Dynamic Topic Detection and Tracking: A Comparison of HDP, C-Word, and Cocitation Methods

Wanying Ding and Chaomei Chen
College of Computing and Informatics, Drexel University, Philadelphia, PA, 19104.
E-mail: {wd78, cc345}@drexel.edu

Cocitation and co-word methods have long been used to detect and track emerging topics in scientific literature, but both have weaknesses. Recently, while many researchers have adopted generative probabilistic models for topic detection and tracking, few have compared generative probabilistic models with traditional cocitation and co-word methods in terms of their overall performance. In this article, we compare the performance of hierarchical Dirichlet process (HDP), a promising generative probabilistic model, with that of the 2 traditional topic detecting and tracking methods—cocitation analysis and co-word analysis. We visualize and explore the relationships between topics identified by the 3 methods in hierarchical edge bundling graphs and time flow graphs. Our result shows that HDP is more sensitive and reliable than the other 2 methods in both detecting and tracking emerging topics. Furthermore, we demonstrate the important topics and topic evolution trends in the literature of terrorism research with the HDP method.

Introduction

Topic detection and tracking have an essential role in bibliometrics, data mining, and many other areas. Topic detection aims to identify significant topics from a document collection, whereas topic tracking aims to follow the evolution of an identified topic. Identified topics and patterns are essential to the understanding of a subject matter.

Cocitation analysis and co-word analysis are two widely used traditional methods in bibliometrics. Cocitation analysis was proposed and proved to be informative in revealing semantic relationships among documents by Henry Small (1973). It has been used widely to detect specialties of topics in a document collection (Ding, 2011; Ding, Yan, Frazho, & Caverlee, 2009; Yan & Ding, 2012). Researchers have examined the validity of these methods. Healy, Rothman, and Hoch (1986) studied both cocitation and co-word methods and identified a number of limitations. For example, a cocitation analysis is relatively restricted to a particular type of scientific communication and could favor mature scientific fields over emerging areas.

Michel Callon introduced the method of co-word analysis (Callon, Courtial, Turner, & Baun, 1983) to study the content of a document directly. This method is a powerful way to reveal semantic relations in scientific literature and is still widely used (Munoz-Leiva, Viedma-del-Jesus, Sanchez-Fernandez, & Lopez-Herrera, 2012; Ronda-Pupo & Guerras-Martin, 2012; Zhang et al., 2012). However, co-word analysis has a limitation: it can not handle synonym and polysemy terms very well because the method primarily relies on the appearances of words, and thus does not lead itself as an ideal tool for subsequent clustering tasks.

With the rapid development of probabilistic methods, generative probabilistic models have drawn researchers’ attention. Since David Blei introduced the latent Dirichlet model (LDA) (Blei, Ng, & Jordan, 2003), researchers have applied LDA to topic detection and tracking. LDA has been shown to be a competent approach, and has stimulated many applications in finding topics and specialties in an area (Maowen, Zhang, Weiyao, & Qiang, 2012; Xin, Xiang, Chen, Huang, & Hao, 2013; Yu, Agichtein, & Benzi, 2012). However, LDA requires researchers to specify the number of topics in advance (Griffiths & Steyvers, 2004; Wallach, Mimno, & McCallum, 2009). It is hard for users to come up with a specific number of topics without a good understanding of the underlying data. In the context of dynamic topic analysis, the problem becomes even more critical. New topics will emerge while existing topics may disappear from time to time. The number of topics will change, thus it would be too much of a burden on the users to specify the number of topics to begin with. In order to address this problem, researchers have proposed three main strategies:

1. Try out different numbers of topics, and choose the best one in terms of the accuracy (Griffiths & Steyvers, 2004).
2. Start with a large number of topics in LDA, and then aggregate similar topics as the final result (Song, 2010).
3. Optimize the generative probabilistic model itself, using a nonparametric method to automatically generate the number of topics in the sampling process.

The Hierarchical Dirichlet process (HDP) is the most widely accepted model for the third purpose (Teh, Jordan, Beal, & Blei, 2006).

Although it is possible to apply LDA repeatedly and check perplexity values in order to find an optimal number of topics, it has a major practical limitation in that one may have to make a large number of trials in order to narrow down the range of the potentially promising numbers of topics. Furthermore, LDA itself does not provide readily usable guidance for making such trials.

The second strategy heavily depends on the clustering method used, and is found not to be superior to LDA. Therefore, in this paper we focus on the third strategy and choose HDP as the primary data-driven approach for detecting and tracking topics associated with a document collection.

