J Vis
DOI 10.1007/s12650-016-0349-7

REGULAR PAPER

Hui Lei · Jing Xia · Fangzhou Guo · Yaoyao Zou · Wei Chen · Zhen Liu

Visual exploration of latent ranking evolutions in time series

Received: 3 August 2015 / Revised: 24 October 2015 / Accepted: 2 December 2015

Abstract Rankings are everywhere in the world and they change constantly. Detecting and analyzing ranking changes in a ranked list is of great importance for recommendation and information retrieval tasks. Common to existing approaches is that the latent correlations and trends of ranked lists are not taken into account. This paper introduces *RankEvo*, an integration of rank structuring and visualization techniques, for detecting and analyzing latent evolutions in ranking time series. We characterize the ranking changes by computing the similarities among the time series of ranked items and organizing similar items into itemsets, and further forming ranking evolutions. The integrated RankEvo system provides visualization and intuitive interactions for exploring correlated itemsets, concurrent ranking evolutions, as well as outlier items of ranked lists. The system also employs additional information windows on demand for evolution elaboration and verification. Case studies are conducted to demonstrate the effectiveness and usability of the RankEvo system in assisting users to understand ranking changes.

Keywords Time-series · Ranking · Rank changes · Temporal evolutions

Supported by National 973 Program of China (2015CB352503), Major Program of National Natural Science Foundation of China (61232012), National Natural Science Foundation of China (61422211, 61202279), Zhejiang Provincial Natural Science Foundation of China (LR13F020001, LQ12F02003) and the Fundamental Research Funds for the Central Universities.

H. Lei
Changsha University of Science and Technology, Changsha, China
E-mail: leihui88@163.com

J. Xia · F. Guo · Y. Zou · W. Chen
State Key Lab of CAD&CG, Innovation Joint Research Center for Cyber-Physical-Society System,
Zhejiang University, Hangzhou, China
E-mail: jjane.summer@gmail.com

F. Guo
E-mail: guofz1234@gmail.com

Y. Zou
E-mail: poluo@vip.qq.com

W. Chen
E-mail: chenwei@cad.zju.edu.cn

Z. Liu (✉)
School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou, China
E-mail: liuzhen@hdu.edu.cn

Published online: 17 February 2016

1 Introduction

Ranking is a basic and important tool applied almost everywhere, especially in fields such as web search, business intelligence, and marketing. Typical ranking operations take two modes: ranking by preference and ranking by process (Marden 1995). The former is identical to the statistical rating, and indicates the distribution of user preferences, such as the presidential election, hot search queries, or popular movie evaluation. The latter compares the item values subject to certain metrics and produces an ordered list. Well known examples include the US Fortune-500 and top-rated stocks.

Regardless of the mode it takes, a ranked list may vary frequently over time due to either the raters' changeable preferences or the evolving ranked items, both of which convey insightful temporal information for understanding ranked items and latent ranking patterns. For instance, in a phrase-based search application, the rankings of specific items (queries) may imply the time when these items become popular, the variation degrees of their rankings, and their temporal trends. Existing solutions for analyzing temporal ranking changes commonly track the ranking changes of each individual item over time, representing them with line charts. However, visualizing the curves or their similarities altogether can often cause heavy visual clutter if there are a large number of items. Another approach (Shi et al. 2012) emphasizes the ranking changes between successive time points by grouping ranked items into multiple segments at each time point. Although this method successfully depicts the ranking changes within and across multiple categories, it is hardly capable of disclosing complicated correlations of ranking changes. We argue that an ideal analytical solution of ranking changes should take the following aspects into consideration:

– Investigating the ranking change of a single item is not the ultimate goal. It is highly preferred to characterize a group of items that may have concurrently or correlatively evolved over time.
– Salient ranking changes of some ranked items may indicate an interesting pattern. However, it is quite challenging to disclose the potential theme by solely leveraging the ranking changes.
– Visualizing ranking changes in large time series can easily lead to visual clutter. Yet, designing an expressive visualization approach is a non-trivial task.

We present a novel visual exploration method for detecting and analyzing latent ranking evolutions in large time series. We define latent ranking evolutions as the ranking trends of ranked lists, including trends of a single item and multiple items of similar ranking fluctuations. We use the term "latent ranking evolutions" because it is important to distinguish our method of discovering concurrent or correlated ranking changes from typical time-of-rank line charts. And the term "latent" here indicates the correlations between items that cannot be revealed with typical methods. It should be noted that while there has been strong interest in representing ranking changes (Shi et al. 2012), to our knowledge this is the first approach for discovering latent rank evolutions and ranked items with concurrent trends. The key to discover item-wise concurrency is to group the ranked items into bundles based on the similarity of the item time series. The dynamic ranking evolution over time is visualized with multiple views in an integrated interface, RankEvo. The entire system has advantages over common methods as it illustrates the concurrent rank changes and interesting patterns indicated by salient ranking changes. The design of the RankEvo system can be easily adapted to work on large scale data or even streaming data with minor changes. In summary, the presented RankEvo system shows three main novelties:

– A general framework for analyzing latent ranking evolutions in time series that only handles the orders of ranked items,
– A novel and uniform visual representation for ranking changes that characterizes the occurrence, evolution, splitting, merging, and decline of an item group,
– A user-guided interactive process that allows dynamic manipulation of the ranked items of interest.

The rest of this paper is organized as follows. Related work is reviewed in Sect. 2. The flowchart of the proposed system is presented in Sect. 3. Sections 4 and 5 describe the representation and visual exploration of the latent ranking evolutions. Case studies are elaborated in Sect. 6, followed by discussions in Sect. 7. Finally we conclude the paper in Sect. 8.

2 Related work

Methods for statistical modeling of ranking can be categorized into two classes (Marden 1995): modeling the ranking process and modeling the population of rankers. The first one is the focus of our approach. Representative methods include the central ranking extraction and ranking distance models. For ranking modeling, various ranking distances (Critchlow 1985) are defined to measure the similarity between two rankings such as Kendall's tau distance and Spearman's rho distance. Special care is paid to incomplete ranking or ranking with tied items (Kidwell et al. 2008; Alvo and Cabilio 1985). In particular, expected weighted Hoeffding distance (Sun et al. 2010) (EWHD for short) can be used to estimate rankings for missing items.

