Computational Social Indicators: A Case Study of Chinese University Ranking

Fuli Feng
National University of Singapore
fulifeng93@gmail.com

Liqiang Nie
Shandong University
nieliqiang@gmail.com

Xiang Wang
National University of Singapore
xiangwang@u.nus.edu

Richang Hong
Hefei University of Technology
hongrc.hfut@gmail.com

Tat-Seng Chua
National University of Singapore
dcscts@nus.edu.sg

ABSTRACT

Many professional organizations produce regular reports of social indicators to monitor social progress. Despite their reasonable results and societal value, early efforts on social indicator computing suffer from three problems: 1) labor-intensive data gathering, 2) insufficient data, and 3) expert-relied data fusion. Towards this end, we present a novel graph-based multi-channel ranking scheme for social indicator computation by exploring the rich multi-channel Web data. For each channel, this scheme presents the semi-structured and unstructured data with simple graphs and hypergraphs, respectively. It then groups the channels into different clusters according to their correlations. After that, it uses a unified model to learn the cluster-wise common spaces, perform ranking separately upon each space, and fuse these rankings to produce the final one. We take Chinese university ranking as a case study and validate our scheme over a real-world dataset. It is worth emphasizing that our scheme is applicable to computation of other social indicators, such as Educational attainment.

CCS CONCEPTS

•**Information systems** → *Web mining; Information retrieval; Retrieval models and ranking;*

KEYWORDS

Computational Social Indicators, University Ranking

ACM Reference format:
Fuli Feng, Liqiang Nie, Xiang Wang, Richang Hong, and Tat-Seng Chua. 2017. Computational Social Indicators: A Case Study of Chinese University Ranking. In *Proceedings of SIGIR17, August 7–11, 2017, Shinjuku, Tokyo, Japan, ,* 10 pages.
DOI: http://dx.doi.org/10.1145/3077136.3080773

SIGIR17, August 7–11, 2017, Shinjuku, Tokyo, Japan

DOI: http://dx.doi.org/10.1145/3077136.3080773

1 INTRODUCTION

Social indicators are defined as statistical measures and analytics that describe social trends and conditions that would impact social well-being [14]. A social indicator is usually in the form of a ranking list that orders the entities of interests according to some pre-defined rules. In the past few decades, professional organizations, such as mass media, academic institutes, and government agencies, have calculated and released hundreds of social indicators on different facets of our society, including cost of living [7], health expenditure [17], happiness index [35], and university quality [23]. Generally, social indicators have some key functions, spanning from providing information for decision-makers, monitoring and evaluating policies, to searching for a common good [3]. For instance, university ranking plays a pivotal role in selecting universities for high school students. Meanwhile, university rankings are mirrors for university themselves to improve their education and research quality. Therefore, the accuracy and timely creation of these indicators are extremely useful to a wide variety of users and applications, including the formulation of government policies and planning of social services.

Most of the released social indicators are typically computed in two steps: given a set of entities to be ranked, they first calculate the scores of these entities according to several factors related to the desired social indicator and then fuse the scores using hand-crafted weights to rank the entities. For instance, universities in QS World University Ranking 2016/17 are ranked by scores weighted upon six factors: academic reputation, employer reputation, student-to-faculty ratio, citations per faculty, international student ratio, and international faculty ratio[1]. However, such computation process usually suffers from the following problems: 1) Labor-intensive data collection. Data used to calculate social indicators usually rely heavily on user studies like questionnaire, especially, for subjective factors, such as the academic and employment reputation in the QS Ranking. It thus requires a lot of human resources and the collected data can hardly be applied to compute other social indicators. 2) Data insufficiency. Existing social indicators usually only cover a small fraction of target entities. For example, there indeed exist 2,553 universities in China[2], while most university rankings involve only less than 800 universities. That is because it is non-trivial to carry out a large-scale user study to gather comprehensive information for each target entity. 3) Expert-relied data fusion.

[1] http://tinyurl.com/zj9vgnj/.
[2] http://tinyurl.com/zcbumn3/.

Factor weighting policies rely heavily on experts and different weighting policies may lead to distinct social indicator results. Although we believe that we can find the outstanding experts and generate reasonable social indicator results, it is extremely resource-consuming.

With the fast development of Internet, we are able to collect large-scale and multi-facet data to describe any given entities from the Web, such as interactions and opinions shared in social networking services (SNSs), timely news reports in online mass media, and purchase history in e-commerce platforms. In a sense, the publicly accessible online data enable us to alleviate the aforementioned data collection and scarcity problems, and save human labors. Considering the university ranking as an example, rich data from multiple channels can be gathered to comprehensively describe each university: 1) Official statistics about students and teachers are available in platforms of the Ministry of Education (MOE) and various educational organizations. 2) Important events related to universities are updated on the website of mass media in real time. 3) Academic records are accessible through online bibliographic database like Microsoft Academic[3]. 4) Employment status of graduate students are shared in business and employment-oriented SNSs, such as LinkedIn[4]. 5) University-related comments and opinions from general users are shared in the mainstream social media like Twitter[5].

Much related works have been conducted to rank entities with multi-channel data. For example: 1) Early fusion, which concatenates all the extracted features from different channels into a single feature vector before feeding it into ranking models [9]; 2) Late fusion, which analyzes data from each channel separately and then aggregates their ranking results [13]; 3) Joint learning that simultaneously learns ranking from each channel and encourages the rankings to be consistent with others [16]; 4) Subspace learning, which derives compact latent representations by taking advantage of inherent structures and relations across multiple channels before ranking the entities based upon the latent representations [10, 28, 36]. However, none of these methods is suitable to compute social indicators, since social indicator computation has the following characteristics: 1) Complex channel relations. The correlation may be strong among some channels, while it may be very weak among others. Therefore, it will cause information loss if all channels are equivalently projected into the same space. 2) Data heterogeneity. There are both semi-structured and unstructured data on the Web. For instance, the tables and statistics in webpages are semi-structured; whereas, texts, images, and videos in mass media reports and social media posts are unstructured. 3) Ranking smoothness. Generally speaking, the latest social indicator is consistent with the last update to some extents, because the target entities in the ranking progress relatively slowly during a short duration. 4) Insufficiency and block-wise missing data for entities. It is not unusual that a social indicator involves up to only hundreds of entities, which constraints the usability of complicated methods relying on large-scale training samples, such as deep learning models. Besides, some channel data may be missing for some entities. For example, in university ranking, academic records of some unpopular universities, such as the Taishan University[6], may not be available, as they seldom publish papers in international conferences or journals.

[3]https://academic.microsoft.com/.
[4]https://www.linkedin.com/.
[5]https://twitter.com/.

To address the aforementioned problems, we present a novel graph-based multi-channel ranking scheme (GMR). In particular, we first collect multi-channel Web data corresponding to the given social indicator and extract a set of features from each channel to represent the candidates. For each channel, we construct a simple graph and a hypergraph on its features from semi-structured and unstructured data, respectively. Following that, we calculate the graph Laplacian for each graph, and cluster all the graphs into groups based on the correlations between their Laplacian matrices[7]. Thus, the involved channels in each cluster are strongly correlated. In the light of this, we derive a common space for each cluster and perform a graph-based ranking upon this common space. It is worth mentioning that we differ the entities with all channel data from those with block-wise missing data, when learning the cluster-wise common space. This strategy can avoid baised common representations caused by data incompletion [42]. Simultaneously, we fuse ranking results learned from different clusters to produce the final one. To enforce ranking smoothness, the aggregated result over different clusters is further regularized by the historical rankings. In this work, we apply the proposed generic scheme to address the Chinese university ranking problem, as shown in Figure 1. In particular, this scheme first collects multi-channel Web data, ranging from official data, mass media reports, academic records, employment status of graduate students, to public comments. It then extracts a rich set of features from each channel to comprehensively represent the universities and then feeds the features into the model of GMR to generate the university ranking. Extensive experiments have well-verified our approach. It is worthwhile highlighting that our ranking scheme is extendable to other social indicator computation, such as the cost of living.

The main contributions of this paper are threefold:

- We present a novel graph-based multi-channel ranking scheme towards social indicator computation. It inherits the advantages of late fusion and subspace learning by performing ranking in the cluster-wise common space.
- We successfully take the Chinese university ranking as a case study of social indicator computation.
- We released the involved codes and our constructed data to facilitate the research community[8].