Theoretically, HDP has many advantages over cocitation and co-word methods; for example, HDP is a method based on intrinsic content-specific signals to detect topics rather than on the extrinsic information such as citations of references. As a result, it may identify topics directly from the text with a more specific focus than the potentially diverse range of topics derived from cocitations. HDP maps terms to a semantic space by using Dirichlet prior, alleviating the synonym problem faced by a co-word analysis. To our knowledge, no study has empirically demonstrated this advantage of HDP in practice. In this article, we compare the performance of cocitation, co-word, and HDP methods in topic detection and tracking. The goal is to establish the extent to which the new HDP method performs with reference to the two classic methods. To our knowledge, this is the first study to compare a generative probabilistic model with traditional topic detecting methods.

This article is organized as follows: Related Work reviews research related to all three methods. Data and Method introduces the data and method we used in this study. The Results and Discussion presents the findings and discusses their implications.

Related Work

Cocitation analysis and co-word analysis are two classical methods for topic detection and tracking. Cocitation analysis utilizes the cocitation relationship to construct a network among citing and cited documents. In addition, one may use clustering algorithms to identify a number of clusters in the network. Each cluster corresponds to an underlying topic. Co-word analysis, which differs from cocitation analysis, constructs a matrix based on words’ co-occurrence frequencies. Then it applies matrix decomposition methods to divide the matrix to a number of sub-matrices and each of them represents a topic. These two methods have been widely used for decades. Even though these two methods have been proved to be useful in topic detection and tracking, they both have some weaknesses.

Recently, LDA has been commonly adopted in bibliometrics to detect topics, but some deficiencies of LDA have undermined LDA’s role in topic detection. The first deficiency is that LDA requires the topics to be input in advance. However, it is often hard to do so. More important, this prerequisite becomes even less justifiable if our interest is in finding emerging topics and how they evolve over time. It is not practically feasible for researchers to provide the exact number of topics for each year without any prior knowledge of the data. The second deficiency is that LDA tends to distribute the topics evenly. Figure 1a shows the topic distribution generated by LDA, and it is clear that every topic has a share of approximately 1/K of the whole (K is the number of topics). However, in the real world an even distribution is not common, and resembles a random distribution. HDP, on the other hand, is a nonparametric Bayesian model (Ferguson, 1973), which can automatically decide the number of topics. In addition, HDP tends to distribute the data according to the Pareto law. Figure 1b shows the topic distribution generated by HDP, and we can see that there are two obviously predominant topics with many secondary topics. One can draw down to lower-level structures by repeatedly applying the HDP process. Given these reasons, HDP is considered more competent than LDA in dynamic topic detecting and tracking.

HDP is a probabilistic model based on a Dirichlet process (DP). There are many perspectives to explain DP, and in this paper we consider DP as a Chinese restaurant process (CRP) (Blei, 2007). In a CRP, a document is seen as a restaurant and each word in this document as a customer. Customers will arrive at the restaurant one by one, and each of them will choose a table, and the table here represents a topic. We can operate the distribution process as follows (Blei, 2007):

1. The first customer always chooses the first table to sit down.
2. The nth customer chooses an unoccupied table with probability of \( \frac{\alpha}{n-1+\alpha} \), and chooses an occupied table with the probability of \( \frac{c}{n-1+\alpha} \), where c represents the number of people who have already chosen that table, and \( \alpha \) is the parameter to control the table assignment.

If we consider that the words in a document arrive one by one, we can construct a successive conditional distribution of the nth word \( \theta_n \) given \( \theta_1, \theta_2, \ldots, \theta_{n-1} \). We can get \( \theta_n \) from the following formula (Teh et al., 2006):

\[
\theta_n | \theta_1, \theta_2, \ldots, \theta_{n-1}, \alpha, G \sim \frac{\sum_{i=1}^{c_n} \frac{c_i}{n-1+\alpha} \delta_{\theta_i}}{n-1+\alpha} + \frac{\alpha}{n-1+\alpha} \ast G
\]

(Formula 1)

Where \( c_i \) represents the number of customers at that table, and \( \delta \) is the word distribution of a topic which this table belongs to, and G is a probability used to generate a topic for the new table.
When it comes to generating a topic for a new table, there are two choices: (a) choose a topic from the existing topics; and (b) generate a totally new topic. So, we need a distribution to control this kind of topic assignment. Thus, we add another distribution over the table distribution described in Formula 1, and the topic assignment can be described as the Formula 2 (Teh et al., 2006):

$$\psi \mid \psi_1, \psi_2, \ldots, \psi_{i-1}, \gamma, H = \sum_{k=1}^{K} \frac{m_k}{m-1+\gamma} \delta_k + \frac{\gamma}{n-1+\gamma} * H$$  
(Formula 2)

Where $m$ represents how many tables have been assigned, and $m_k$ represents how many tables have been assigned to topic $k$, and $\delta_k$ is the word distribution on the topic. $H$ is the baseline distribution to generate new topics.

