

Visualizing Rank Time Series of Wikipedia Top Viewed Pages

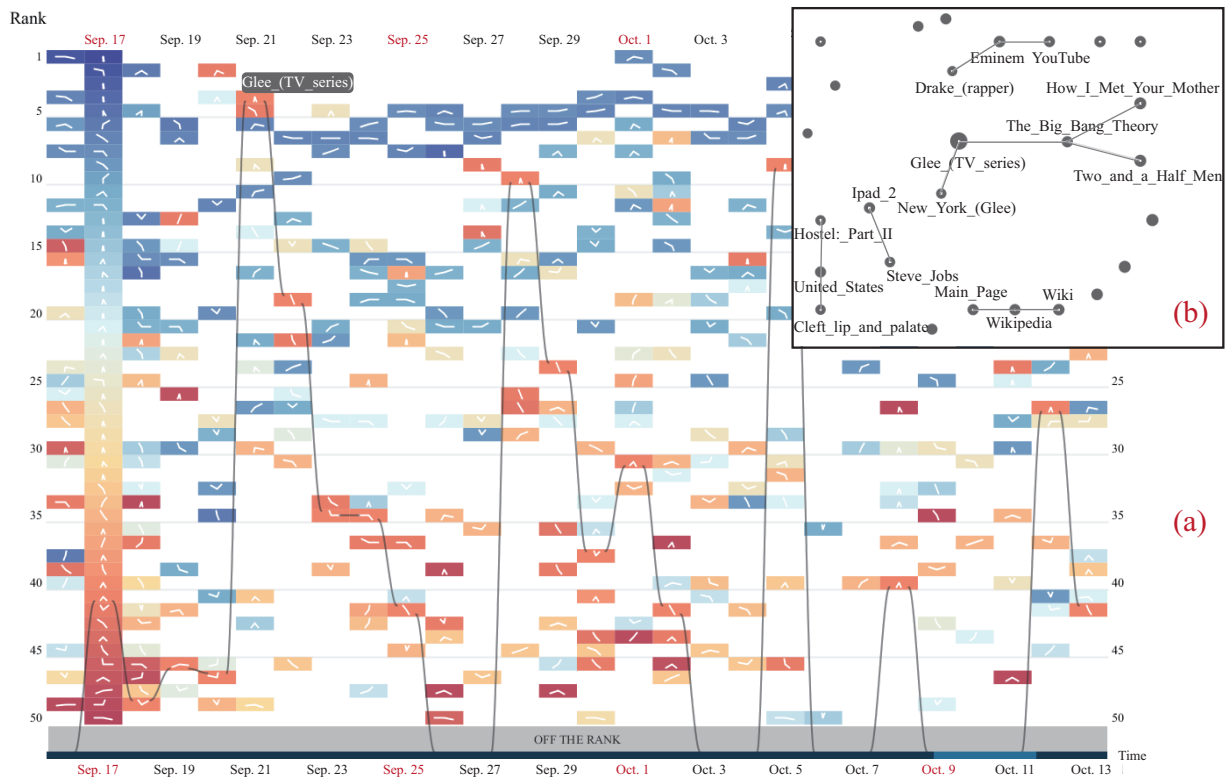


Figure 1: The evolving ranking trends of weekly TV series on Wikipedia. (a) A blue to red color scheme is applied to each page from top to bottom on the selected day (September 17th). The same pages on other days carry the same colors with those on this day while pages not included on this days are cast away. The rank series of page “Glee” (an American TV series) goes up every Wednesday, indicating the day when Glee is on TV, and goes down afterward. (b) By exploring the page’s similar pages, we get other pages for Glee, and other weekly TV series such as “How I Met Your Mother”, “The Big Bang Theory” and “Two and a Half Men” in the page link network.

ABSTRACT

Visual clutter is a common challenge in visualizing large rank time series data. Following the Gestalt’s law of continuity [11], we try out a variety of visual design approaches on large rank time series datasets. We use Wikipedia top page view statistics to test and evaluate these approaches. The data is a set of top viewed pages over time, which is of great importance in analyzing viewers’ interest in current affairs. Based on the visual designs we implement **Wiki-TopReader**, a reader of Wikipedia page rank, with which users are able to explore connections among those top viewed pages by connecting the page rank behaviors with the page link relations. Such a combination enhances the unweighted Wikipedia page link network and brings users’ page of interest to a broader attention. The evaluation shows that the system is effective on representing evolving ranking patterns and page-wise correlation. The design is intuitive and visually appealing.

Keywords: Rank time series, page view, page link, visualization.

1 INTRODUCTION

Large rank time series datasets have always posed a challenge to data analysts. In addition to the time-varying property, each of the items at every time point can be a complex object that has multi-variate properties, relation properties or both. In the scope of this paper, we study rank time series to explore their evolving patterns and the relations among the top items.

Wikipedia is considered to be the biggest online encyclopedia, whose everyday page view throughput can be more than 600M across languages¹. It has become a knowledge gathering platform where users learn and contribute their knowledge. Due to the sustained large scale knowledge accumulation, Wikipedia has also become a huge and growing knowledge warehouse.

It would take more than a human lifetime to digest all the knowledge collected in Wikipedia. But if we just want to understand current news and events, we could just read through the top viewed pages. According to the 80/20 principle, the ranking data of Wikipedia’s top page view statistics (**Wikipedia page rank** for short) reflects users’ major interests in Wikipedia and furthermore in ongoing affairs indicated by the top queries. Therefore, the time series of Wikipedia page rank (**Wikipedia page ranking trends**)

¹Wikipedia article traffic. <http://stats.grok.se>

indicate how users' interests evolve over time.

An enormous amount of work has focused on understanding Wikipedia contents, such as natural language processing (NLP) and resource description framework (RDF) database. However, in terms of analyzing Wikipedia contents, data mining strategies may not satisfy users' various needs for investigating the long term Wikipedia page ranking trends. On the other hand, existing visualization solutions can satisfy different user's various need, but usually fall into two categories with severe limitations: scattered solutions and continuous solutions. Scattered solutions, such as a scatter plot of rank orders is too simple to reveal rank trends. It requires many user interactions to explore the entire dataset. Continuous solutions, such as band or river-like designs can reduce interactions, but they suffer heavily from visual clutter that results from overwhelming crossings caused by item series.