The remainder of this paper is structured as follows. Section 2 reviews the related work. In Section 3, we introduce our proposed scheme. In Section 4, we apply our scheme to Chinese university ranking. Experimental settings and results are reported in Section 5, followed by conclusion and future work in Section 6.

2 RELATED WORK

Our work is related to recent studies on multi-view subspace learning, unsupervised ranking, and university ranking.

[6]http://www2.tsu.edu.cn/www/ywbtsu/.
[7]It is worth emphasizing that some channels may be placed into two clusters since they have one simple graph and one hypergraph Laplacian matrices. This is reasonable since semi-structured and unstructured data may convey different topics and hence may have different correlations with others.
[8]https://github.com/hennande/cur/.

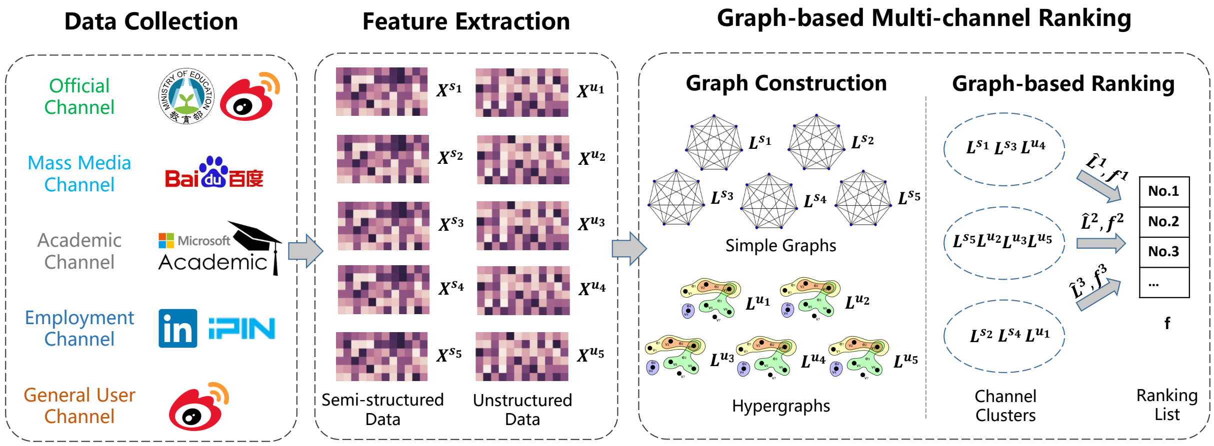

Figure 1: Schematic illustration of social indicator computing and a case study of Chinese university ranking.

2.1 Multi-view Subspace learning

Subspace learning is a widely explored technique to analyze multi-view data. It aims to obtain compact latent representations by leveraging underlying structures and relations across multiple views. Typically, multiple views are mapped into a common space by different algorithms [40], including canonical correlation analysis [15], dictionary learning [4], matrix factorization [20, 29, 41]. In addition, the latent representations are further regularized to be sparse with different norms [34]. Apart from the shallow learning methods, subspace learning is also explored with deep learning models, such as deep restricted Boltzmann machines [31], deep feedforward networks [2, 43], and deep autoencoders [39]. In summary, although great success has been achieved by these models, few of them simultaneously consider the difference between unstructured and semi-structured data, let alone block-wise data missing.

2.2 Unsupervised Ranking

Unsupervised ranking is a popular technique to produce permutation of entities without labled data. Studies on unsupervised ranking are roughly separated into two categories based on whether the entities have direct linkages: 1) Linkage-based ranking. These methods infer rank of entities from the link structure information. For example, PageRank [33] and HITS [24] estimate the importance of webpages from the hyperlinks jumping to the given page. Standing on the shoulder of them, a couple of improvements have been presented. For instance, PopRank [32] further handles Web spam and heterogeneous graphs. BrowseRank [27] integrates the metadata of user behaviors. BiRank [19] expanded it to the bipartite graph. And 2) similarity-based ranking. Similarity-based ranking algorithms enforce that similar entities obtain close ranks. For example, Agarwal [1] constructed a graph, where vertices and edges respectively represent entities and similarity between them, and derived rankings from the Laplacian of the graph. Zhou et al. [44] replaced the conventional graph Laplacian with an iterated and unnormalized one to improve the robustness. Cheng et al. [11] further considered the entity redundancy with sink points in the Laplacian. In addition, Bu et al. [9] utilized the hypergraph instead of the simple one to represent the entities. Yet, most of the aforementioned methods are designed to process single view data.

2.3 University Ranking

Traditional university rankings, such as the U.S.News & World Report[9], Times Higher Education[10], and QS[11], usually measure the qualities of universities with a few pre-defined factors, such as research reputation and academic reputation. These factors are then fused with human designated weights to obtain the final ranking scores. In China, several university rankings are calculated in a similar process by distinct organizations like the Chinese Universities Alumni Association (CUAA)[12], Research Center for China Science Evaluation (RCCSE)[13], and Chinese Academy of Management Science (CAMS)[14]. It is clear that the performance of these ranking systems highly depends on these pre-defined factors and heuristic weights.

Instead of heuristic weights, some researchers attempted to fuse factors with statistical methods. Guarino et al. [18] applied the Bayesian latent variable analysis to learn the weights. Dobrota et al. [12] used I-distance values to estimate the weights based on data from previous years. In addition, some attempts have been done to rank universities with new factors. Lages et al. [25] ranked universities by the importance of their corresponding Wikipedia pages. Kapur et al. [23] utilized LinkedIn Economic Graph data to rank universities by employment of graduates. To sum up, these aforementioned ranking methods pay more attention to weight tuning or calculating specific factors. With the multi-channel Web data, our ranking method explores multi-facets of universities and thus ranks the university precisely.

3 METHODOLOGY

We first define some notations. In particular, we use bold capital letters (e.g., $\mathbf{X}$) and bold lowercase letters (e.g., $\mathbf{x}$) to denote matrices

[9]http://www.usnews.com/rankings.
[10]https://www.timeshighereducation.com/.
[11]http://www.qs.com/.
[12]http://www.cuaa.net/cur/.
[13]http://www.nseac.com/html/168/.
[14]http://edu.sina.com.cn/gaokao/wushulian/.

and vectors, respectively. We employ non-bold letters (e.g., x) to represent scalars, and Greek letters (e.g., λ) to represent parameters. In addition, $tr(\mathbf{X})$ denotes the trace of $\mathbf{X}$. If not clarified, all vectors are in column forms.

The social indicator computation is formalized as: given a list of N entities, a historical ranking list of all the entities $\mathbf{y} \in \mathbb{R}^N$, and the latest entity descriptions from M channels, $\{[\mathbf{X}^{\mathbf{s}_1}, \mathbf{X}^{\mathbf{u}_1}], [\mathbf{X}^{\mathbf{s}_2}, \mathbf{X}^{\mathbf{u}_2}], \cdots, [\mathbf{X}^{\mathbf{s}_M}, \mathbf{X}^{\mathbf{u}_M}]\}$, social indicator computation is to learn a new ranking list $\mathbf{f} \in \mathbb{R}^N$ by harvesting the current data and the historical ranking list. Thereinto, $\mathbf{X}^{\mathbf{s}_\mathbf{m}} \in \mathbb{R}^{N \times D^{sm}}$ and $\mathbf{X}^{\mathbf{u}_\mathbf{m}} \in \mathbb{R}^{N \times D^{um}}$ are the features extracted from semi-structured and unstructured data from the m-th channel; and $\mathbf{y}$ refers to the latest released social indicator by professional organizations. For example, if the desired social indicator is Chinese university ranking in 2017, $\mathbf{y}$ will be the ranking results in 2016. To compute social indicators, we present a novel graph-based multi-channel ranking framework: 1) We first construct a simple graph on the semi-structured data and a hypergraph on the unstructured data for each channel. 2) We then cluster all the graphs into groups based on the correlation of their Laplacian matrices. 3) We ultimately learn a cluster-wise ranking list and fuse them together within a tailored objective function.

3.1 Graph Construction

In some channels, there indeed exist both semi-structured and unstructured data to describe the given entities. Semi-structured ones are of higher quality and thus more discriminative. On the contrary, the unstructured data are more noisy. As their distinct structures and features, we refuse to naively merge the semi-structured and unstructured data. Inspired by that simple graph is sensitive to the data noise; whereas the hypergraph is typically more robust but less discriminative than the simple one [15], we leverage the simple graph and hypergraph to represent the entities and their relations. For each channel, we construct a simple graph over the semi-structured data and a hypergraph over the unstructured ones so that we neither sacrifice the discrimination of semi-structured data nor be affected by the noisy unstructured ones.