Here the two-level DP is known as an HDP. The HDP can be represented as a graphical model shown in Figure 2. In Figure 2, $H$ and $\gamma$ controls the topic distribution $G_0$, which is shared across the entire document collection. For each document $j$, a local distribution $G_j$, drawn from $G_0$, controls how a word distributes over the existing topics—$\beta$ affects the $\delta$ distribution in Formulas 1 and 2.
By manipulating the parameters of $\alpha$, $\beta$, and $\gamma$, we can obtain different distributions: word-topic distribution, word-document distribution, table-topic distribution, and document-topic distribution and so on. Based on these distributions, one may compare the results generated by HDP with that by cocitation and co-word methods.

**Data and Method**

**Data Collection**

Given the objectives of our study, an ideal data set would contain cited references for cocitation analysis, and various components of text for co-word analysis and topic modeling. We are familiar with the research on terrorism through our prior studies on the structure and dynamics of relevant research (Chen, 2006). Thus, we constructed our data set on the general topic of terrorism to be analyzed by all three analytic methods. We obtained our data set by using a topic search for publications between 1995 and 2012 on the topic of “terrorism” in the Web of Science (WoS), which is a widely used source for scientometric and bibliometric studies. The resultant 9,033 bibliographic records formed our data set.

The time frame is chosen so that it will cover major terrorist events such as the Oklahoma City bombing and the September 11, 2001 terrorist attacks. Although this is not a standard data set, it is straightforward to reconstruct by anyone who has access to the WoS using the same query.

We considered three kinds of information sources to analyze: title, keyword, and abstract. An abstract contains too much noise to generate an accurate result, so we decided to exclude abstracts from the subsequent analysis. As for keywords, we found that within 9,033 documents, only 4,978 of them have keywords. The incompleteness of keywords makes them unreliable to provide a full picture of the document collection. Finally, we focus on the title as the primary source of information for our study.

**Tools and Methods**

We used CiteSpace (http://cluster.cis.drexel.edu/~cchen/citespace/) to perform the cocitation and co-word analyses, and we developed a special-purpose computer program to perform topic modeling with HDP. Unlike traditional HDP, which uses a word as a unit of analysis, we used a noun phrase as a unit of analysis in our study. Noun phrases provide more specific information than a single word, thus they are easier to interpret. For example, the phrase “posttraumatic-stress-disorder” represents the complete concept of a psychological condition caused by a trauma. If we use words as a unit of analysis, the connections between words and the underlying concepts may not be as explicit as using phrases. Before removing common stop words, we POS-tagged all the titles, and kept all the phrases for HDP.

As shown in Figure 2, the performance of HDP varies with different combinations of three parameters, $\alpha$, $\beta$, and $\gamma$. We need to find the combination that produces the best HDP performance. We consider two indicators of the performance of HDP: the perplexity of the result and the number of topics.

Perplexity is defined based on the theory of entropy as a measure of the quality of a probabilistic model. The smaller the value of perplexity, the better the result is considered to be. Perplexity can be calculated as follows:

$$p(p) = 2^H(p) = 2^{-\sum_x p(x) \log_2 p(x)}$$  \hspace{1cm} (Formula 3)

The number of topics in HDP is automatically determined. In cocitation and co-word methods, the number of topics is selected as follows: According to the results generated by CiteSpace, both cocitation and co-word methods on average generate 10–20 clusters from every year’s data. In order to keep the results comparable, we choose combinations of the three parameters such that the number of HDP topics is within the range 10–20. Within this range, the least number of topics is considered optimal.

Based on these two reasons, we measured the performance of HDP with a function defined in Formula 4. We chose the parameter combination that generates the lowest $S$ value as the optimal parameter combination of HDP.

$$S(\alpha, \beta, \gamma) = \sqrt{(\text{perplexity})^2 + (\text{number of topic})^2}$$  \hspace{1cm} (Formula 4)

Then we compared the three methods on two tasks: topic detecting and topic tracking. For different tasks, we used different comparison and visualization methods.
Methods for Topic Detection

In topic detection, if a method can find most topics detected by other methods, this method is considered to be powerful and its strength is measured by a “coverage” indicator (see Formula 5). The higher the coverage score of a method, the more powerful the method is in topic detecting.