To address these challenges, we design the WikiTopReader, a visual analytical system that not only visualizes Wikipedia page ranking trends, but also constructs a semantic network for every given Wikipedia page of interest. We construct a semantic network that associates Wikipedia pages of similar ranking trends to form an ongoing affair. While designing the visualization, we keep a simple principle: allow the user to see the trends of Wikipedia pages without causing unnecessary perceptual complexity. Our visualization design avoids visual clutter by breaking the band or river-like visualization into scattered glyphs while keeping users' perceptual continuity towards item series.

Although only demonstrated with a Wikipedia dataset, our scattered glyph design can be directly applied to other rank time series such as stock price datasets. For rank items with semantic relations, the page link relation in the Wikipedia case can also be easily adjusted to other relations to construct a user-aware network. We summarize our contributions as follows:

- Three glyph designs that portray rank time series data with perceptual continuity.
- A mashup of rank time series and semantic relations that enhance users' understanding of rank items.
- A Wikipedia page rank application that demonstrates both the glyph designs and the mashup framework.

The rest of this paper is organized as follows. Related work is reviewed in Section 2. Section 3 elaborates all the Wikipedia page rank representations. Section 4 explains how users connect the page ranking trends with the page link relations. Case studies are elaborated in Section 5, followed by user studies and discussions in Section 6. Finally, we conclude the paper in Section 7.

2 RELATED WORK

The visualization of rank time series falls into three categories: visualization without explicit representation, curve representation and glyph representation.

2.1 Visualization Without Explicit Representation

Low-dimensional embedding for visualizing rank time series is to embed each individual rank in a low-dimensional space. The ranks then are encoded with a heatmap [8], dots [7] and clickstream sequences [5]. Even without proper layout, selective interactions can also significantly improve users' efficiency in pattern exploration. With the system, users are able to rank the data based on different scales and semantics. Vuillemot and Perin [9] proposed a navigating technique that allows line charts overlaying on the top of the rank time series table by direct manipulation such as dragging. However, low-dimensional embedding causes great information loss due to the embedding. Our designs are all glyph-based representations that maintain items' rank orders.

2.2 Curve Representation of Rank Time Series

Typical curve representations of rank time series are lines and bands (of the same width or river-like). RankClock [6] positioned rank trends as spirals on a radial coordinates for better periodical pattern revealing. The "table-graphic" [4] by Edward Tufte, or now known as slope graph, connects time-varying items with lines. Sorting items by rank, it can also present rank time series. The work by Chris Harrison² can be considered an application of the slope graph on Wikipedia hot queries. Visualizing rank times series with band³ is another option but failed to present a clear display. Existing representation of rank items as lines, bands or rivers can often cause heavy visual clutter when it comes to a large number of items. On the other hand, our proposal breaks the curves into glyphs to avoid crossings while keeps the perceptual continuity.

2.3 Glyph Representation of Rank Time Series

The simplest glyph representation is a scatter plots of rank time series enhanced with user selections⁴. RankExplorer [3] segmented rank time series by rank and encoded each segment with glyphs to reveal inflow and outflow changes between segments. Although this method successfully depicts the rank changes within and across multiple categories, it is hardly capable of disclosing pattern of individual items, let alone complicated correlations of rank changes. There are two techniques not specially designed for rank times series but they also share similar design with ours. LineUp [10] is an outstanding multi-attribute ranking exploration system that integrates visual forms such as slope graphs and bar charts with flexible interactions including filtering and reordering. It may suffer from edge crossing when the ranking size is large because the rank items are exactly connected with slope graph. Parallel tag clouds [2] combines tag clouds with parallel coordinates to visualizes document collections. The same words in documents are explicitly linked as in slope graphs. Our proposal has much in common with Parallel tag clouds in terms of breaking band or river-like visualization while keeping users' perceptual continuity. However, our visualization design concentrates more with ranking trends rather than the item itself and the semantics part is represented additionally. The color schemes between the two designs are also different.

3 RANK TIME SERIES REPRESENTATIONS

The general goal of our visualization is to describe the evolving local ranking trend of rank items. Based on the principle that a good description should not cause unnecessary perceptual complexity, we discard the obscure low-dimensional embedding and the overly cluttered curve representation. Our proposed alternative solution is a glyph representation that can reveal evolving patterns of individual items, as well as the correlations of ranking trends. For visualizing rank time series, we follow the Gestalt law of continuity and make three representation proposals: the well-adopted sparkline design (Figure 2), the badge design (Figure 3) and the badge with patch design (Figure 4).

The three representations all share a general timeline layout. The horizontal axis is the time axis with a fixed time window (28 time steps for 4 weeks) and a drag-able time slider. Each time step is a column, with all items placed in the column according to their rank order. The exact rank curve of the item as well as its label will be highlighted when users hover over. Labels are not shown by default but triggered with user interactions for two reasons. First, for the case of Wikipedia as well as other similar cases, labels are not just

²WikiTop-50 visualization. <http://www.chrisharrison.net/index.php/Visualizations/WikiTop50>

³Fortune-500 visualization. <http://in.somniac.me/2010/01/fortune-500-visualization/>

⁴Fortune-500 visualization. <http://fathom.info/fortune500/>

one word but can also be a long phrase (as shown in Figure 1), which causes clutter because of its length. Second, positioning of labels can further cause visual clutter and hinder users' performance in tracing the trends.

When users focus on one day, a blue to red color scheme is applied to each item from top to bottom in the corresponding column. The same items on other days carry the same colors as that of the selected day, while items not included on that day are cast away. We apply a diverging color scheme to highlight the hot queries and disappearing queries. The filtering function helps users concentrate on items of interest because it assists users in locating a certain item by both the glyph and the color.

3.1 The Sparkline Visualization

The simplest solution of time series would be sparklines of local ranking trends of items (see Figure 2), which is widely used in stock price displays or in click rate displays. The sparkline visualization depicts the rank orders of the item 5 days before and 5 days after the day of the column. In this way, users are able to recognize the same item in the sparkline design by looking for glyphs of the same color and similar sparkline shape. In addition, general ranking trends such as steady, increasing and decreasing trends are easy to identify with sparklines.