3.1.1 Simple Graph Construction. In a simple graph, vertices represent entities and edges refer to their pairwise similarities. A simple graph with N vertices is represented by an incidence matrix, $\mathbf{W} \in \mathbb{R}^{N \times N}$ and a diagonal vertex degree matrix, $\mathbf{D} \in \mathbb{R}^{N \times N}$, where W_{ij} is the similarity between the i-th and j-th vertices; $D_{ii} = \sum_{j=1}^{N} W_{ij}$ is the degree of the i-th vertex. Given $\mathbf{X}^{\mathbf{s}_\mathbf{m}}$, $\mathbf{W}$ is estimated as,

$$W_{ij} = \begin{cases} exp(-\|\mathbf{x}_\mathbf{i}^{\mathbf{s}_\mathbf{m}} - \mathbf{x}_\mathbf{j}^{\mathbf{s}_\mathbf{m}}\|^2/2\sigma^2), & \text{if } i \neq j, \\ 0, & \text{otherwise,} \end{cases} \tag{1}$$

where the radius parameter σ is simply set as the median of the Euclidean distances of all pairs. Following [1], we then calculate the normalized graph Laplacian matrix as,

$$\mathbf{L}^{\mathbf{s}_\mathbf{m}} = \mathbf{D}^{-1/2}(\mathbf{D} - \mathbf{W})\mathbf{D}^{-1/2}. \tag{2}$$

3.1.2 Hypergraph Construction. Generalized from a simple graph where an edge links pairwise vertices, an edge in a hypergraph connects a set of vertices to represent the finitary relations. A hypergraph with N vertices and P edges is represented by an incidence matrix $\mathbf{H} \in \mathbb{R}^{N \times P}$, edge degree matrix $\mathbf{E} \in \mathbb{R}^{P \times P}$, edge weight matrix $\mathbf{W} \in \mathbb{R}^{P \times P}$, and vertex degree matrix $\mathbf{V} \in \mathbb{R}^{N \times N}$, where $H_{ij} = 1$ if the i-th vertex is connected by the j-th edge, otherwise $H_{ij} = 0$; $\mathbf{E}$, $\mathbf{W}$, and $\mathbf{V}$ are diagonal matrices with $E_{jj} = \sum_{i=1}^{N} H_{ij}$, W_{jj} as the weight of the j-th edge, and $V_{ii} = \sum_{j=1}^{P} W_{jj} H_{ij}$. Following [9], given $\mathbf{X}^{\mathbf{u}_\mathbf{m}}$, we calculate the hypergraph Laplacian matrix with,

$$\mathbf{L}^{\mathbf{u}_\mathbf{m}} = \mathbf{V}^{-1/2}(\mathbf{V} - \mathbf{H}\mathbf{W}\mathbf{E}^{-1}\mathbf{H}^T)\mathbf{V}^{-1/2}. \tag{3}$$

In particular, we construct the j-th edge by connecting the k-most similar vertices to the j-th vertex $\mathcal{N}_j(\mathbf{x}_\mathbf{j}^{\mathbf{u}_\mathbf{m}})$ and estimate the weight of the j-th edge by,

$$W_{jj} = \sum_{\mathbf{x}_\mathbf{i}^{\mathbf{u}_\mathbf{m}} \in \mathcal{N}_j(\mathbf{x}_\mathbf{j}^{\mathbf{u}_\mathbf{m}})} exp(-\|\mathbf{x}_\mathbf{i}^{\mathbf{u}_\mathbf{m}} - \mathbf{x}_\mathbf{j}^{\mathbf{u}_\mathbf{m}}\|^2/2\sigma^2). \tag{4}$$

3.2 Channel Clustering

After graph construction, the original multi-channel descriptions $\{[\mathbf{X}^{\mathbf{s}_1}, \mathbf{X}^{\mathbf{u}_1}], [\mathbf{X}^{\mathbf{s}_2}, \mathbf{X}^{\mathbf{u}_2}], \cdots, [\mathbf{X}^{\mathbf{s}_M}, \mathbf{X}^{\mathbf{u}_M}]\}$ are mapped to the Laplacian representations $\{[\mathbf{L}^{\mathbf{s}_1}, \mathbf{L}^{\mathbf{u}_1}], [\mathbf{L}^{\mathbf{s}_2}, \mathbf{L}^{\mathbf{u}_2}], \cdots, [\mathbf{L}^{\mathbf{s}_M}, \mathbf{L}^{\mathbf{u}_M}]\}$. It is not wise to directly fuse all graphs by conventional multi-view ranking techniques, because the correlation among some channels may be very strong, and it may be very weak among others. Therefore, some information may be lost if they are indiscriminately projected to a common space. Towards this end, we first divide all the graphs into groups based on the correlations between their Laplacian matrices with spectral clustering [37]. During the clustering, the distance between two Laplacian matrices is estimated by Hilbert-Schmidt Independence Criterion (HSIC),

$$dis(\mathbf{L}^\mathbf{i}, \mathbf{L}^\mathbf{j}) = HSIC(\mathbf{L}^\mathbf{i}, \mathbf{L}^\mathbf{j}, \phi, \varphi) = (N-1)^2/tr(\mathbf{PHQH}), \tag{5}$$

where ϕ and φ are the kernel functions of the i-th and j-th matrices; $\mathbf{P}$ and $\mathbf{Q} \in \mathbb{R}^{N \times N}$ are the Gram matrices with $P_{mn} = \phi(\mathbf{l}_\mathbf{m}^\mathbf{i}, \mathbf{l}_\mathbf{n}^\mathbf{i})$ and $Q_{mn} = \varphi(\mathbf{l}_\mathbf{m}^\mathbf{j}, \mathbf{l}_\mathbf{n}^\mathbf{j})$; $\mathbf{H} = \mathbf{I} - N^{-2}\mathbf{I}^1 \in \mathbb{R}^{N \times N}$ centers the Gram matrix to have zero mean, where $\mathbf{I}$ and $\mathbf{I}^1$ respectively denote identity and all-one matrices.

3.3 Objective Function

Given the historical ranking list $\mathbf{y}$ and the clustered Laplacian matrices in K groups, $\{\{\mathbf{L}^{\mathbf{1}_1}, \cdots, \mathbf{L}^{\mathbf{1}_{S^1}}\}, \cdots, \{\mathbf{L}^{\mathbf{K}_1}, \cdots, \mathbf{L}^{\mathbf{K}_{S^K}}\}\}$, where S^k denotes the number of matrices in the k-th cluster. The desired ranking list $\mathbf{f}$ is learned via the following function:

$$\begin{aligned} \Gamma = \min_{\hat{\mathbf{L}}^\mathbf{k}, \mathbf{f}, \mathbf{f}^\mathbf{k}} \frac{1}{2} \sum_{k=1}^{K} l_{intra}(\hat{\mathbf{L}}^\mathbf{k}, \{\mathbf{L}^{\mathbf{k}_1}, \cdots, \mathbf{L}^{\mathbf{k}_{S^k}}\}) + \\ \frac{\lambda_1}{2} \sum_{k=1}^{K} l_{man}(\hat{\mathbf{L}}^\mathbf{k}, \mathbf{f}^\mathbf{k}) + \frac{\lambda_2}{2} l_{inter}(\mathbf{f}, \mathbf{y}, \{\mathbf{f}^\mathbf{1}, \cdots, \mathbf{f}^\mathbf{k}\}), \end{aligned} \tag{6}$$

where l_{intra}, l_{man}, and l_{inter} respectively denote the loss of: 1) intra-group fusion, 2) manifold ranking, and 3) inter-group fusion. The intra-group fusion aims to learn a common Laplacian matrix $\hat{\mathbf{L}}^\mathbf{k} \in \mathbb{R}^{N \times N}$ to fuse the Laplacian matrices $\{\mathbf{L}^{\mathbf{k}_1}, \cdots, \mathbf{L}^{\mathbf{k}_{S^k}}\}$ in the k-th group. Upon the k-th common Laplacian matrix, the manifold ranking learns a ranking list $\mathbf{f}^\mathbf{k} \in \mathbb{R}^N$. Inter-group fusion combines rankings from different groups into the final ranking $\mathbf{f}$. $\mathbf{f}$ is further regularized by the historical ranking result $\mathbf{y}$ so that it satisfies the ranking smoothness. λ_1 and λ_2 are hyper-parameters to balance the three kinds of loss.