However, all methods mentioned above deal with a single ranked list. Little attention has been paid to rank time series despite a number of visualization approaches for time-oriented data (Aigner et al. 2011). The time dimension of data can be linear or cyclic. The standard visualization for linear time dimension would be a two dimensional plot: one axis for time, the other for data value. Weber et al. (2001) designed a spiral-shaped time axis where careful selection of cycle length could reveal the cyclic pattern of the data. If the time dimension refers to date, a weight-encoded calendar view (Wijk and Selow 1999) can be adopted to visualize the value changes in different days. Tominski et al. (2004) made use of parallel coordinates to represent high dimensional time series. They placed other dimension axes around the central time axis so that the distribution of values in other dimensions could be clearly observed. Visualization applying time line structure (Plaisant et al. 1996) is also common in representing events, activities or even status, such as personal health records. For time-varying evolving topics in text data, TextFlow (Cui et al. 2011) is a technique for visualizing various evolution patterns from multiple topics. Also, time-oriented data visualization in diverse applications integrates all kinds of techniques corresponding to the datasets. Good examples include LiveRAC (McLachlan et al. 2008): time series data visualization of system management , Whisper (Cao et al. 2012): visual tracing tool of information diffusion, Storyline (Ogawa and Ma 2010): software evolution visualization, and Timebox (Hochheiser and Shneiderman 2004): an interactive tool for time series exploration.

This paper concerns the visualization of rank time series. The simplest way for visualizing rank time series is to embed each individual rank in a low-dimensional space and encode it with a heatmap representation. Similar models have been proposed in Sun et al. (2010) and Wei et al. (2012). By emphasizing the subtle rank changes, RankExplorer (Shi et al. 2012) adaptively segmented the ranking time series into groups, and extends ThemeRiver with color bars and glyphs to convey adjacent ranking changes and overall trends. To present ranking trends with spirals, RankClock (Batty 2006) employed radial coordinates to represent the ranks of different time points, however it can hardly scale to long rankings. Slope graph (Tufte 1986) connects time-varying items with lines to reveal temporal changes of item values. If the rank evolution of each ranked item is considered, it can be visualized as a rank band.[1] A similar visualization was adopted in TrajRank Lu et al. (2015), which ranks trajectory clusters by overall ranking score so that users are able to explore trajectories of higher importance. Another similar visualization is achieved in WikiTop50,[2] which depicted the variation of the top-50 ranked items one by one and connected the same items with curves. Even the simplest scatter plots assisted with user selections[3] can be applied to explore temporal trends of ranked items. Our work differs from existing solutions that our focus is the representation, visualization and interaction of the latent rank evolutions instead of a single item.

3 Approach overview

To clear up the confusion over similar terms used in this paper, we first provide several definitions related to ranking (examples shown in Fig. 1) before the overview of our RankEvo system.

Ranked list A ranked list is a list of items ordered by certain measurement at a certain time point.
Ranked item A ranked item is an item with its order in a ranked list.
Ranking time series A ranking time series is the series of ranked lists varying over time, like the four Top-10 rankings plotted in Fig. 1.
Item time series An item time series is the order series of an item in a ranking time series, plotted as the curve in blue in Fig. 1.
Concurrent itemset A concurrent itemset is a collection of ranked items on a ranked list that have similar ranking changes in their recent history and that may indicate an interesting activity or event. The two items in Fig. 1 form a concurrent itemset since their historical ranking time series follow a similar trend. If a ranked item's time series does not belong to any concurrent itemset, the item is called an isolated ranked item.

[1] Fortune 500 visualization. http://in.somniac.me/2010/01/fortune-500-visualization/. Accessed: 2015-04-11.

[2] Wikipedia Top 50. http://www.chrisharrison.net/index.php/Visualizations/WikiTop50. Accessed: 2015-04-11.

[3] Fortune 500. http://fathom.info/fortune500/. Accessed: 2015-04-11.

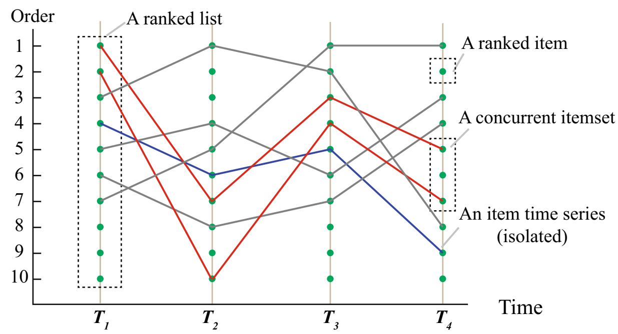

Fig. 1 Basic definitions related to time-varying ranking

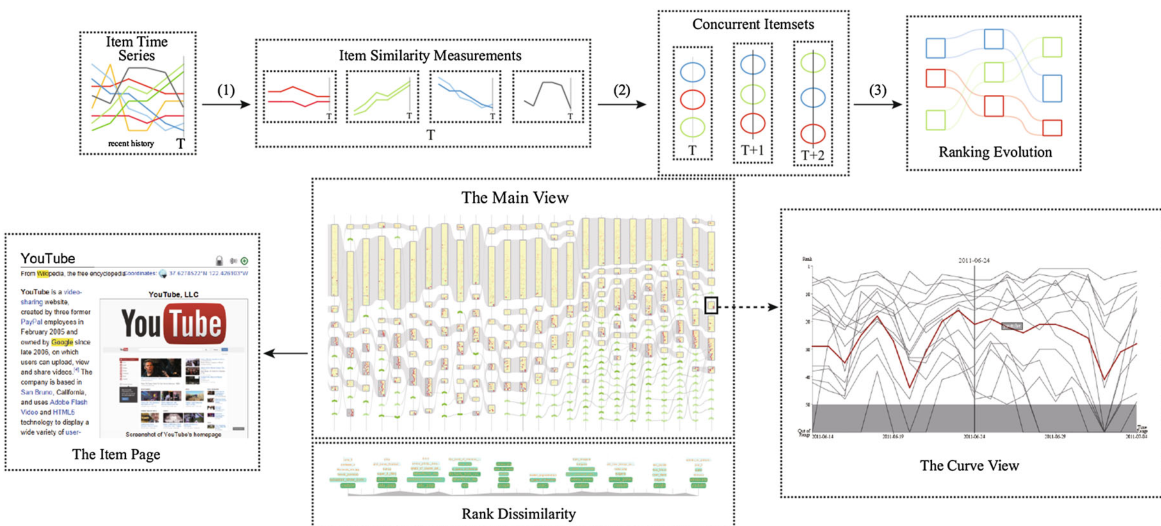

Fig. 2 The design model of our RankEvo system. The data preprocessing includes three stages, the dissimilarity measurement of the item time series, clustering items into concurrent itemsets, and the generation of ranking evolution. The RankEvo system visualizes items, itemsets and evolutions in a single view. The system also employs multiple extra windows to elaborate and verify item concurrency and independence with the item page and the curve view