However, the detailed sparkline shape may also cause visual distraction towards detailed information. Subtle changes, such as ranking changes smaller than 10, are not easy to distinguish. Users may find it difficult to distinguish the ranking orders precisely with one sparkline alone, and to recognize the same item in different columns. Therefore, we feel it reduces pre-attentive pattern recognition and is not very aesthetically pleasing.

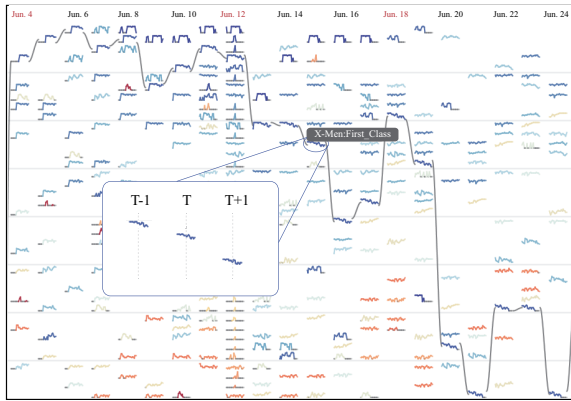


Figure 2: The sparkline representation of Wikipedia page rank time series. A blue to red color scheme is applied to each item from top to bottom on the selected day (June 12th). The same items on other days carry the same colors with those on this day while items not included on this days are cast away. Users may check a single glyph to see its local evolving trend (e.g., “X-Men: First Class”).

3.2 The Badge Visualization

Usually lines or arrows that connect to a page's next rank orders can lead users attention and provide a continuous perceptual experience. Guided by the Gestalt law of continuity, we get rid of the lines or streams that cause visual clutter but retain the item rank glyph. Here we call the glyph a **badge**. A badge is a simple glyph representation of the item's rank with two edges (see Figure 3): one pointing to its last rank position and the other pointing to its next rank position. Its direction and color are carefully computed to enhance the recognition of the items.

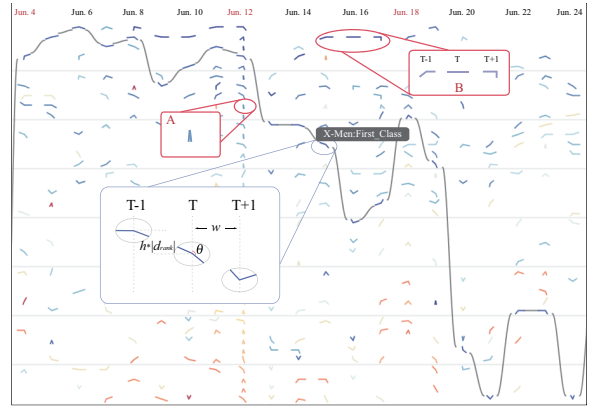


Figure 3: The badge representation of Wikipedia page rank time series. A blue to red color scheme is applied to each item from top to bottom on the selected day (June 12th). The same items on other days carry the same colors with those on this day while items not included on this days are cast away. By following the color and the directions of the two edges, users can see its local evolving trend (e.g., “X-Men: First Class”). Meanwhile, glyphs of page A and page B show two different rank evolving patterns.

To maintain a consistent shape for the badge, we constrain the area for a badge in an ellipse with the same width and height. The rank difference and the time interval make an angle θ .

$$\theta = \begin{cases} \pi - \arctan(h * |d_{rank}|/w) & \text{left} \\ -\arctan(h * |d_{rank}|/w) & \text{right} \end{cases}$$

where d_{rank} is the rank difference, w and h are the width and height of the display region, respectively. Thus, a badge can be represented with two line segments from the center to the edge of the ellipse, respectively. The coordinates of the two points on the edge can be represented with polar coordinates as follows:

$$\begin{cases} x = w'/2 * \cos \theta, \\ y = -h'/2 * \sin \theta \end{cases}$$

where w' and h' are the width and height of the ellipse and are usually a little smaller than the display area of the items.

The shape of a badge (the angle size) naturally depicts the general ranking trends of the items. As shown in Figure 3, Item A is more likely to be a new top page, flashing in and flashing out of the top page rank list. In contrast, Item B is a popular page with steady evolving pattern during that period. Following its trend, Item B disappears from the rank after June 17th.

3.3 The Badge with Patch Visualization

The glyphs in the badge design are small to accommodate many items. Thus, it becomes hard to distinguish the neighboring colors. To further highlight the distinct colors, we make a modification of the badge visualization. Instead of coloring the glyph, we embed a color-filled patch to the glyph (see Figure 4) that we have named a **badge with patch**, to reveal how items on the selected day survive in other days. In both the badge and badge with patch design, the empty slots tell two things: 1. the current topics are disappearing, and 2. new items that are evolving should fill in these slots. Compared with the badge design, we think the filled patch can be a great improvement in visual recognition of the items.

4 SEMANTIC EXPLORATION

We explore page-wise semantics with two integrated relations: the **page link relation (PLR)** and the **similarity relation of rank time**

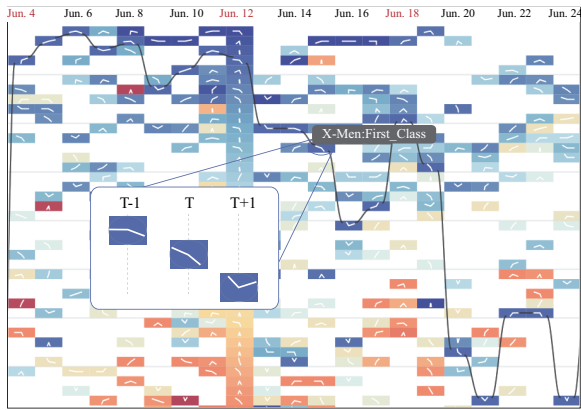


Figure 4: The badge with patch representation of Wikipedia page rank time series. A blue to red color scheme is applied to each item from top to bottom on the selected day (June 12th). The same items on other days carry the same colors with those on this day while items not included on this days are cast away. Users can notice how pages on June 12th remained on the rank list in other days.

series (SRR). PLR denotes a semantic relation explicitly, while SRR is more implicit. For an ongoing topic, users tend to query several key words that are related to the affair, resulting in concurrent rank time series. Under this assumption, pages with similar rank series are likely but not necessarily to be related. The two relations can be combined to enhance the Wikipedia page link network with user behaviors and capture users' limited attention to a condensed network.