3.3.1 Intra-group Fusion. The intra-group fusion leans a common Laplacian matrix $\hat{\mathbf{L}}^{\mathbf{k}}$ to fuse the Laplacian matrices in the k-th group $\{\mathbf{L}^{\mathbf{k_1}}, \cdots, \mathbf{L}^{\mathbf{k_{S^k}}}\}$ by minimizing l_{intra},

$$\frac{1}{2}\sum_{i=1}^{S^k} tr\left((\hat{\mathbf{L}}^{\mathbf{k}} - \mathbf{L}^{\mathbf{k_i}})^T \mathbf{S}^{\mathbf{k_i}} (\hat{\mathbf{L}}^{\mathbf{k}} - \mathbf{L}^{\mathbf{k_i}})\right). \tag{7}$$

Thereinto, $\mathbf{S}^{\mathbf{k_i}} \in \mathbb{R}^{N\times N}$ is a diagonal matrix with,

$$S_{jj}^{k_i} = \begin{cases} 0, & \text{if the } j\text{-th entity misses the } i\text{-th channel,} \\ 1, & \text{otherwise.} \end{cases} \tag{8}$$

It is a selector to avoid the biases in the common Laplacian caused by data missing. A toy example with two graphs shown in Figure 2 illustrates the effects of the selector. In the example, data of the n-th entity in the k_1-th graph are missing. Thus, $\mathbf{S}^{\mathbf{k_1}}$ and $\mathbf{S}^{\mathbf{k_2}}$ are set as $\mathbf{I} \in \mathbb{R}^{N\times N}$ except $S_{nn}^{k_1} = 0$. So entries related to the n-th entity in the common Laplacian learned by the intra-group fusion are the same as those in $\mathbf{L}^{\mathbf{k_2}}$. However, those entries could be bias towards zero if there is no selectors in the intra-group fusion. This is why we claim by integrating selectors, our intra-group fusion alleviates the impacts of data missing.

3.3.2 Manifold Ranking. Given a common Laplacian matrix $\hat{\mathbf{L}}^{\mathbf{k}}$, manifold ranking learns a ranking list $\mathbf{f}^{\mathbf{k}}$, where similar entities obtain close ranks, via,

$$\min_{\mathbf{f}^{\mathbf{k}}} l_{man}(\hat{\mathbf{L}}^{\mathbf{k}}, \mathbf{f}^{\mathbf{k}}) = \mathbf{f}^{\mathbf{k}T} \hat{\mathbf{L}}^{\mathbf{k}} \mathbf{f}^{\mathbf{k}}. \tag{9}$$

3.3.3 Inter-group Fusion. As aforementioned, different local ranking lists are learned from different clusters, i.e., we have $\{\mathbf{f}^1, \mathbf{f}^2, \cdots, \mathbf{f}^{\mathbf{K}}\}$. The inter-group fusion learns a set of weights $\mathbf{b} = [b_1, b_2, \cdots, b_K] \in \mathbb{R}^K$ to get the desired ranking $\mathbf{f} = \sum_{k=1}^{K} b_k \mathbf{f}^{\mathbf{k}}$ and regulates the fused ranking to be smooth with the historical one by minimizing l_{inter},

$$(\sum_{k=1}^{K} b_k \mathbf{f}^{\mathbf{k}} - \mathbf{y})^T \mathbf{C} (\sum_{k=1}^{K} b_k \mathbf{f}^{\mathbf{k}} - \mathbf{y}),\ s.t.\ \sum_{k=1}^{K} b_k = 1, \tag{10}$$

where $\mathbf{C} \in \mathbb{R}^{N\times N}$ is diagonal matrix with $C_{jj} = c_j$. c_j is the pre-calculated weight of the j-th entity controlling the entity-aware ranking smoothness. Taking university ranking as an example, c_j is large for top universities while small for bottom ones.

3.4 Optimization

We adopt the alternating strategy to optimize the proposed model, until it converges.

3.4.1 Computing $\hat{\mathbf{L}}^{\mathbf{k}}$. To ease the optimization of $\hat{\mathbf{L}}^{\mathbf{k}}$, we set each common Laplacian as,

$$\hat{\mathbf{L}}^{\mathbf{k}} = \sum_{i=1}^{S^k} a_i^k \mathbf{L}^{\mathbf{k_i}},\ s.t.\ \sum_{i=1}^{S^k} a_i^k = 1, \tag{11}$$

and optimize each $\hat{\mathbf{L}}^{\mathbf{k}}$ independently keeping $\mathbf{f}$ and $\mathbf{b}$ fixed. After removing the fixed parts and substituting the constraint $\sum_{i=1}^{S^k} a_i = 1$ with Lagrange multipier δ, the objective function is rewritten as,

$$\min_{\mathbf{a}^{\mathbf{k}}} \frac{1}{2}\sum_{i=1}^{S^k} tr\left((\sum_{j=1}^{S^k} a_j^k \mathbf{L}^{\mathbf{k_j}} - \mathbf{L}^{\mathbf{k_i}})^T \mathbf{S}^{\mathbf{k_i}} (\sum_{j=1}^{S^k} a_j^k \mathbf{L}^{\mathbf{k_j}} - \mathbf{L}^{\mathbf{k_i}})\right) + \frac{\lambda_1}{2} \mathbf{f}^{\mathbf{k}T} \sum_{i=1}^{S^k} a_i^k \mathbf{L}^{\mathbf{k_i}} \mathbf{f}^{\mathbf{k}} + \delta(1 - \mathbf{e}^T \mathbf{a}^{\mathbf{k}}), \tag{12}$$

where $\mathbf{e} = [1, 1, \cdots, 1]^T \in \mathbb{R}^{S^k}$. We then take the derivative of Eqn.(12) regarding a_i^k, as follows,

$$\sum_{j=1}^{S^k} (a_j^k S^k - 1) tr\left(\mathbf{L}^{\mathbf{k_i}} \mathbf{S}^{\mathbf{k_i}} \mathbf{L}^{\mathbf{k_j}}\right) + \frac{\lambda_1}{2} \mathbf{f}^{\mathbf{k}T} \mathbf{L}^{\mathbf{k_i}} \mathbf{f}^{\mathbf{k}} - \delta. \tag{13}$$

Setting it to zero and rearranging the terms, all a_i^k's and δ can be learned by solving the following linear system,

$$\mathbf{M}\widehat{\mathbf{a}^{\mathbf{k}}} = \mathbf{u}, \tag{14}$$

where $\widehat{\mathbf{a}^{\mathbf{k}}} = [a_1^k, a_2^k, \cdots, a_{S^k}^k, \delta]^T \in \mathbb{R}^{S^k+1}$, $\mathbf{u} = [u_1, u_2, \cdots, u_{S^k}, 1]^T \in \mathbb{R}^{S^k+1}$, and $\mathbf{M} \in \mathbb{R}^{(S^k+1)\times(S^k+1)}$. M_{ij} and u_i are defined as follows,

$$\begin{cases} M_{ij} = S^k tr\left(\mathbf{L}^{\mathbf{k_i}} \mathbf{S}^{\mathbf{k_i}} \mathbf{L}^{\mathbf{k_j}}\right), & i, j \neq S^k + 1, \\ M_{ii} = 0, & i = S^k + 1, \\ M_{ij} = 1, & \text{otherwise,} \\ u_i = \sum_{j=1}^{S^k} tr\left(\mathbf{L}^{\mathbf{k_i}} \mathbf{S}^{\mathbf{k_i}} \mathbf{L}^{\mathbf{k_j}}\right) - \frac{\lambda_1}{2} \mathbf{f}^{\mathbf{k}T} \mathbf{L}^{\mathbf{k_i}} \mathbf{f}^{\mathbf{k}}. \end{cases} \tag{15}$$

