathbf{f}_{\text{X}}, \tag{28}
$$

where $\alpha = \frac{\mu}{1+\mu}$, $\mathrm{S}_{11} = \mathbf{D}_1^{-1/2}\mathbf{W}_{11}\mathbf{D}_1^{-1/2}$, $\mathrm{S}_{12} = \mathbf{D}_1^{-1/2}\mathbf{W}_{12}\mathbf{D}_1^{-1/2}$, $\mathrm{S}_{21} = \mathbf{D}_2^{-1/2}\mathbf{W}_{21}\mathbf{D}_2^{-1/2}$, and $\mathrm{S}_{22} = \mathbf{D}_2^{-1/2}\mathbf{W}_{22}\mathbf{D}_2^{-1/2}$.

D DERIVATION OF THE ITERATION FORM

Without loss of generalization, we suppose $\mathbf{f}_{\text{v-ManX}}(0) = \mathbf{f}_{\text{X}}$. According to the iteration $\mathbf{f}_{\text{v-ManX}}(t+1) = \alpha(\frac{1}{2}\mathrm{S}_{11} + \frac{1}{2}\mathrm{S}_{12} + \frac{1}{2}\mathrm{S}_{21} + \frac{1}{2}\mathrm{S}_{22})\mathbf{f}_{\text{v-ManX}}(t) + (1-\alpha)\mathbf{f}_{\text{X}}$, we have:

$$
\begin{aligned}
\mathbf{f}_{\text{v-ManX}}(t) &= \left(\alpha(\frac{1}{2}\mathrm{S}_{11} + \frac{1}{2}\mathrm{S}_{12} + \frac{1}{2}\mathrm{S}_{21} + \frac{1}{2}\mathrm{S}_{22})\right)^{t-1}\mathbf{f}_{\text{X}} + \\
&(1-\alpha)\sum_{i=0}^{t-1}\left(\alpha(\frac{1}{2}\mathrm{S}_{11} + \frac{1}{2}\mathrm{S}_{12} + \frac{1}{2}\mathrm{S}_{21} + \frac{1}{2}\mathrm{S}_{22})\right)^{i}\mathbf{f}_{\text{X}}.
\end{aligned}
\tag{29}
$$

Because $0 < \alpha < 1$ and the eigenvalues of $\alpha(\frac{1}{2}\mathrm{S}_{11} + \frac{1}{2}\mathrm{S}_{12} + \frac{1}{2}\mathrm{S}_{21} + \frac{1}{2}\mathrm{S}_{22})$ is within $[-1, 1]$, the the optimal solution $\mathbf{f}^*_{\text{v-ManX}}$ is:

$$
\begin{aligned}
\mathbf{f}^*_{\text{v-ManX}} &= \lim_{t\to+\infty}\mathbf{f}_{\text{v-ManX}}(t) = \lim_{t\to+\infty}\left(\alpha(\frac{1}{2}\mathrm{S}_{11} + \frac{1}{2}\mathrm{S}_{12} + \frac{1}{2}\mathrm{S}_{21} + \frac{1}{2}\mathrm{S}_{22})\right)^{t-1}\mathbf{f}_{\text{X}} \\
&+ \lim_{t\to+\infty}(1-\alpha)\sum_{i=0}^{t-1}\left(\alpha(\frac{1}{2}\mathrm{S}_{11} + \frac{1}{2}\mathrm{S}_{12} + \frac{1}{2}\mathrm{S}_{21} + \frac{1}{2}\mathrm{S}_{22})\right)^{i}\mathbf{f}_{\text{X}} \\
&= \mathbf{0} + (1-\alpha)\mathbf{f}_{\text{X}}\frac{\mathbf{I} - \left(\alpha(\frac{1}{2}\mathrm{S}_{11} + \frac{1}{2}\mathrm{S}_{12} + \frac{1}{2}\mathrm{S}_{21} + \frac{1}{2}\mathrm{S}_{22})\right)^{+\infty}}{\mathbf{I} - \alpha(\frac{1}{2}\mathrm{S}_{11} + \frac{1}{2}\mathrm{S}_{12} + \frac{1}{2}\mathrm{S}_{21} + \frac{1}{2}\mathrm{S}_{22})} \\
&= (1-\alpha)\left(\mathbf{I} - \alpha(\frac{1}{2}\mathrm{S}_{11} + \frac{1}{2}\mathrm{S}_{12} + \frac{1}{2}\mathrm{S}_{21} + \frac{1}{2}\mathrm{S}_{22})\right)^{-1}\mathbf{f}_{\text{X}},
\end{aligned}
\tag{30}
$$

which is the same as that in (28). Thus, the iteration form holds.