4.1 Finding Pages with Similar Trends

We evaluate the dissimilarity between two pages with the proximity of their associated rank time series, under the assumption that two Wikipedia pages are potentially but not necessarily correlated if they share similar trends during their recent history. We first define the dissimilarity of two ranked pages (A and B) at a time point as the dissimilarity of the rank time series of their consecutive days. Then we apply a curve matching technique [1] that employs the dynamic time warping to compute the dissimilarity of the rank time series. We do not employ the Euclidean distance because it is not suitable for cases where two time series have a similar trend but are offset from each other.

We also adopt the entropy-based evaluation of cluster quality described in Chen et al.'s paper [12] in finding the most similar pages. The similarity-aware entropy score calculates the entropy of a bunch of time series based on their pairwise similarities. We then add pages with the most similar rank series to the page cluster, and compute the entropy score of the growing cluster iteratively until the score exceeds a certain threshold. This is superior to the k -nearest neighbors because it gets all pages whose rank series are similar with that of this page. We plot the local rank trend of these pages in a curve view so that users can compare how and when the pages have similar patterns. Figure 5 (a) shows the similar patterns such that the pages stayed on top for several days, then went down rapidly and disappeared from the Top-50 rank list.

4.2 Exploring Semantics

Because similar pages do not contain semantic correlations among pages, we integrate a page link network to additionally represent relationships between pages. For all found similar pages, we construct a network based on PLR. The page of interest is placed in the middle while pages linked to it are connected via edges. To

further draw users' attention, pages not linked to other pages are represented as scattered points and randomly placed.

Figure 1 (b) shows pages whose rank trends are similar with that of the page "Glee (TV series)". In the page link network, there are other pages of Glee ("New York (Glee)"), and other weekly TV series such as "How I Met Your Mother", "The Big Bang Theory" and "Two and a Half Men". The node size indicates the degree of similarity as defined in Section 4.1.

4.3 Implementation

The page view statistics for Wikimedia projects⁵ maintains raw page access records for all Wikipedia projects across all languages. We have collected 14-month English page view statistics dataset (from June to October in 2011 and from January to September in 2014) and generated daily Top-1000 page views in a MySQL data archive. To focus on ongoing affairs, we have removed the index page, the portal pages, the error pages and other specific pages from our statistical results. The visualization only shows top-50 page views and pages with similar rank series are fetched among the Top-1000 pages. We also have collected the page link dataset via Wikipedia page link APIs⁶ and maintained data persistence via Neo4j⁷, a graph database.

5 CASE STUDIES

The WikiTopReader system can be used to identify the evolving patterns of ranking trends, as well as the underlying events composed by similar pages. In this section, we describe two cases that show the usefulness of WikiTopReader on spotting patterns of concurrent events and different evolution patterns of one topic. Typical user exploration styles are time-based, item-based and glyph-based.

5.1 Pattern of Concurrent Events

Pages with similar ranking trends are not necessarily related, which may mislead users when they find two pages of similar trends unrelated. The page link network avoids such misunderstanding. The first case shows that WikiTopReader can be used to recognize two distinct concurrent events with similar ranking trends.

As shown in Figure 5, on July 23rd, 2011, four pages: "Amy", "27 Club", "Anders" and "2011 Norway Attack" appeared on the top of the rank simultaneously. They stayed on top for several days, then went down rapidly and disappear from the top-50 rank list. We click one page "27 Club" for further exploration.

Figure 5 (a) is the rank curve view of pages sharing similar ranking pattern with "27 Club", while Figure 5 (b) shows the network constructed by the pages with both similar ranking trends and page link relations. In the network, the node size represents the similarity shared with the central node, and the edge between two nodes depicts the existence of page link relation between the pages.

We find that most similar pages share a similar changing pattern. They suddenly appeared on a specific day, stayed for several days and then went down. When we study the page link network view (Figure 5 (b)) for more information, we get two networks of pages related to the four pages. N1 and N2 are obviously independent events, but they get high attention simultaneously. N1 is about "Amy" and "27 Club", and N2 is about "Norway" and "Anders".

Integrating the Wikipedia contents of the pages, we realize that "27 Club" is a term that refers to the group of rock/popular musicians who die at age 27. And the pages such as "Jim Morrison" and "Brian Jones" are names of those who belong to the 27 Club. On July 23rd 2011, "Amy", a famous British rock musician, died at her age of 27. Thus, the related pages constructed a network centered

⁵Page view statistics for Wikimedia projects. <http://dumps.wikimedia.org/other/pagecounts-raw/>