3.4.2 Computing $\mathbf{f}$. By fixing $\hat{\mathbf{L}}^{\mathbf{k}}$'s and $\mathbf{b}$, we take the derivative of Eqn.(6) regarding $\mathbf{f}^{\mathbf{k}}$ and then reach the following linear system,

$$\mathbf{W}\widehat{\mathbf{f}} = \mathbf{t}, \tag{16}$$

which can be restated as,

$$\begin{bmatrix} \mathbf{W_{11}} & \mathbf{W_{12}} & \dots & \mathbf{W_{1K}} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{W_{K1}} & \mathbf{W_{K2}} & \dots & \mathbf{W_{KK}} \end{bmatrix} \begin{bmatrix} \mathbf{f}^1 \\ \vdots \\ \mathbf{f}^{\mathbf{K}} \end{bmatrix} = \begin{bmatrix} \mathbf{t_1} \\ \vdots \\ \mathbf{t_K} \end{bmatrix}, \tag{17}$$

where $\mathbf{W} \in \mathbb{R}^{KN\times KN}$ is a block matrix with $K \times K$ blocks; $\widehat{\mathbf{f}} = [\mathbf{f^1}^T, \mathbf{f^2}^T, \cdots, \mathbf{f^S}^T]^T \in \mathbb{R}^{KN}$ and $\mathbf{t} = [\mathbf{t_1}^T, \mathbf{t_2}^T, \cdots, \mathbf{t_K}^T]^T \in \mathbb{R}^{KN}$ are both block vectors with K blocks; $\mathbf{W_{kj}}$ and $\mathbf{t_k}$ are defined as follows,

$$\begin{cases} \mathbf{W_{kk}} = \lambda_1 \mathbf{L}^{\mathbf{k}} + \lambda_2 b_k^2 \mathbf{C}, & k = j, \\ \mathbf{W_{kj}} = \lambda_2 b_k b_j \mathbf{C}, & \text{otherwise,} \\ \mathbf{t_k} = \lambda_2 b_k \mathbf{C}\mathbf{y}. \end{cases} \tag{18}$$

As $\mathbf{t}$ can be treated as a constant vector as $\mathbf{b}$ is fixed, $\mathbf{W}$ is apparently invertible. We thus can derive the closed-form solution of $\widehat{\mathbf{f}}$ as,

$$\widehat{\mathbf{f}} = \mathbf{W}^{-1}\mathbf{t}. \tag{19}$$

Finally, $\mathbf{f}$ is updated based on the solved $\widehat{\mathbf{f}}$ as,

$$\mathbf{f} = \sum_{k=1}^{K} b_k \mathbf{f}^{\mathbf{k}}. \tag{20}$$

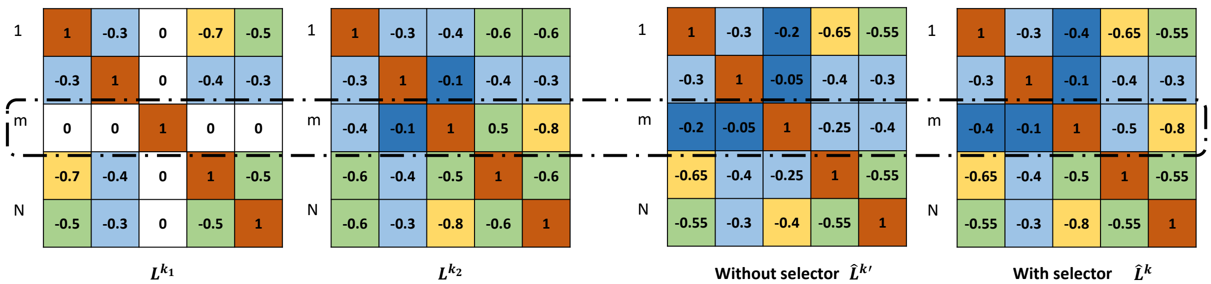

Figure 2: A toy example to illustrate the impact of missing data. $\mathbf{L^k}$ and $\mathbf{L^{k'}}$ are the common Laplacian learned by the intra-group fusion with and without selectors.

Table 1: Statistics of the collected multi-channel data.

Channels	Sources	#Universities	#Items	Duration
Official Channel	MOE	743	96,551	13.06-15.06
	Sina Weibo	721	10,912,234	15.01-16.05
Mass Media Channel	Baidu News	743	508,851	15.01-16.05
Academic Channel	Microsoft Academic	456	1,211,102	11.01-16.03
Employment Channel	LinkedIn	411	411	-
	iPIN	722	722	-
General User Channel	Sina Weibo	573	2,025,777	15.01-16.05

3.4.3 Computing $\mathbf{b}$. We first fix $\hat{\mathbf{L}}^{\mathbf{k}}$'s and $\mathbf{f^k}$'s, and then substitute the constraint $\sum_{k=1}^{K} b_k = 1$ into the objective function with Lagrange multiplier δ and rewrite it without fixed part as,

$$\min_{\mathbf{b}} \frac{\lambda_2}{2}(\sum_{k=1}^{K} b_k\mathbf{f^k} - \mathbf{y})^T\mathbf{C}(\sum_{k=1}^{K} b_k\mathbf{f^k} - \mathbf{y}) + \delta(1 - \mathbf{e}^T\mathbf{b}), \tag{21}$$

where $\mathbf{e} = [1, 1, \cdots, 1]^T \in \mathbb{R}^K$. We then take the derivative of Eqn.(21) regarding b_k and obtain,

$$\mathbf{M}\widehat{\mathbf{b}} = \mathbf{u}, \tag{22}$$

where $\widehat{\mathbf{b}} = [b_1, b_2, \cdots, b_K, \delta]^T \in \mathbb{R}^{K+1}$, $\mathbf{u} = [\lambda_2\mathbf{f^1}^T\mathbf{C}\mathbf{y}, \lambda_2\mathbf{f^2}^T\mathbf{C}\mathbf{y}, \cdots, \lambda_2\mathbf{f^K}^T\mathbf{C}\mathbf{y}, 1]^T \in \mathbb{R}^{K+1}$, and $\mathbf{M} \in \mathbb{R}^{(K+1)\times(K+1)}$ with M_{kj}, as follows,

$$\begin{cases} M_{kj} = \lambda_2\mathbf{f^k}^T\mathbf{C}\mathbf{f^j}, & k, j \neq S+1, \\ M_{kk} = 0, & k = S+1, \\ M_{kj} = 1, & \text{otherwise}. \end{cases} \tag{23}$$

4 CHINESE UNIVERSITY RANKING

4.1 Data Collection

In this work, we take the Chinese university ranking as a case study of social indicator computation. For each university, we first collected Web data from five channels. They are the official, mass media, academic, employment, and general user channels. The statistics of the collected data are summarized in Table 1.

4.1.1 Official Channel. Official channel contains the primary information of universities, such as student quality, official activities, and development plans, which plays a pivotal role in inferring university quality. Data in official channel are usually released by government agencies and university themselves. They includes: 1) MOE[15]. From the platform of MOE, we collected university profiles, such as location and category, and the enrollment score of universities from 2013-2015[16]. 2) Sina Weibo[17]. Sina Weibo is one of the most popular SNSs in China. Most Chinese universities publicize their official activities and announcements through their official Sina Weibo accounts. We thus crawled the historical posts from such accounts.

4.1.2 Mass Media Channel. It contains insights of mass media which uncovers the hot topics, events, discoveries, and even criticisms related to universities. News reports from mass media are usually formalized by professional journalists with incisiveness of arguments, and hence their opinions are objective. To take full advantages of such opinions, we collected news reports mentioned the universities of interest from Baidu News[18].

4.1.3 Academic Channel. This channel contains academic records of universities showing their academic contributions and influences. Such records are available from online bibliographic databases. In this work, given a university, we collected papers whose authors' affiliation is the given university and papers' citations from Microsoft Academic.

4.1.4 Employment Channel. Employment channel contains employment status of universities' graduate students. This is one of the key factors related to university quality, because most of students pursue higher education for better employment. The employment data are accessible through employment-oriented SNSs and third party data analysis companies. They include: 1) iPIN[19]. We collected employment data of the university's graduate students, including average salary, working location distribution, and male-female ratio, from its homepage in iPIN, a data analysis company in China. 2) LinkedIn. We collected *People also viewed* information from universities' homepage in LinkedIn to infer employment similarity among universities.

[15] http://gaokao.chsi.com.cn/.
[16] In China, last year high school students first take part in the National College Entrance Examination (NCEE). They then apply for universities based on their NCEE scores. Regarding applications from students, the university selects students by their scores from high to low. The lowest score of the selected students is released as the enrollment score of the university.
[17] http://weibo.com/.
[18] http://news.baidu.com/.
[19] http://www.ipin.com/.

Table 2: Features extracted from the multi-channel data.