We have designed our RankEvo system (Fig. 2) to accomplish the three main tasks for visual exploration of latent ranking evolutions: generation, visualization and elaboration of the ranking time series. Our approach first extracts relations of ranked items according to ranking time series, capturing concurrent ranked items. We then visualize the concurrency and the independence of the ranked items at each individual time point and visually aggregate the concurrency to represent temporal ranking evolutions. Concurrency elaboration and verification are supported via additional information windows on demand.

We adopt a band-based visual design to reveal the latent ranking evolutions of the ranked items. Users are able to highlight any items to display its evolving ranking orders, or expand any itemset to discover their concurrency as well. It also provides ranking line chart windows and item detail pages so that users can easily make comparisons among items or itemsets.

4 Representation of latent rank evolutions

To present the latent rank evolutions, our approach first calculates the dissimilarity of each pair of ranked items on a ranked list according to their recent ranking trends. An aggregation of the ranked items is conducted with an agglomerative clustering method to join items into concurrent itemset. In addition, similar concurrent itemsets are further associated over time to form the latent evolutions.

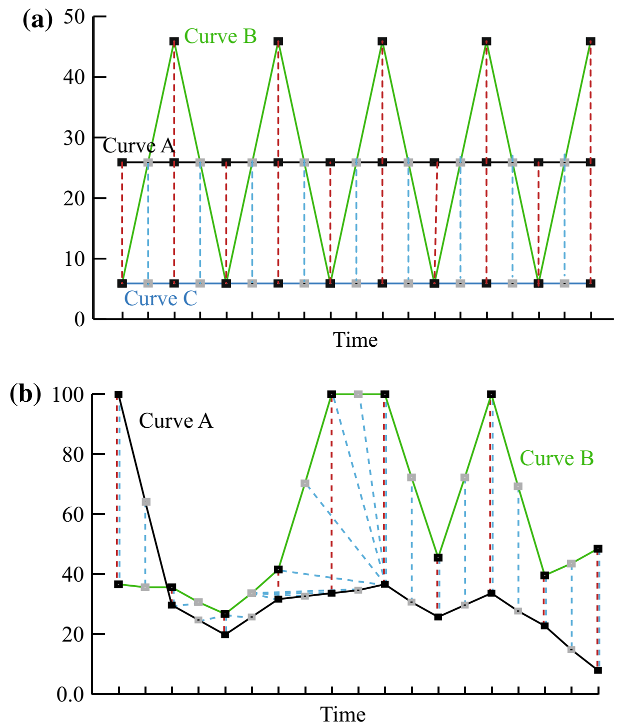

Fig. 3 Computing dissimilarity of time-of-rank curves. **a** Without interpolation, red dashed lines indicate the distances between curve A to curve B and curve A to curve C. The shortest paths and the standard deviations are the same for both of them. With interpolation (the *gray dots*), *blue dashed lines* are added for distance between curve A and curve C. Although the shortest paths remain the same, the length of the path and the standard deviation of the distances between curve A and curve C is much larger than that between curve A and curve B. **b** *Red dashed lines* indicate distance before interpolation and *blue lines* after interpolation. With interpolation (the *gray dots*), the shortest paths take a different route

4.1 Dissimilarity computation of ranked items

We evaluate the dissimilarity between two ranked items by the proximity of their associated item time series, with the assumption that two ranked items are potentially (but not necessarily) correlated if they share similar trends during their recent history. since dynamic time warping method (DTW) (Vintsyuk 1968) is widely used in time series distance computation, we adopt this method as the main factor for the similarity of two ranked items, while considering other factors as well. In specific, the dissimilarity of two ranked items at a certain time point is defined as the weighted sum of three factors: the dynamic time warping factor f_{dtw}, the average order factor f_{avgo}, and the loss compensation factor f_{comp}.

$$Dissim = \frac{w_{\text{dtw}} * f_{\text{dtw}} + w_{\text{comp}} * f_{\text{comp}} + w_{\text{avgo}} * f_{\text{avgo}}}{w_{\text{dtw}} + w_{\text{comp}} + w_{\text{avgo}}}$$

where w_{dtw}, w_{comp} and w_{avgo} are adjustable weights in correspondence with the three factors.

The dynamic time warping factor f_{dtw} estimates the temporal trends dissimilarity of two ranked items by computing the cost for matching their item time series curves. For a given pair of curves, a set of the shortest distances and the associated shortest paths are computed with DTW alignment. It is apparent that two item time series with similar trends have stable shortest distances, which indicates small standard deviation. Thus, we define f_{dtw} as the standard deviation of the shortest distances after DTW alignment (see the red alignments in Fig. 3b). However, in some cases such as that shown in Fig. 3a, the same length of shortest paths and standard deviation may result from completely dissimilar curves (curve A to curve B and curve A to curve C) due to a lack of distinction between positive and negative distance. We address the problem by adding interpolated points into the time series (gray dots in Fig. 3b). The interpolation scheme adjusts the alignment paths and its distance standard deviation. We do not apply the common offset practice, which substrates the time series to a constant offset, because it differs the result greatly from different offset distances (the mean order, the median order or the minimum order of the time series). And we would also distinguish two time series of large offset from two of small offset because the latter ones are relatively more correlated on the ranked list. Besides, we also assign a large number (e.g., twice of the item number in a rank) to discontinuous item time series, which drops off the ranked list and comes back sometimes. The time complexity of DTW alignment is $O(n^2)$, where n is the length of the time series.