⁶MediaWiki API. <http://en.wikipedia.org/w/api.php>

⁷Neo4J, a graph database. <http://neo4j.com>

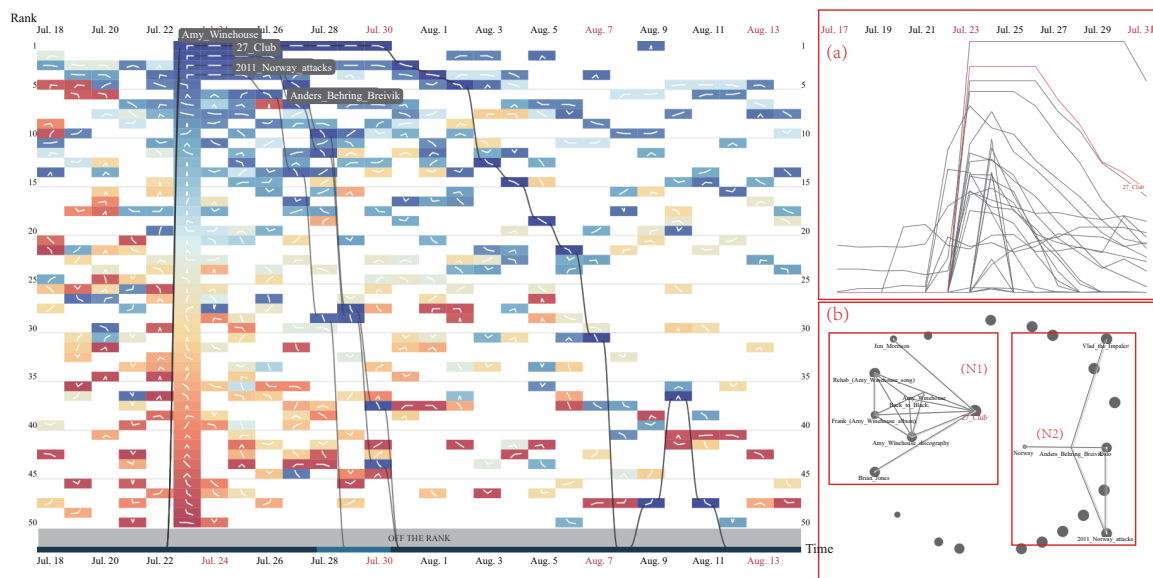


Figure 5: The evolving ranking trends of page “27 Club” and other pages of similar ranking pattern. (a) is the rank curve view of pages sharing similar ranking pattern with “27 Club”. (b) depicts the network constructed by the pages with both similar ranking trends and page link relations. It shows two networks, of which one indicates members of the “27 Club”, and the other denotes the “2011 Norway Attack”.

with “27 Club”. In terms of N2, on July 23rd 2011, a terrorist attack conducted by “Anders” happened in “Oslo”, “Norway”, which was known as the “2011 Norway Attack”.

5.2 Evolution Pattern of One Topic

The second case shows how WikiTopReader traces the evolution pattern of a specific topic: the missing flight Malaysia Airlines Flight 370 (MH370).

On March 8th, 2014, a new page, “Malaysia Airlines Flight 370” (MH370), suddenly appeared to the top of the page view rank of Wikipedia (see Figure 6 (a)). We trace the page named “Malaysia Airlines Flight 370” for its evolution pattern. Its first appearance began on March 8th, 2014, lasting for 31 days and ended on April 8th, 2014. Then, on July 16th, 2014, it reappeared on the Top-50 rank and disappeared 6 days later (see Figure 6 (b)).

To explore the different patterns of its two appearances, we study the page “Malaysia Airlines Flight 370” on March 8th and July 16th, respectively. Figure 6 (a) shows the page-view frequency curve (top right) and the page link network of pages sharing similar trends with MH370 in March (bottom right). Figure 6 (b) shows the results of similar pages of MH370 in July.

As shown in the bottom right of Figure 6 (a), the similar pages of MH370 are mainly about locations (e.g. “South China Sea”, “Malaysia”) and aviation (“Boeing 777”, “Malaysia Airlines”, “Asiana Airlines Flight 214”, etc.). The pages showed up simultaneously and stayed top for weeks, sharing a similar evolution pattern as MH370 (see top right of Figure 6 (a)).

Wikipedia describes the event of missing flight MH370 in March 2014. MH370 was a scheduled international passenger flight that disappeared on March 8th, 2014, while flying from “Kuala Lumpur International Airport”, “Malaysia” to “Beijing Capital International Airport”, “People’s Republic of China”. The aircraft, a “Boeing 777-200ER”, was carrying 12 Malaysian crew members and 227 passengers from 15 nations.

Thus, we could draw the conclusion that the similar pages were related based on the context of the vanished flight. This event was mysterious and catastrophic and gained a long lasting high attention and triggered a wide discussion. It resulted in a shared pattern of a steady top trend (Figure 6 (b) (B)) and generated a complicated

semantic page link network centered with “Malaysia Airlines Flight 370” (Figure 6 (b) (C)).

In contrast with the long lasting popularity of MH370 in March, the second appearance of MH370 in July (represented by Figure 6 (b)) is just a flash. Additionally, the semantic page link network is quite simple (see Figure 6 (b) (C)). The event behind the pages is another accident with the same flight. Another scheduled international passenger flight of Malaysia Airlines from Amsterdam to Kuala Lumpur crashed on July 17th, 2014. The page link network indicates some relevance with this event: “Iran Air Flight 370” (another flight shot down in wars) and “September 11 Attacks” (Figure 6 (b) (C)).

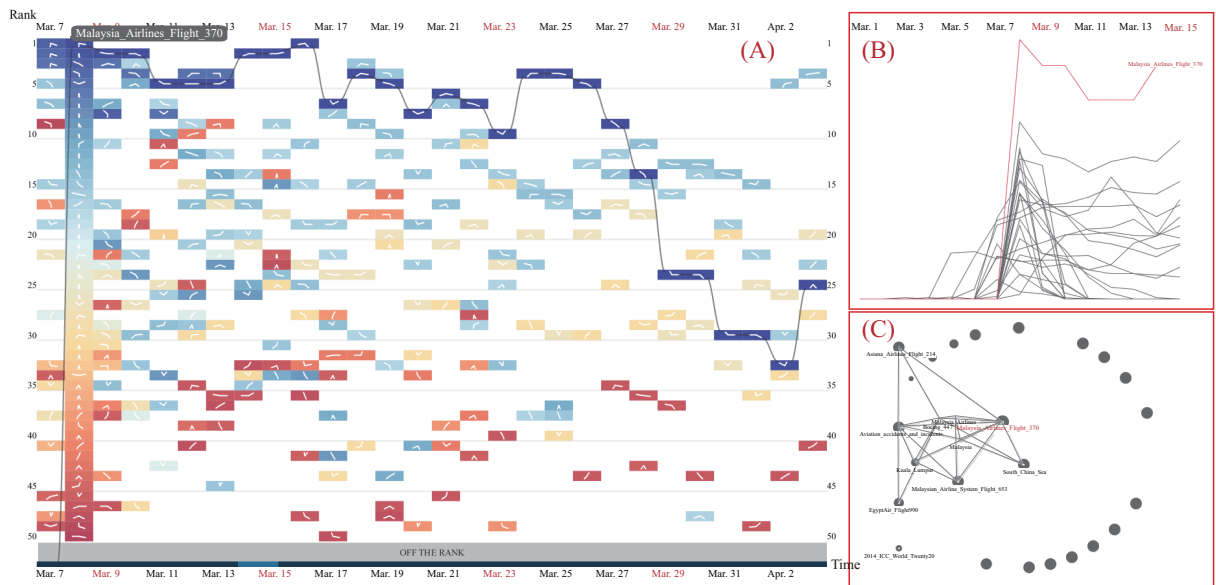
Based on the distinct patterns of the same page, we generate two hypotheses. First, although centered on the same topic (MH370), two events have different backgrounds, which lead to the distinct evolution patterns. Second, according to the agenda-setting theory, it is also possible that media gives different perspectives towards the two events. Nevertheless, the final conclusion can only be drawn when more controlled experiments are conducted with enriched contextual data.