Channels	Semi-structured Data	Dimension	Unstructured Data	Dimension
Official Channel	NCEE_enrollment_line, category, is_985, is_211, key_subjects_count, city, fans_count, followers_count, posts_count, comments_count, likes_count, etc.	78	topics	56
Mass Media Channel	monthly_reports_count	16	topics, sentiment	95
Academic Channel	papers_count, first_author_papers_count, cooperated_papers_count, authors_count, citations_count, citations_author, citations_paper	13	-	0
Employment Channel	average_sallary, average_sallary_top5_subjects, working_city, male_female_ratio, similar_universities	443	-	0
General User Channel	posts_count, reposted_count, likes_count, comments_count,	4	topics, sentiment	81

4.1.5 General User Channel. It contains public impressions, attitudes, and sentiment polarities of universities shared in SNSs posts, signaling the reputation of universities. We hence collected posts mentioning the given university from Sina Weibo.

4.1.6 Historical Ranking Result. The historical ranking result $\mathbf{y}$ is estimated from three most popular Chinese university rankings: CUAA, RCCSE, and CAMS (Wu Shulian). To generate a relatively objective historical ranking list, we averagely fused ranking results in 2015 of these three traditional rankings. It should be noted, in the future, the historical ranking result can be obtained from our previous release rather than the result of traditional rankings.

4.2 Feature Extraction

Regarding the collected multi-channel data, we extracted three types of features to describe each university: 1) Sentiment features. We noticed that data in mass media and general user channels convey the attitude and sentiment of users [26]. We thus utilized the Chinese microblog sentiment analysis tool [22] to judge the polarity of contents from the mass media and general user channels. For each given input, this tool would generate a three dimension distribution to denote its probability to be negative, neutral, and positive. 2) Topic features. According to our observation, contents in the official, mass media, or general user channel about similar universities are likely to express similar topics. For instance, reports from mass media may have a higher probability to report the topics of "research achievements" and "technologies" for top universities. Inspired by this, we explored the topic distributions over official, mass media, and general user channel. In particular, we generated topic distributions using Latent Dirichlet Allocation [6], which has been widely used in topic modeling. 3) Statistic features. Quality of universities are directly reflected by the volume of statistics, for instance, the average salary of graduate students, the number of publications, and the NCEE enrollment scores. Together with the sentiment and topic features, the statistical features are summarized in Table 2.

5 EXPERIMENT

5.1 Experimental Settings

5.1.1 Entity-aware Ranking Smoothness. As our historical ranking $\mathbf{y}$ is estimated from CUAA, RCCSE, and CAMS, the ranking smoothness weight of the i-th univeristy C_{ii} in Eqn.(6) is assigned as the ratio of ranking lists containing the i-th university among CUAA, RCCSE, and CAMS.

Table 3: Statistics of the ground truth.

Universities	University Pairs			
	Label 1	Label 0	Label -1	All
640	178,342	48,448	178,342	405,132

5.1.2 Ground Truth. Establishing the ground truth for university ranking from scratch by ourselves is extremely resource consuming and not reliable. We thus turn to justify the 2016 university ranking results by our model in a pair-wise fashion. In particular, although the traditional university ranking results of CUAA, RCCSE, and CAMS are time- and resource-consuming, they were generated by experts with sufficient domain knowledge. They are hence remarkably reasonable. We established the pair-wise ground truth upon their 2016 results. Given a pair of universities $< u_i, u_j >$, if all CUAA, RCCSE, and CAMS rank u_i as better or worse than u_j, then the pair is labeled as 1 or -1, respectively. Otherwise, it is labeled as 0, meaning u_i and u_j are not distinguished. The statistics of the constructed ground truth are shown in Table 3.

5.1.3 Evaluation Metrics. The performance of our model and the baselines was measured by Cohen's kappa coefficient (κ) [30], macro-averaged precision (Pre), macro-averaged recall (Rec), macro-averaged F1 score (F1), and micro-averaged F1[20] [5]. We also carried out the significance test and reported the p-values.

5.1.4 University Pair Tagging. Regarding the learned ranking list $\mathbf{f}$, the label of the i-th and j-th universities was set as,

$$\begin{cases} sign(f_i - f_j), & \text{if } |f_i - f_j| > \theta, \\ 0, & \text{otherwise.} \end{cases} \tag{24}$$

θ was set as 0.004, as it outperformed the others in $\{0.001, 0.002, \cdots, 0.01\}$ during our preliminary experiments.

5.1.5 Compared Methods. To show the effectiveness of our scheme, we compared it with the following state-of-the-art methods,

- **Historical Ranking (HR)**: It takes the historical ranking list $\mathbf{y}$ as current ranking, i.e., $\mathbf{f} = \mathbf{y}$.

[20]Regarding our ground truth, the micro-averaged Pre, Rec, and F1 are equal, we thus only reported the F1.

- **NCEE Enrollment Scores (NES)**: It ranks universities with higher average NCEE enrollment scores in the front.
- **Early Fusion (EF)**: It first concatenates features of all channels, constructs a simple graph as described in Section 3.1.1, and then performs manifold ranking with the simple graph [21].
- **Late Fusion (LF)**: It separately performs simple graph construction and manifold ranking upon each channel, and then averagely combines the generated ranking lists [8].
- **Joint Learning (JL)**: **JL** constructs simple graph on each channel and then learns a common ranking list by jointly regulating it on simple graph Laplacian matrices of all channels [38].
- **Subspace Learning (SL)**: It first maps multi-channel data to subspaces with the same dimension by dictionary learning [4]. Based on the representations in subspaces, it then performs ranking via **JL**.

It should be noted that **EF**, **lF**, **JL**, and **SL** also encourage the final ranking results to be close to the initial ranking one.

5.2 Parameter Tuning and Sensitivity Analysis

In the proposed **GMR**, we have two implicit parameters and two explicit parameters. They are the number of nearest neighbors k in hypergraph construction, the number of clusters K, λ_1 and λ_2. During the experiments, we heuristically set k as 5 based on our observation on the data. The optimal values of the remaining parameters were carefully tuned on the development set. In particular, in each round of the 5-fold cross-validation, we divided our dataset into two parts: 80% of the universities pairs were used for tuning, 20% were used for testing. We employed grid search to select the optimal parameters with a small but adaptive step size. The search ranges for λ_1, λ_2, and K are $[0.1, 100]$, $[10, 10, 000]$, and $[1, 8]$. The parameters corresponding to the largest micro-averaged F1 were used to report the final results. For other compared methods, the procedures of parameter tuning are the same to ensure fair comparison.

Take the tuning procedure of one round in the 5-fold cross validation as an example, we observed that our model reached the optimal performance when $K = 3$, $\lambda_1 = 9$, $\lambda_2 = 500$. Figure 3 illustrates the performance of our model with respect to these three parameters. This was accomplished by varying one and fixing the others with optimal values.

5.3 Performance Comparison

The comparison results between our proposed **GMR** and baselines are summarized in Table 4. From this table, we have the following observations: 1) **NES** and **HR** perform worse than the others. This tells us that the graph-based ranking methods successfully leverage the multi-channel Web data and hence improve the ranking performance. 2) **LF** performs worse than the other multi-channel ranking methods. This is because it equally fuses channels instead of distinguishing them with different confidences. 3) **GMR** shows superior performance to the others. This justifies the importance of integrating the block-wise data completion, cluster-wise ranking, and ranking results fusion within a unified model. 4) All the p-values of the pairwise significance t-test based on 5-fold evaluation are greatly much than 0.05. This demonstrates that the performance improvements achieved by our model over the baselines are statistically significant.

5.4 Component-wise Comparison

We also carried out experiments to justify the effectiveness of each component in the proposed **GMR**. In particular, we compared the following methods by disabling some terms of our objective function in Eqn.(6).

- **GMR-HRC**: We set $\mathbf{C}$ to an identity matrix to ignore the historical ranking confidence. **GMR-MD**: We set all $\mathbf{S}^{\mathbf{k_i}}$'s to identity matrices to ignore the missing data problem.
- **GMR-DH**: In this method, the semi-structured and unstructured data from one channel are directly concatenated and used to construct the simple graph.
- **GMR-CC**: It learns a common space from all channels and then performs ranking on the common representations to ignore inter-group fusion.

Table 5 displays the performance of the above methods. From this table, we observed that: 1) **GMR** performs better than the remaining methods. It confirms the effectiveness of jointly considering of the block-wise data completion, cluster-wise ranking, and ranking results fusion. 2) **GMR-HRC** performs much worse than **GMR**. It shows the importance of carefully setting entity-aware ranking smoothness, and hence assigning identical ranking smoothness to all the entities may lead to suboptimal performance. 3) It is interesting to see that **GMR-DH** achieves better macro-averaged Rec since it predicts more university pairs to be 0, a relatively rare class. **GMR** fails to identify such pairs since the similarity between entities in the pair is carved by their nearest neighbors. Although sacrificing such pairs, **GMR** successfully recalls more pairs in total, and hence verifies the robustness of hypergraph.