Although f_{dtw} considers the influence of discontinuous item time series, the factor confidence is low when the original time series has too many missing values. To compensate this loss, we define a compensation factor f_{comp} as the sum of the missing items in two underlying time series. To further separate items with high and low rankings, an average order factor f_{avgo} is adopted. Note that the three factors f_{dtw}, f_{avgo} and f_{comp} are only related to the orders of the ranked items on the same ranked list, and are normalized to [0,1].

Adjustment of the three factors depends on specific applications. Applications with moderate ranking changes should weight more on f_{avgo} and less f_{comp}. And f_{dtw} is always the most essential factor in the dissimilarity computation. The parameter settings are discussed with more details in Sect. 7.2.

4.2 Concurrent itemset generation

The dissimilarities among item time series often indicate the trend differences of the underlying ranked items. Additional clustering of the series based on item dissimilarity can help users identify the ranked items and their time points that reveal dramatic ranking changes. The analysis of the corresponding causes and effects uncovers features of the latent ranking evolutions. In our clustering case, the cluster number is not fixed and an item time series is not necessarily assigned to a cluster. Therefore, we conduct an agglomerative clustering method as described in Algorithm 1, which groups ranked items based on pair-wise dissimilarity and filters small clusters. The time complexity of Algorithm 1 is $O(n^3)$ where n is the number of items on the a ranked list (e.g. k in a top-k rank).

Algorithm 1 Agglomerative Clustering

```
Initialization: Generate a set of clusters R by taking each item as a cluster
repeat
   Compare the distance between every pair of clusters in R
   get the minimum distance d of the corresponding clusters C_i and C_j
   if d <= the_distance_threshold then
      Merge C_i and C_j in R
   end if
until size(R) does not change
for each cluster C in R do
   if size(C) too small then
      Remove C from R
   end if
end for
return R
```

4.3 Latent evolution formation

If two concurrent itemsets, which are successive over time, share plenty of items in common, then we say that the two itemsets are consecutive. Consecutiveness between two itemsets is defined by the Jaccard Index (Jaccard 1901), which estimates the similarity of the two itemsets as an index between common ranked items and universe ranked items. The consecutiveness is filtered with a threshold, such that the smaller the threshold is, the more tightly related the two itemsets are.

We also apply the expected weighted Hoeffding distance (EWHD) (Sun et al. 2010) to compute the dissimilarity between two adjacent ranked lists (see the gray area of rank similarity in Fig. 2). Two ranked lists are of low EWHD distance if most of the items on the list maintains steady orders. The comparison between adjacent ranked lists is adopted to measure rapid changes of overall rankings quantitatively. EWHD can effectively handle partial or missing ranking information and only requires the orders of the items. Computing EWHD is fast for ranked lists with constant item numbers (e.g., the top-k ranks). To accelerate the computation, a lookup table can be built and reused by pre-computing all combinations of pairs of the underlying ranked lists.

5 Visual exploration of latent ranking evolutions

Our RankEvo system visualizes the latent ranking evolutions on a scrollable canvas, which gives an overview to all ranked lists with concurrent itemsets and isolated ranked items in all time. Users can track the evolution of one itemset on demand, explore correlations and validate the similarity of the items in extra linked windows. This exactly follows Ben Schneiderman's visualization mantra, "overview first, zoom and filter, then details on demand (Shneiderman 1996)". In addition to ranking evolutions, supplementary information is provided via extra linked windows instead of inline views or zoomable user interface, to allow parallel comparisons of multiple small patterns in large ranking time series space. Users can open multiple extra windows to make flexible comparisons among different ranking trends of different itemsets without consuming too much visual working memory (Ware 2012).

5.1 Visualizing ranking evolutions

First of all, the ranked list at each time point is represented as a ranking column evenly placed in parallel along the time axis, which form a time-rank coordinates system (see ranking columns in Fig. 2). The coordinates system uniformly locate all the concurrent itemsets and isolated ranked items as well as their evolving relations. Due to limited display space, RankEvo has ranked lists of a limited period of time (28 time steps) shown in one frame and enables users to select any other period with a scrollable slider.

Each concurrent itemset is represented as a rectangle, filled with small rectangles whose colors are encoded with ranking changes of each item member. Thus, the rectangle reflects recent change patterns of its item members, and further, the quality and different patterns of item clustering. As it shows in the Fig. 4, the change of ranking orders are represented from rising order to falling order as red to blue, light gray if the item falls off the ranking. Itemset with good clustering displays a uniform color distribution. The length of the itemset rectangle is set to be proportional to the size of the items in the cluster. RankEvo employs the average order of all the ranked items in the cluster to locate the concurrent itemset at the associated rank column.

To enhance the temporal relations, the RankEvo system employs a smooth curve-shaped band to connect the concurrent itemset between adjacent rank columns. The top and bottom boundaries of the band are represented with two 3-order Bezier curves, respectively, whose control points are set as two points on the individual rectangles and as two additional points between them. In addition, the vertical boundaries of two rectangles are used as the left and right boundaries of the band. The edge band in Fig. 5 illustrates the edges between consecutive itemsets. Normal bands and highlighting bands are encoded in green and blue, respectively.

The isolated ranked items are located on the rank columns according to their orders posterior to itemsets. To show the order change compared to its previous time point, an upwards or downwards glyph is used to encode each isolated item (see the isolated items in Fig. 6, the same item time series with that in Fig. 1). Also, transparency of color encodes the overall lifespan of the item. Item time series of isolated ranked item is shown with an edge connecting to the corresponding glyphs. After highlighting, item's original ranking order as well as relative position fixed by layout is plotted with a red curve and a green curve, respectively.