6 EVALUATION

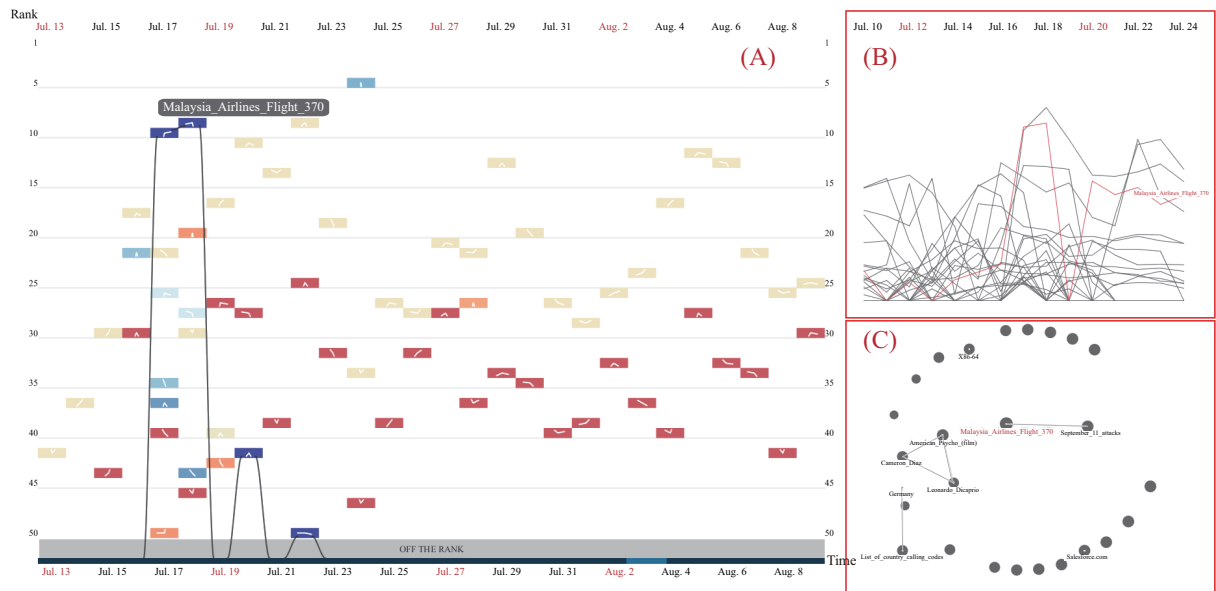
To evaluate the visual designs as well as the usability of the system, we carried out two experiments. The first aims at the effectiveness and efficiency, including speed and accuracy, on locating rank items. The second evaluates the user experience about the WikiTopReader system, especially on how they find pages of similar trends. Note that our targeted user are not domain experts but rather regular Wikipedia users.

6.1 Experiment 1: Evaluation on Visual Designs

Experiment 1 is designed to compare the accuracy and speed of three visual designs in identifying different rank patterns: locating items at successive time steps, determining the life span of the items, locating greatly changing items and depicting ranking trends. We are not certain of their specialties in various tasks. But as being described in Section 3.1, our hypothesis is that, the sparkline has poorer performance (in accuracy as well as in time of completion) than the other two in depicting ranking orders.



(a) The evolving ranking trends of page “Malaysia Airlines Flight 370” on March 8th, 2014, and other pages of similar ranking pattern. Among the similar pages, there are important locations such as “Kuala Lumpur International Airport”, “Malaysia”, “Beijing Capital International Airport” and “People’s Republic of China”; and other aircraft accident evolving “Asiana Airlines Flight 214” and “Air France Flight 447”.



(b) The evolving ranking trends of page “Malaysia Airlines Flight 370” on July 17th, 2014, and other pages of similar ranking pattern. Compared with that of the same page on March 8th, 2014, the 17th July page has much simpler pattern and page link network.

Figure 6: Page “Malaysia Airlines Flight 370” has two different evolving patterns on the Wikipedia page rank list. The evolving ranking trends of page “Malaysia Airlines Flight 370” on (a) March 8th, 2014, and on (b) July 17th, 2014.

20 subjects aged between 18 and 40 participated in the first experiment online. Among them, 8 were female and 12 were male. The first experiment is composed of tasks that seek to evaluate the visual designs with only static images. After a brief description of 3 designs, our participants were asked to finish 4 tasks. Before each task, our participants first learned the task with a representative glyph (a gray ellipse). For each task, the participants had to complete 3 exercises with 3 visual designs respectively in a random order and from different data subsets. The tasks are described as follows:

- T1** Choose the glyph on the previous/coming day of the selected item.
- T2** Locate which day contains the most/least items of the selected day.
- T3** Choose greatly changing items in the current time scope.
- T4** Given a selected item, describe its possible rank series during the 12 time steps on the time axis below. For this task, our participants were asked to click on 12 time axes to plot the

rank orders to form the rank series. A distance score between time series and ones drawn by users is calculated to evaluate their performance.

After the exercises of each task, the participants were asked to order their preferences towards 3 visual designs for the specific task. They should order the designs from the most helpful to the least (e.g., *sparkline* > *badge* > *badge with patch*).

6.1.1 Accuracy and time of completion of tasks in Experiment 1

Figure 7 shows the accuracy of T1, T2 and T3. Table 1 indicates the accuracy of T4. Figure 8 shows the completion time in Experiment 1 for 4 tasks, with outliers plotted explicitly.