5.5 Channel Comparison

To measure the representation ability of each channel, we held one channel out and fed the others into our **GMR** model. The experimental results are displayed in Table 6. We observed: 1) The performance of **GMR** decreases more when the official channel is not fed into. This suggests that the official channel provides more informative and important cues for university ranking. 2) With all channels fed into, **GMR** performs best, which indicates that universities can be comprehensively described by more channels. 3) All the p-values of the pairwise significance t-test are greatly smaller than 0.05, which verifies the significance of performance improvements.

5.6 Development Set Comparison

Parameter tuning of our **GMR** relies on the development set. Towards the whole ranking list generation in 2016, it is arbitrary to directly tune parameters with the ground truth. We thus constructed a new development set (DS2015) from the ranking results of CUAA, RCCSE, and CAMS in 2015 for whole ranking list generation task. University pairs in DS2015 were labeled in the same way as the ground truth. To uncover the effectiveness of the constructed development set, we compared it with arbitrarily tuning parameters based on ground truth (GT). Corresponding performances are shown in Table 7. As can be seen, the performance

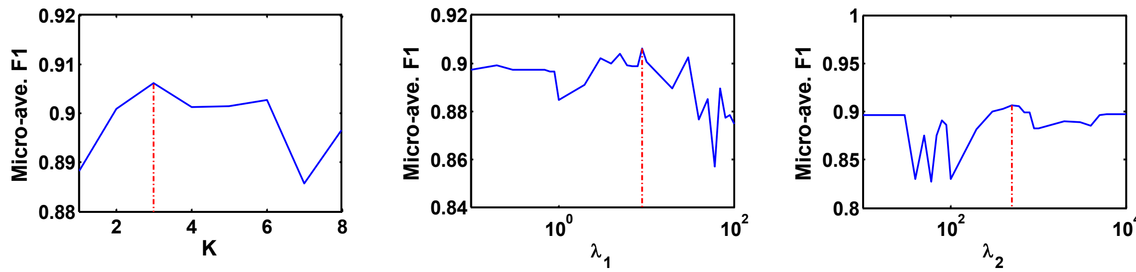

Figure 3: Procedure of parameter tuning by varying one and fixing others. The red dotted line marked the optimal settings.

Table 4: Performance comparison between our method and baselines.

Methods	Macro Averaged				Micro Averaged		κ	p-value@κ
	Pre	Rec	F1	p-value@F1	F1	p-value@F1		
NES	0.507±3e-7	0.576±4e-7	0.540±3e-7	8e-10	0.761±7e-1	4e-9	0.572±2e-6	3e-9
HR	0.776±4e-4	0.658±2e-7	0.618±2e-7	1e-9	0.868±3e-7	2e-7	0.764±1e-6	8e-8
EF	0.833±5e-6	0.786±4e-6	0.801±3e-6	1e-5	0.896±5e-7	2e-6	0.823±1e-6	2e-6
LF	**0.844±6e-5**	0.692±2e-6	0.684±6e-6	2e-8	0.878±7e-7	4e-7	0.784±2e-6	2e-7
JL	0.826±1e-5	0.789±6e-6	0.802±7e-6	2e-4	0.894±3e-6	1e-5	0.820±8e-6	1e-5
SL	0.822±1e-5	0.788±2e-6	0.800±4e-6	8e-5	0.893±2e-6	6e-6	0.818±6e-6	6e-6
GMR	0.841±1e-5	**0.797±4e-6**	**0.812±4e-6**	-	**0.906±2e-6**	-	**0.840±4e-6**	-

Table 5: Performance comparison among components in our GMR model.

Methods	Macro Averaged				Micro Averaged		κ	p-value@κ
	Pre	Rec	F1	p-value@F1	F1	p-value@F1		
GMR-HRC	0.838±9e-6	0.783±2e-6	0.800±2e-6	2e-6	0.898±8e-7	2e-6	0.825±2e-6	2e-6
GMR-MD	0.834±8e-6	0.794±5e-6	0.809±5e-6	8e-4	0.902±1e-6	3e-5	0.834±3e-6	3e-5
GMR-DH	0.833±3e-6	**0.798±7e-6**	0.811±5e-6	9e-2	0.903±8e-7	6e-4	0.835±2e-6	1e-3
GMR-CC	0.839±1e-5	0.786±2e-5	0.802±2e-5	5e-4	0.903±3e-6	4e-3	0.834±8e-6	3e-3
GMR	**0.841±1e-5**	0.797±4e-6	**0.812±4e-6**	-	**0.906±2e-6**	-	**0.840±4e-6**	-

Table 6: Performance comparison among channels with our GMR model. Official, Media, Academic, Employ, Crowd respectively denote the official, mass media, academic, employment, and general user channels.

Methods	Macro Averaged				Micro Averaged		κ	p-value@κ
	Pre	Rec	F1	p-value@F1	F1	p-value@F1		
No-Official	0.784±5e-6	0.807±7e-6	0.791±6e-6	7e-6	0.867±4e-6	9e-8	0.783±9e-6	1e-7
No-Media	0.810±2e-6	0.800±3e-6	0.805±2e-6	6e-5	0.893±6e-7	6e-7	0.820±2e-6	9e-7
No-Academic	0.828±1e-5	0.809±8e-6	**0.817±9e-6**	1e-3	0.902±3e-6	2e-4	0.835±9e-6	6e-4
No-Employ	0.831±9e-6	0.780±3e-6	0.795±4e-6	6e-5	0.900±1e-6	3e-4	0.829±3e-6	2e-4
No-Crowd	0.815±1e-5	**0.812±8e-6**	0.814±1e-5	8e-2	0.895±4e-6	1e-5	0.825±1e-5	3e-5
All	**0.841±1e-5**	0.797±4e-6	0.812±4e-6	-	**0.906±2e-6**	-	**0.840±4e-6**	-

of DS2015 is comparable to that of GT, which indicates that our scheme is truly usable and works appropriately without the latest ranking results of CUAA, RCCSE, and CAMS.

5.7 User Study

To further investigate the effectiveness of our scheme, we invited 17 volunteers[21] to evaluate our generated ranking list and the ranking results of RCCSM, CAMS, and CUAA in 2016. Each volunteer was presented the top-30 of each ranking list and was required to assign one of eleven scores (ranging from 0 to 10) according to their subjective opinions. These scores represent the strength of the volunteer's agreement with the given ranking list. The volunteer would assign score s to a ranking list, if the number of universities whose ranks are consensus with the expectations of the volunteer belongs to the range $(3(s-1), 3s]$. For instance, if the volunteer thinks that 20 of the top-30 universities are ranked as exepected, then he/she will assign 7 to the given ranking list. The user study results are summarized in Table 8. As can be seen, our ranking achieves higher average score than traditional rankings. Besides, over half of the volunteers assign highest score to our result among all the given rankings. It shows that our ranking results are comparable to those traditional rankings and further verifies the usability of our scheme.

[21]The volunteers are graduate students, research fellows, and visiting professors in different majors of National University of Singapore, coming from mainland China.

Table 7: Performance comparison between development sets towards the whole ranking list generation.

Methods	Macro Averaged			Micro Ave.	κ
	Pre	Rec	F1	F1	
GT	**0.841**	**0.797**	**0.812**	**0.906**	**0.840**
DS2015	**0.841**	0.793	0.809	0.905	0.837

Table 8: Performance comparison among our ranking and traditional Chinese university rankings.

Ranking Results	Ours	RCCSE	CAMS	CUAA
Average Scores	**8.12±0.99**	7.59±1.51	7.71±1.35	8.06±0.81
Highest Score Percentage	53%	18%	35%	**59%**

6 CONCLUSION AND FUTURE WORK

This paper presented a novel and automatic scheme for social indicator computation by exploring multi-channel Web data. This scheme integrates the block-wise data completion, cluster-wise ranking, and ranking results fusion within a unified model. The scheme is successfully applied to Chinese university ranking, a case study of social indicator. We observed that: 1) the official channel dominates the university ranking performance; and 2) the generated ranking results are comparable to the traditional Chinese university rankings, which demonstrates the effectiveness and rationality of our scheme.