To give a summary of the ranked items at each time point, we draw the tags of the itemset's center item at the bottom (see rank dissimilarity in Fig. 2). In addition, the difference between two neighboring ranked

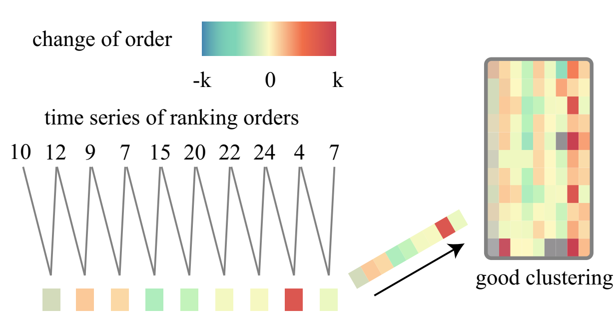

Fig. 4 Design of concurrent itemset. The change of ranking orders are represented from rising order to falling order as *red* to *blue*, *light grey* if the item falls off the ranking. Itemset with good clustering displays a generally uniform color distribution

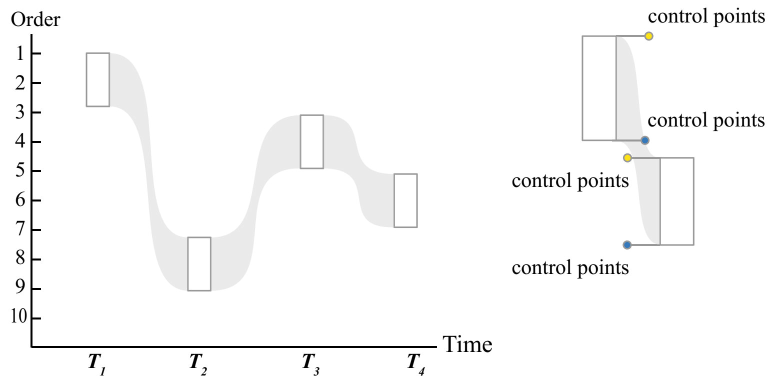

Fig. 5 A smooth curve-shaped band to connect the concurrent itemset between adjacent rank columns

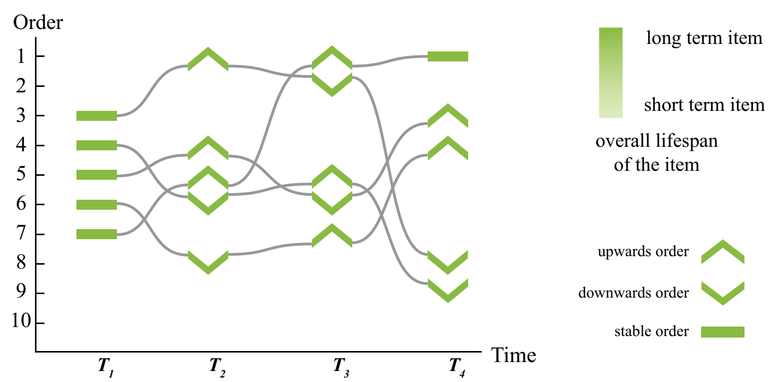

Fig. 6 Visual encoding of isolated ranked items. An upwards or downwards rank glyph is used to encode the ranking trend of the item, and transparency to encode its overall lifespan

lists at each time point is plotted as a stacked graph. The stacked graph enables users to take an overview of the rank variations and detect significant changes.

5.2 Visualizing concurrency of itemsets

The curve view visualizes the recent history and the near future of the item time series of a selected set of ranked items with line chart. Line chart is especially straightforward for displaying the details of item time series for a small number of concurrent items. Users can select a single or a bundle of curves for detailed investigation, comparison, and trend prediction. Because items may fall off the rank at some time points, we add a gray area at the bottom. The light gray axis in the view indicates the current given time point of all the ranked items in the cluster. The visualization represents both the most recent historic rank values (basis for item dissimilarity computation) and their near future values (if they have). Users can either evaluate the quality or discover the evolution of the cluster.

The curve view in Fig. 2 shows a list of Wikipedia queries belonging to the same cluster, where the highlighted curve represents the center of the cluster. We can observe similar ranking trends and subtle details of individual query item.

5.3 Visual exploration of ranking changes

Users may explore ranking changes in two styles: either focusing on a custom item of users' interest, or scanning through the evolution series, looking for specific patterns. For the first style, users first scroll to a period of time of interest, choosing a single item or itemset. Hovering an item highlights its overall time series, while hovering an itemset highlights its evolution bands as well as the tag cloud of all its member items. For the second style, long-lasting itemsets or rapid changes (indicated by rank similarity and the item evolution visualization) of overall patterns can be potentially interesting patterns.

To allow users to evaluate the relative ranking orders of all member items, users can click on the itemset for the curve view of a recent period of the time series of the corresponding items. Good or bad clustering can also be indicated from the overall trends of the line chart. To further explore the semantic correlations of the items, users can click on the highlighted tag or the curve for item's semantic page (the Wikipedia page for example), in which we adopt a degree-of-relevance highlighting technique (Ware 2012) to highlight its semantic-relevant member items in the itemset. Both the curve view and the item pages are extra linked windows triggered on demand so that users can open as many windows as it needs to make easy comparisons among items or itemsets.

6 Use cases

In this section, we demonstrate how the RankEvo system can be used to analyze latent rank evolutions with three real-life datasets: Wikipedia top-50 query series, US Fortune-500 series, and life expectancy at birth series. The construction of time-of-rank tree waves is completed in seconds and the system allows real-time user interactions in exploring latent rank evolutions.

6.1 Wikipedia top-50 query series

The page view statistics for Wikimedia projects[4] maintains raw page access records for all Wikipedia projects in all languages. We have collected the page view statistics for three months (from Jun. 1st to Oct. 27th in 2011) and generated daily top-50 queries as the rank time series data. We set items' latest 10 days as the recent time series in the similarity computation.

We follow the first exploration style and focus our attention on the topic of Steve Jobs's death in October 2011. As shown in Fig. 7, we zoom the RankEvo visualization into a 4-day range band, from the Oct. 6th to the Oct. 9th, when relevant queries appear on the rank. The queries of the cluster on the Oct. 7th (marked by the red rectangle) contain the name of Apple's successor "Tim_Cook", Apple's co-founder "Steve_Wozniak", Jobs's wife "Laurene_Powell_Jobs", the daughter he had with his high school girlfriend "Lisa_Brennan-Jobs", his biological younger sister "Mona_Simpson", "pancreatic_cancer", "Bill_Gates" and "Apple_Inc.". The branch starts from the top on Oct. 6th and falls down day by day, until out of the top-50 queries after Oct. 10th. It is interesting that another isolated cluster on Oct. 6th also includes queries related to Jobs (marked by the gray rectangle), such as "Pixar" graphics studio founded by Jobs, Apple's slogan "think_different" and "macintosh". The queries in this cluster are not as consistent as the the first branch, as they appear only on the first day of this time range. This comparison indicates that Wikipedia users are interested in Jobs's family and famous people related to him more than the technology background. The reason "Steve_Mozniak" is a hotter query than "Tim_Cook" and "Bill_Gates" may be that he played a part in the popular series "Big Bang Theory" and is known to a wider audience.