For accuracy, 3 designs do not yield much difference in T1, T2 and T3. In terms of distinguishing the possible rank series of an item (T4), Table 1 shows the standard deviation between the actual series and that plotted by participants. The average standard deviations are 3.68, 6.89 and 3.63 for the badge, the sparkline and the badge with patch, respectively. Both the badge and the badge with patch worked effectively and significantly outperformed the sparkline (with $p=0.0110$ and $p=0.0032$, respectively). It verifies the limitation of the sparkline in depicting ranking trends. Although users can follow a very rough trend with the sparkline, the accurate directions pointed by 2 badge designs contribute better accuracy.

As for the average completion time, the badge contributes the shortest completion time in T1 while the sparkline contributes the longest (with $p=0.0098$). There is no significant difference in T2 or T3. Although the sparkline is significantly faster than the badge with patch in T4, the accuracy of the sparkline in T4 is the worst according to Table 1.

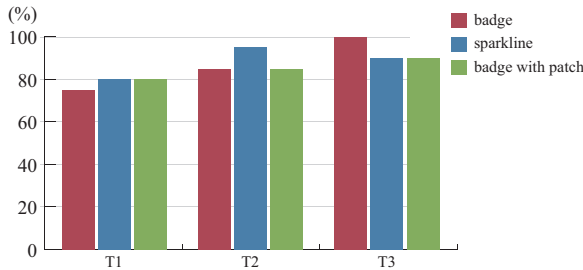


Figure 7: The average accuracy of all participants in T1, T2, T3.

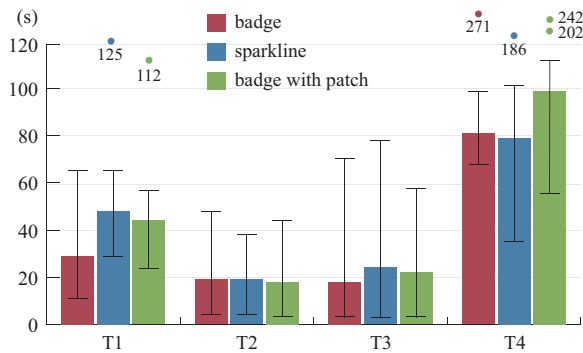


Figure 8: The average completion time of all participants in T1, T2, T3 and T4, with outliers plotted explicitly.

6.1.2 User preference towards designs

We also compare the user preference towards 3 designs on 4 tasks. The most favorable design is scored 3, the medium one scored 2,

and the least one scored 1. Results of the user preference on each task towards 3 designs are illustrated in Figure 9.

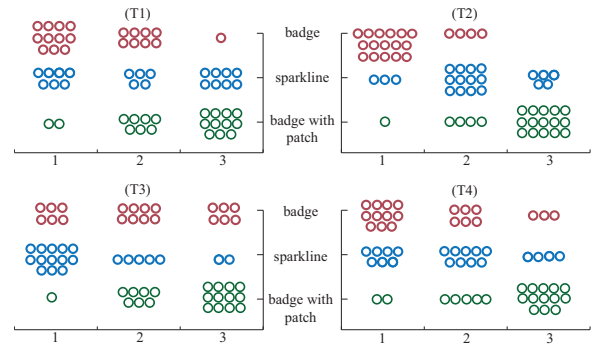


Figure 9: User preference scores (20 participants) towards the 3 designs on the 4 tasks: T1, T2, T3 and T4.

As clearly shown in Figure 9, our participants prefer different designs with different tasks, but the badge with patch generally receives more preference. It is not so significant in T1 but significant in T2, T3 and T4 (with $p = 0.0149$, $p=0.00006$ and $p=0.0094$, respectively) in a way that the badge with patch is better than the sparkline. The badge is the least favorable in determining item life span (T2) (with $p=0.0002$), not so welcome in locating successive item (T1) (with $p = 0.0428$), and slightly better than the sparkline in locating greatly changing items (T3). The badge and the sparkline win similar votes in depicting rank trends (T4).

It is interesting that even though the badge with patch does not help much in assisting analyzing tasks according to the accuracy results, it turns out to be the most popular one. Hence, we interviewed some of the participants for reasons:

“The badge with patch impresses me. It seems to be clearer because the colored background is visually outstanding.”

“The colored background makes comparison much easier than colored strokes.”

Although the difference between the badge and the badge with patch is just the reverse of the background color and the foreground color, users’ preference yields much different results. It becomes understandable when interpreting the answer as the proportion of color counts. Because the badge with patch presents color with a patch-shape block, it brings a sharper sense of intuition by using a seemingly more colorful badge. And for the same reason, participants felt it more visually pleasant. As for the reason why our participants dislike the sparkline, these words may explain:

“The worm-like sparkline glyph makes me uncomfortable when the view is full of them.”

“I feel tired when trying to figure out which glyphs are similar and which are not.”

According to the evaluation of the designs, we cannot differentiate their performance for locating single items. Nevertheless, we confirm our hypothesis that sparkline has limitation in depicting subtle changes. It surprises us that although the time of completion does not differentiate the designs, the badge with patch receives more user preference than the other ones. It encourages us to strengthen both the design and the user interaction of the badge with patch. To further study the user experience aspect of the design as well as the usability of the system, we conducted an additional user study with the badge with patch design.