In future, we plan to apply our scheme to other social indicator applications and consider the complementary relatedness among channels instead of simple correlations.

Acknowledgements We would like to thank the anonymous reviewers for their valuable comments. NExT research is supported by the National Research Foundation, Prime Ministers Office, Singapore under its IRC@SG Funding Initiative.

REFERENCES

[1] S. Agarwal. 2006. Ranking on graph data. In *ICML*. ACM, 25–32.
[2] G. Andrew, R. Arora, J. A. Bilmes, and K. Livescu. 2013. Deep canonical correlation analysis.. In *ICML*. 1247–1255.
[3] A. Armstrong, R. Francis, M. Bourne, and I. Dussuyer. 2002. *Difficulties of developing and using social indicators to evaluate government programs: a critical review.* Ph.D. Dissertation.
[4] S. Bahrampour, N. M. Nasrabadi, A. Ray, and W. K. Jenkins. 2016. Multimodal task-driven dictionary learning for image classification. *TIP* 25, 1 (2016), 24–38.
[5] P. N. Bennett and N. Nguyen. 2009. Refined experts: improving classification in large taxonomies. In *SIGIR*. ACM, 11–18.
[6] D. M. Blei, A. Y. Ng, and M. I. Jordan. 2003. Latent dirichlet allocation. *JMLR* 3, Jan (2003), 993–1022.
[7] M. J. Boskin. 2008. Consumer price indexes. *Concise Encyclopedia of Economics* (2008).
[8] E. Bruno and M. M. Stephane. 2009. Multiview clustering: a late fusion approach using latent models. In *SIGIR*. ACM, 736–737.
[9] J. Bu, S. Tan, C. Chen, C. Wang, H. Wu, L. Zhang, and X. He. 2010. Music recommendation by unified hypergraph: combining social media information and music content. In *MM*. ACM, 391–400.
[10] J. Chen, X. Song, L. Nie, X. Wang, H. Zhang, and T. Chua. 2016. Micro Tells Macro: Predicting the Popularity of Micro-Videos via a Transductive Model. In *MM*. ACM, 898–907.
[11] X. Q. Cheng, P. Du, J. Guo, X. Zhu, and Y. Chen. 2013. Ranking on Data Manifold with Sink Points. *TKDE* 25, 1 (Jan 2013), 177–191.
[12] M. Dobrota, M. Bulajic, L. Bornmann, and V. Jeremic. 2016. A new approach to the QS university ranking using the composite I-distance indicator: Uncertainty and sensitivity analyses. *Journal of the Association for Information Science and Technology* 67, 1 (2016), 200–211.
[13] C. Dwork, R. Kumar, M. Naor, and D. Sivakumar. 2001. Rank aggregation methods for the Web. In *WWW*. ACM, 613–622.
[14] Organisation for Economic Co-operation and Development. 1976. *Measuring social well-being: a progress report on the development of social indicators.* OECD Publications Center.
[15] Y. Fu, T. M. Hospedales, T. Xiang, and S. Gong. 2015. Transductive multi-view zero-shot learning. *TPAMI* 37, 11 (2015), 2332–2345.
[16] W. Gao and P. Yang. 2014. Democracy is good for ranking: towards multi-view rank learning and adaptation in web search. In *WSDM*. ACM, 63–72.
[17] U. Gerdtham and B. Jönsson. 2000. International comparisons of health expenditure: theory, data and econometric analysis. *Handbook of Health Economics* 1 (2000), 11–53.
[18] C. Guarino, G. Ridgeway, M. Chun, and R. Buddin. 2005. Latent variable analysis: a new approach to university ranking. *Higher Education in Europe* 30, 2 (2005), 147–165.
[19] X. He, M. Gao, M. Y. Kan, and D. Wang. 2017. BiRank: Towards Ranking on Bipartite Graphs. *TKDE* 29 (2017), 57–71.
[20] X. He, M. Kan, P. Xie, and X. Chen. 2014. Comment-based Multi-view Clustering of Web 2.0 Items. In *WWW*. ACM, 771–782.
[21] J. Jeon, V. Lavrenko, and R. Manmatha. 2003. Automatic image annotation and retrieval using cross-media relevance models. In *SIGIR*. ACM, 119–126.
[22] F. Jiang, Y. Liu, H. Luan, M. Zhang, and S. Ma. 2014. Microblog sentiment analysis with emoticon space model. In *Chinese National Conference on Social Media Processing*. Springer, 76–87.
[23] N. Kapur, N. Lytkin, B. Chen, D. Agarwal, and I. Perisic. 2016. Ranking universities based on career outcomes of graduates. In *SIGKDD*. ACM, 137–144.
[24] J. M. Kleinberg. 1999. Authoritative sources in a hyperlinked environment. *J. ACM* 46, 5 (1999), 604–632.
[25] J. Lages, A. Patt, and D. L. Shepelyansky. 2016. Wikipedia ranking of world universities. *The European Physical Journal B* 89, 3 (2016), 1–12.
[26] L. Liao, X. He, Z. Ren, L. Nie, X. Huan, and T. Chua. 2017. Representativeness-aware Aspect Analysis for Brand Monitoring in Social Media. In *IJCAI*.
[27] Y. Liu, B. Gao, T. Liu, Y. Zhang, Z. Ma, S. He, and H. Li. 2008. BrowseRank: letting Web users vote for page importance. In *SIGIR*. ACM, 451–458.
[28] X. Lu, F. Wu, S. Tang, Z. Zhang, X. He, and Y. Zhuang. 2013. A low rank structural large margin method for cross-modal ranking. In *SIGIR*. ACM, 433–442.
[29] H. Ma, T. C. Zhou, M. R. Lyu, and I. King. 2011. Improving recommender systems by incorporating social contextual information. *TOIS* 29, 2 (2011), 9:1–9:23.
[30] O. Megorskaya, V. Kukushkin, and P. Serdyukov. 2015. On the relation between assessor's agreement and accuracy in gamified relevance assessment. In *SIGIR*. ACM, 605–614.
[31] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, and A. Y. Ng. 2011. Multimodal deep learning. In *ICML*. 689–696.
[32] Z. Nie, Y. Zhang, J. Wen, and W. Ma. 2005. Object-level ranking: bringing order to Web objects. In *WWW*. ACM, 567–574.
[33] L. Page, S. Brin, R. Motwani, and T. Winograd. 1999. *The PageRank citation ranking: bringing order to the Web.* Technical Report 66.
[34] S. Shekhar, V. M. Patel, N. M. Nasrabadi, and R. Chellappa. 2014. Joint sparse representation for robust multimodal biometrics recognition. *TPAMI* 36, 1 (2014), 113–126.
[35] K. Ura, S. Alkire, and T. Zangmo. 2012. A short guide to gross national happiness index. (2012).
[36] D. Wang, S. C. Hoi, P. Wu, J. Zhu, Y. He, and C. Miao. 2013. Learning to name faces: a multimodal learning scheme for search-based face annotation. In *SIGIR*. ACM, 443–452.
[37] D. Wang, T. Li, S. Zhu, and C. Ding. 2008. Multi-document summarization via sentence-level semantic analysis and symmetric matrix factorization. In *SIGIR*. ACM, 307–314.
[38] M. Wang, H. Li, D. Tao, K. Lu, and X. Wu. 2012. Multimodal graph-based reranking for web image search. *TIP* 21, 11 (2012), 4649–4661.
[39] W. Wang, R. Arora, K. Livescu, and J. Bilmes. 2015. On deep multi-view representation learning. In *ICML*. 1083–1092.
[40] X. Wang, L. Nie, X. Song, D. Zhang, and T. Chua. 2017. Unifying Virtual and Physical Worlds: Learning Toward Local and Global Consistency. *TOIS* 36 (2017), 4:1–4:26.
[41] W. Xu, X. Liu, and Y. Gong. 2003. Document clustering based on non-negative matrix factorization. In *SIGIR*. ACM, 267–273.
[42] Q. Yin, S. Wu, and L. Wang. 2015. Incomplete multi-view clustering via subspace learning. In *CIKM*. ACM, 383–392.
[43] H. Zhang, X. Shang, H. Luan, M. Wang, and T. Chua. 2017. Learning from Collective Intelligence: Feature Learning Using Social Images and Tags. *TOMCCAP* 13 (2017), 1:1–1:23.
[44] X. Zhou, M. Belkin, and N. Srebro. 2011. An iterated graph Laplacian approach for ranking on manifolds. In *SIGKDD*. ACM, 877–885.