With the RankEvo system, we can also find that many of the query inputs are generated from the Internet links than manual inputs. For example, the queries of "Mona_Simpson_(novelist)" are more popular than "Mona_Simpson" in the data. It reveals that much traffic of Wikipedia comes from internal links rather than direct query. Wikipedia users tend to click the words they are interested in to jump between pages. In the example of Steve Jobs, while accessing the page of Steve Jobs they may jump to the web pages of his wife or his daughter via the internal links on his page.

6.2 US fortune-500 series

The US Fortune-100 dataset[5] published by Fortune magazine maintains an annual list of the top-500 companies in the United States during 1955 to 2012. We set items' latest 15 years as the recent time series in the similarity computation, and visualize the top 100 ranking variations since 1970. Following the second exploration style (both from the low similarity in the rank similarity or from the item/itemset visualization), we discovered the rapid change pattern in 1995 (in red rectangle).

4 http://dumps.wikimedia.org/other/pagecounts-raw/.

5 http://fortune.com/fortune500/.

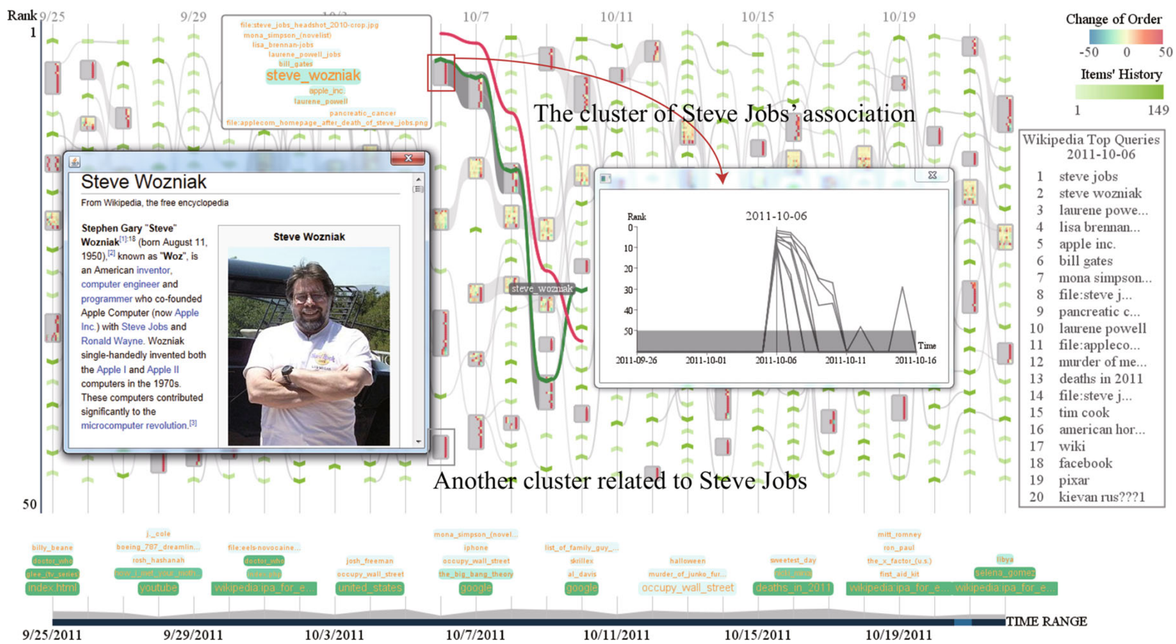

Fig. 7 The RankEvo visualization of Wikipedia top-50 queries related to Steve Jobs. The *red rectangle* contains Jobs's family and famous people related to him, such as his biological younger sister "Mona_Simpson" and his successor "Tim_Cook". The *gray rectangle* contains other elements associated with Jobs, such as "Pixar" graphics studio and "macintosh". The temporal evolutions of these two groups are significantly different. The line chart shows the ranking patterns of those items

Figure 8 reveals one significant pattern change in 1995, that many top ranking companies isolate themselves with others in 1995. This is caused by the addition of service-oriented companies such as "Walmart". This event is also recorded by several other visualization applications for this dataset (Shi et al. 2012). In the RankEvo system, the new companies introduced in 1995 become the roots of new branches in later years.

The RankEvo system also reveals that the rankings of Fortune-100 companies before 1995 are stable. Most of the businesses of these companies are in the fields of manufacturing, mining, or energy exploration. We can also recognize the rising of electronics industry, especially the IT industry. As shown in Fig. 8, the branch smoothly rising and growing strongly is the cluster of IT companies including "HP", "Motorola", "Digital Equipment", "Emerson Electric", and "Unisys".

6.3 Life expectancy at birth series

The life expectancy at birth dataset[6] published by the World Bank includes the expected life span of 198 countries during 1950 to 2011. First we convert the exact value into ranking orders from longer to shorter life span of all the countries. Then we set items' latest 10 years as the recent time series in the similarity computation. and visualize the ranking evolution since 1960.

Life expectancy at birth statistically indicates the life span of a newborn infant if prevailing patterns of mortality stay the same. As it shows in Fig. 9 that the clusters remain relatively steady while some items drop out of their historic clusters due to local rising or dropping of ranking orders. Such patterns as items isolating themselves from or joining other clusters usually reflect certain events about the corresponding countries.

Going through all the time range an obvious outlier item at the bottom of the rankings(the green curve) is uncovered. By highlighting the item, we find that its historic value series follows a rising and sharp dropping pattern while others are generally steady, which makes the item an outlier. The red curve reflects that the ranking order of the item rose greatly from the years 1953 to 1958, however, declined even more from 1959 to 1961. After 1961, the order slowly climbed back and maintained steady. Referring to the history of the item—"China", it reflects the development of the country in the 1950s. China put forward the first 5-year

[6] http://data.worldbank.org/indicator/SP.DYN.LE00.IN.