6.2 Experiment 2: Evaluation on System Usability

Our target users are regular Wikipedia readers, but not researchers or domain experts. Thus the user experience is an important aspect for evaluating the system. Other than objective measurements on

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
badge	0.91	<u>1.47</u>	0.82	1.08	1.71	0.82	<u>21.93</u>	11.83	3.04	1.76
sparkline	<u>1.26</u>	0.65	<u>3.48</u>	<u>11.10</u>	1.12	<u>4.31</u>	12.51	<u>16.30</u>	<u>3.59</u>	<u>9.73</u>
badge with patch	1.04	0.65	0.91	0.71	<u>2.43</u>	0.76	17.76	11.49	0.82	4.91
	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20
badge	0.96	1.04	0.58	1.55	1.63	13.78	<u>1.38</u>	1.15	<u>0.96</u>	5.14
sparkline	<u>2.45</u>	<u>1.66</u>	<u>0.65</u>	<u>12.20</u>	<u>17.27</u>	<u>16.32</u>	1.04	<u>2.99</u>	0.87	<u>18.47</u>
badge with patch	0.65	0.58	0.41	2.90	2.16	7.33	1.76	1.89	0.76	12.77

Table 1: The standard deviations between the actual series and series stroked by the 20 participants. The worst performance by each participant is underlined. The average standard deviations are 3.68, 6.89 and 3.63 for the badge, the sparkline and the badge with patch, respectively. Both the badge and the badge with patch worked effectively and largely outperformed the sparkline.

efficiency of speed and accuracy, in the second experiment we try to record users’ reactions on aesthetics, intuitiveness, and learnability of the system.

In the second experiment, we recruited another 20 participants to our lab to explore the pages of similar trend or of PLR with the specific pages, and to deduce the related topic with the relations with our system. To focus on the evaluation of the system instead of the visual designs, we choose the mostly adopted design—the patch with patch in Experiment 2. In the end, a questionnaire was presented to the participants, with six questions regarding the representation of our visualization designs and the usability of the system. They were asked to answer the questions in a Likert scale (see Figure 10), ranging between 1 to 5 (negative to positive).

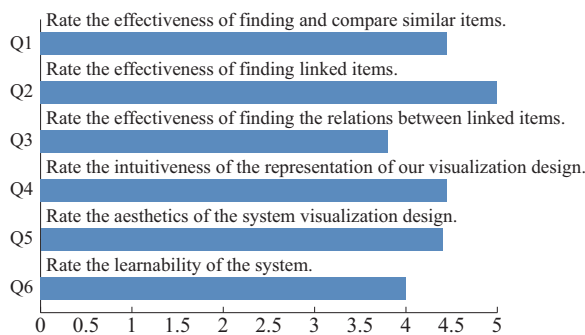


Figure 10: Likert scale questions and results on system user friendliness and usability evaluation. The results ranges between 1 to 5 (negative to positive).

Figure 10 shows the results of the six questions regarding user friendliness and system usability. Overall, the participants consider that our system is highly effective, intuitive, visually appealing, and relatively easy to learn, especially in the aspects such as to find the page link items (Q2: 100% 5 points), to compare the similar items (Q1), the intuition (Q4) and aesthetics (Q5) of our visualization design. The participants also state that there are some rooms to improve especially on learnability (Q6) and semantic exploration (Q3). To be specific, they want to infer a general context with the PLR key word network. But without supplementary searches, they sometimes cannot discover the story behind the key words. They suggest that we integrate the external search into the system for convenience.

7 CONCLUSIONS

Visualizing ranking trends of large rank time series data is a challenging task, let alone understanding the correlations between rank items. In this paper we propose three glyph representations of large rank time series, which avoid visual clutter while maintain the continuity of evolving time series. In terms of the Wikipedia top view

statistics case, we construct semantic networks for pages with similar ranking trends and characterize the correlations with page link information. Based on the WikiTopReader system, we conduct case studies and two user studies, which verify its effectiveness, efficiency, and user experience issues in detecting evolving ranking patterns and page-wise correlations.

For future work, we intend to strengthen the badge with patch design with more flexible user interactions. We want to enhance the semantic network by explaining why two pages are linked together. We also plan to apply the application on real streaming data, updating daily reports of Wikipedia top query.

REFERENCES

- [1] Alon Efrat, Quanfu Fan, and Suresh Venkatasubramanian. Curve matching, time warping, and light fields: New algorithms for computing similarity between curves. *Journal of Mathematical Imaging and Vision*, 27(3):203–216, 2007.
- [2] Christopher Collins, Fernanda B Viegas and Martin Wattenberg. Parallel tag clouds to explore and analyze faceted text corpora. In *Visual Analytics Science and Technology, 2009. VAST 2009. IEEE Symposium on*, pages 91–98. IEEE, 2009.
- [3] Conglei Shi, Weiwei Cui, Shixia Liu, Panpan Xu, Wei Chen, and Huamin Qu. Rankexplorer: Visualization of ranking changes in large time series data. *IEEE Transactions on Visualization and Computer Graphics*, pages 2669–2678, 2012.
- [4] Edward R Tufte, and PR Graves-Morris. *The visual display of quantitative information*, volume 2. Graphics press Cheshire, CT, 1983.
- [5] Jishang Wei, Zeqian Shen, Nee Sundaresanl, and Kwan-Liu Ma. Visual cluster exploration of web clickstream data. In *Visual Analytics Science and Technology (VAST), 2012 IEEE Conference on*, pages 3–12. IEEE, 2012.
- [6] Michael Batty. Rank clocks. *Nature*, 444:592–597, 2006.
- [7] Mingxuan Sun, Guy Lebanon, and Kevyn Collins-Thompson. Visualizing differences in web search algorithms using the expected weighted hoeffding distance. In *Proceedings of the 19th international conference on World Wide Web*, pages 931–940, New York, NY, USA, 2010.
- [8] Paul Kidwell, Guy Lebanon, and William Cleveland. Visualizing incomplete and partially ranked data. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1356–1363, 2008.
- [9] Romain Vuillemot and Charles Perin. Investigating the direct manipulation of ranking tables for time navigation. In *Proceedings of the 2015 Annual Conference on Human Factors in Computing Systems (CHI 2015)*. ACM, 2015.
- [10] Samuel Gratzl, Alexander Lex, Nils Gehlenborg, Hanspeter Pfister, and Marc Streit. Lineup: Visual analysis of multi-attribute rankings. *Visualization and Computer Graphics, IEEE Transactions on*, 19(12):2277–2286, 2013.
- [11] Willis D. Ellis. *A source book of Gestalt psychology*. Kegan Paul, Trench, Trubner & Company, 1938.
- [12] Xi C Chen, Abdullah Mueen, Vijay K Narayanan, Nikos Karampatziakis, Gagan Bansal, and Vipin Kumar. Online discovery of group level events in time series. *Proceedings of the 2014 SIAM International Conference on Data Mining*, pages 632–640, 2014.