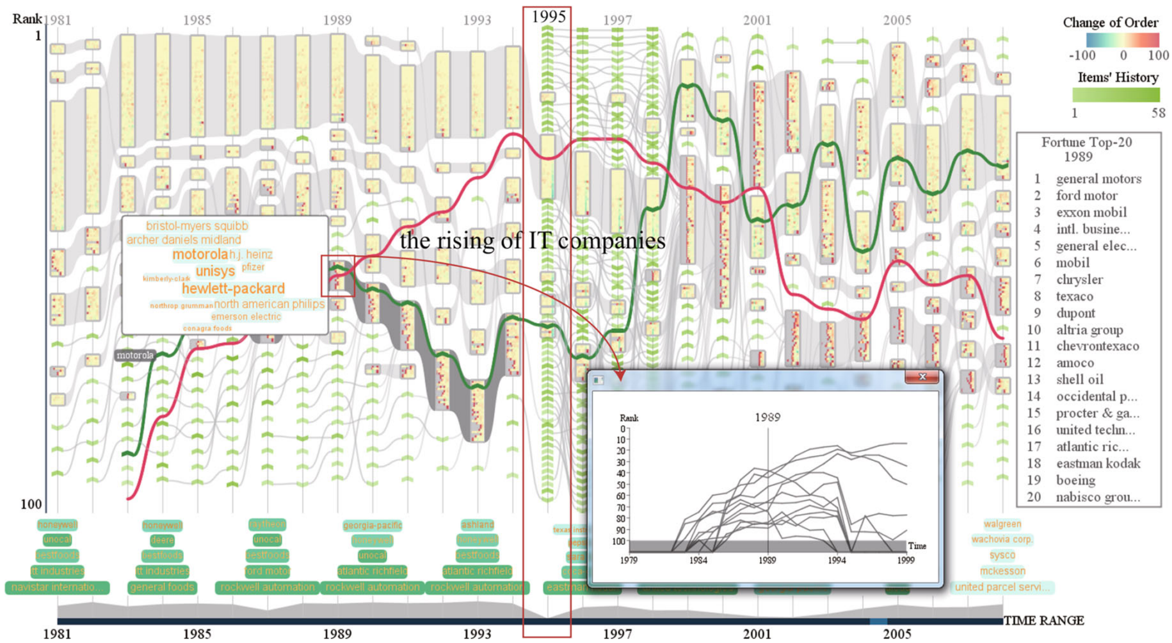

Fig. 8 The RankEvo visualization of fortune-100 companies. The itemsets represented by the gray bands and the associated line chart indicate a rising of IT companies since 1983. The representative company "Motorola" is marked with a *green curve*. Also, there reflects a significant evolution change in 1995 in two ways: the sudden dropping of ranked list similarity at the bottom and the many isolated items on the list of 1995. This evolution reveals the addition of service-oriented companies such as "Walmart"

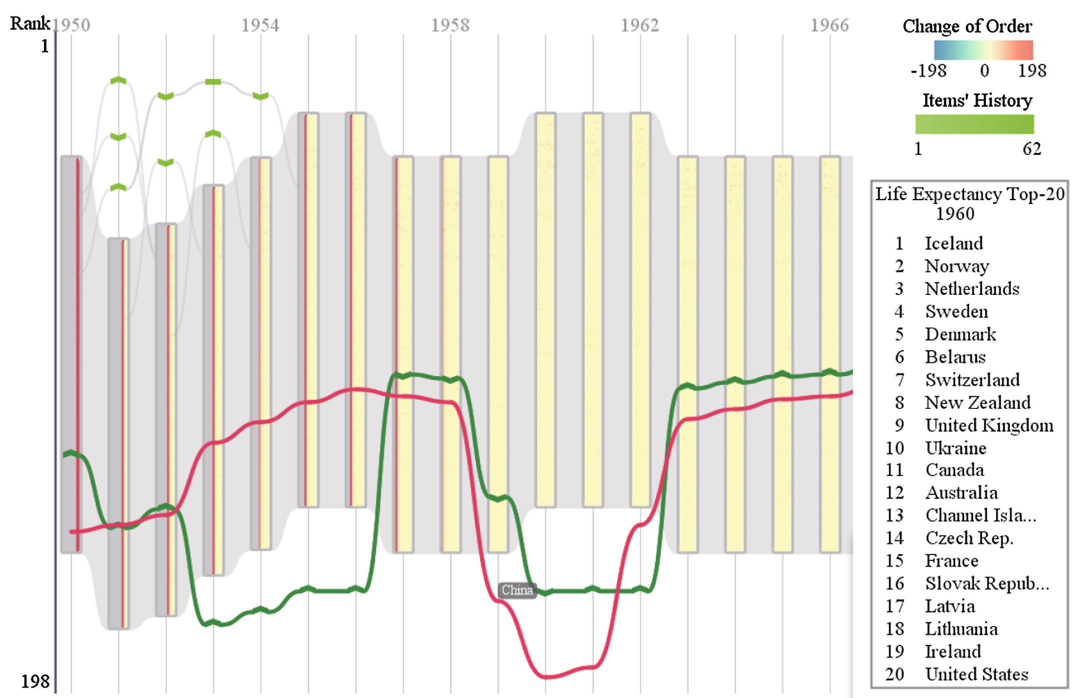

Fig. 9 The RankEvo visualization of time-varying life expectancy at birth. The life expectancy decline of China reveals the success of Chinese socialist transformation between the years 1953 and 1958, and the following 3 years of great Chinese famine between the years 1958 and 1961

plan in 1953 and carried out the socialist industrialization and the socialist transformation successfully. However, due to natural disasters and government policy errors, China experienced the 3 years of great Chinese famine between the years 1958 and 1961.

7 Discussion

7.1 User evaluation

We conducted a user evaluation study with 5 participants for the effectiveness of our system in exploring ranking evolution. All participants expressed their interests in discovering the correlations and trends of the itemsets with the RankEvo system. They indicated that our methods are able to separate different evolution features and visualize temporal patterns intuitively. They are happy that unfamiliar items in the Wikipedia page can be explained by the item page. The participants also pointed out that they were confused why sometimes the rising trend of items and itemset are encoded in a decline trend in the main view, such as the IT companies case. The red ranking curve and the extra curve view only mitigate the problem to a limited extent. Also, they found some items in the itemset share a common ranking trend but they're not much related. This is due to the weak assumption that similar ranking trend indicates semantic correlation of items. In the future work, we plan to introduce complementary semantic support to the similarity measurement process. For the Wikipedia case, complementary semantics such as the page-link between items can greatly enhance the correlations among itemsets. However, such semantic support is case-dependent and cannot be applied to other applications.

7.2 Parameter discussion

According to different features of various datasets, different ratios of the three parameters: the dynamic time warping factor, the compensation factor, the average rank factor, are applied to explore latent rank evolutions. We tend to use larger compensation factor for datasets with more short-term ranked items. For example, the compensator factor in Wikipedia Top-50 dataset is large while that in life expectancy dataset is set to 0 because the ranks are complete all the time. Also, because most ranks are steady in the life expectancy dataset, a higher average rank factor is required to further separate the items. However the short-term popular ranked items, due to the lack of similarities with historic data, are often hard to identify and require user interactions to analyze.

Although the parameters are crucial for the latent ranking evolutions, users are not allowed to adjust them interactively. Because each adjustment results to overall data reconfiguration and layout reconstruction once the parameters are changed, the overall ranking patterns changes so disorderedly that users cannot follow the visualization smoothly.

7.3 Limitations

According to users' feedback, our system still has some limitations. First, the visualization reveals the concurrent ranking pattern rather than the absolute rank orders. An emergence or expansion of a cluster sometimes yields the layout offset of other clusters and other items, which often confuses users when items' positions do not indicate their rank orders. Second, because the common ranking trend is automatically computed based on the rank time series, the semantic correlation is not covered. It still depends on users' personal knowledge or external query to supplement the semantic correlation among items in one cluster. Thus, we would enhance the layout of items and itemsets and semantic support of our RankEvo system in the future work.

8 Conclusions

Studying the ranking evolutions has attracted much attention in the data analysis and visualization communities. This paper presents a novel scheme for discovering latent ranking evolutions by integrating ranking structuring algorithms with visualization techniques. The ranking structuring process includes dissimilarity computation of the ranked items and tree wave structure generation of the concurrent itemsets.

Our RankEvo system employs three coordinated visualizations to analyze ranking time series in different aspects and provide a set of interactive analysis interactions. The three case studies verify its efficiency in detecting both short-term and long-term evolution patterns of ranked items, as well as patterns of isolated items.

Concerning the future work, we plan to enrich our visualization with level-of-detail clustering of ranked items based on custom related parameters. Such clustering approach should be integrated into the system with additional flexible user interactions. Also, we will leverage other analysis approaches for item properties to enhance the analysis efficiency.

Acknowledgments This research was supported by National Natural Science Foundation of China (61202279) and Zhejiang Provincial Natural Science Foundation of China under Grant No. LQ12F02003.

References

Aigner W, Miksch S, Schumann H, Tominski C (2011) Visualization of Time-Oriented Data. Springer
Alvo M, Cabilio P (1985) Rank correlation methods for missing data. Can J Stat 23(4):345–358
Batty M (2006) Rank clocks. Nature 444:592–597
Cao N, Lin YR, Sun X, Lazer D, Liu S, Qu H (2012) Whisper: Tracing the spatiotemporal process of information diffusion in real time. IEEE Trans Vis Comput Graph 18(12):2649–2658
Critchlow DE (1985) Metric Methods for Analyzing Partially Ranked Data, vol 34. Springer
Cui W, Liu S, Tan L, Shi C, Song Y, Gao Z, Qu H, Tong X (2011) Textflow: Towards better understanding of evolving topics in text. IEEE Trans Vis Comput Graph 17(12):2412–2421
Hochheiser H, Shneiderman B (2004) Dynamic query tools for time series data sets: timebox widgets for interactive exploration. Inf Vis 3(1):1–18
Jaccard P (1901) Etude comparative de la distribution florale dans une portion des Alpes et du Jura. Impr, Corbaz
Kidwell P, Lebanon G, Cleveland W (2008) Visualizing incomplete and partially ranked data. IEEE Tran Vis Comput Graph 14(6):1356–1363
Lu M, Wang Z, Yuan X (2015) Trajrank: Exploring travel behaviour on a route by trajectory ranking. In: Proceedings of the IEEE Pacific Visualization Symposium, IEEE
Marden JI (1995) Analyzing and modeling rank data. Chapman&Hall, London
McLachlan P, Munzner T, Koutsofios E, North S (2008) Liverac: interactive visual exploration of system management time-series data. In: Proceedings of the 26th SIGCHI conference on Human factors in computing systems, ACM, pp 1483–1492
Ogawa M, Ma KL (2010) Software evolution storylines. In: Proceedings of the 5th international symposium on Software visualization, ACM, pp 35–42
Plaisant C, Milash B, Rose A, Widoff S, Shneiderman B (1996) Lifelines: visualizing personal histories. In: Proceedings of the SIGCHI conference on Human factors in computing systems, ACM, pp 221–227
Shi C, Cui W, Liu S, Xu P, Chen W, Qu H (2012) Rankexplorer: Visualization of ranking changes in large time series data. IEEE Trans Vis Comput Graph 2669–2678
Shneiderman B (1996) The eyes have it: a task by data type taxonomy for information visualizations. In: Proceedings of the IEEE Symposium on Visual Languages, IEEE, pp 336–343
Sun M, Lebanon G, Collins-Thompson K (2010) Visualizing differences in web search algorithms using the expected weighted hoeffding distance. In: Proceedings of the 19th international conference on World Wide Web. NY, pp 931–940
Tominski C, Abello J, Schumann H (2004) Axes-based visualizations with radial layouts. In: Proceedings of the 2004 ACM symposium on Applied computing, ACM, pp 1242–1247
Tufte ER (1986) The visual display of quantitative information. Graphics Press, Cheshire
Vintsyuk TK (1968) Speech discrimination by dynamic programming. Cybern Syst Anal 4(1):52–57
Ware C (2012) Information visualization: perception for design, 3rd edn. Elsevier
Weber M, Alexa M, Müller W (2001) Visualizing time-series on spirals. In: Proceedings of the IEEE Symposium on Information Visualization, p 7
Wei J, Shen Z, Sundaresanl N, Ma KL (2012) Visual cluster exploration of web clickstream data. In: Proceedings of the IEEE Conference on Visual Analytics Science and Technology, IEEE, pp 3–12
Wijk JJV, Selow ERV (1999) Cluster and calendar based visualization of time series data. In: Proceedings of the IEEE Symposium on Information Visualization, IEEE, pp 4–9