We also introduced the concept of an important topic associated with an individual method. The importance of a topic is the proportion of the topic among all the topics identified by the method. Besides, we may be more concerned about the topics with high proportions in each year than the topic with low proportions. We applied the coverage measurement not only on the whole set of topics, but also the set of important topics to see whether HDP is better in both of these two kinds of coverage:

\[
\text{Coverage}(\text{Method}_A \text{ to } \text{Method}_B) = \frac{|T_A \cap T_B|}{|T_B|} \tag{Formula 5}
\]

where \(T_A\) represents the topics identified by method A, and \(T_B\) represents the topics identified by method B. \(T_A \cap T_B\) represents the similar topics identified by A and B, respectively. The similarity between two detected topics is defined in terms of the overlapping terms between the two topics. After normalization, we retained only two topics as similar topics if their similarity is larger than 0.2.

Then we used a hierarchical edge bundling (Holten, 2006) graph to show the relationships between topics detected by the three methods. If two topics are similar, they would be shown with a connecting link in the hierarchical edge-bundling graph; otherwise, the relationship would not be shown. In addition, the bundling coefficient was set to 0.85 to obtain a clearer visualization. The hierarchical edge bundling is shown in Figure 3.

The visualizations in Figure 3 reveal three types of patterns. Topics detected by all three methods are considered significant; those detected by two methods as ordinary; and topics only detected by one method are defined as exclusive. Examples of these patterns can be seen in Figure 4.

Methods for Topic Tracking

In topic tracking, we developed a time flow graph to depict topic evolution over time (see Figure 5). In Figure 5,
each color represents a year, and the topics in each year are ranked by their proportion in the entire topic set of that year; the most important topic is placed on in the top of the rectangle, and the least important one is shown at the bottom. The lines represent the similarities between topics. The thicker a line is, the more similar the two topics are.

In this part, we linked up the similar topics in two consecutive years, so that we could track the topics from time to time.
We defined two indicators to measure each method's ability in tracking topics over time. The similarity between topics was calculated by Kullback-Leibler (KL) divergence, which is shown as Formula 6:

\[ \text{Dis}(P, Q) = \frac{1}{2} \left( \sum_i \ln \left( \frac{P(i)}{Q(i)} \right) P(i) + \sum_i \ln \left( \frac{Q(i)}{P(i)} \right) Q(i) \right) \]

(Formula 6)

We defined two indicators to measure each method’s ability in topic tracking capability: topic sensitivity and topic persistence.

Topic sensitivity is used to check whether the method is sensitive to topics. It is measured by the time interval, which is the time length between the time when a topic was detected and the time when the topic turned out to be a critical topic.

Topic persistence checks whether the method can keep tracking the topic over time. It is measured by the time interval between the birth and death of a topic. The longer the time interval, the more persistent a method is at tracking.

In addition, we identified three kinds of topic evolution patterns from Figure 6. They are one-to-one evolution, one-to-many evolution, and many-to-one evolution. Generally speaking, all three kinds of evolution are very common, so a competent method should have the ability to find all of these three kinds of topic evolution patterns.

**Results and Discussion**

**Topic Detection**

Comparison of cocitation, co-word, and HDP in topic detecting. We applied cocitation, co-word, and HDP methods to detect topics from the terrorism document collection year by year, and then visualized the relationship between those topics in hierarchical edge-bundling graphs. The graphs are shown in Figure 3. The red arc represents HDP, the green arc represents cocitation, and the blue one represents co-word. Each label outside of the ring represents a topic identified by the corresponding method. Topics near the beginning of an arc are more important than topics that are farther away from the beginning. The lines between two topics indicate that the two topics are similar. The thickness of a bundled edge is proportional to the strength of similarity. The similarity is calculated using the Jaccard coefficient in terms of overlapping terms between the two topics.

Figure 7 shows a brief comparison of the similarities between the three methods from 1995 to 2012. As shown in Figure 7, all of these three methods are very closely related, but there still exist some differences:

1. When detecting topics, an ideal method would be able to separate topics apart from each other. In other words, an ideal method would minimize within-group topic correlations. Both cocitation and co-word analysis have such correlations (for example, in the result of 1998, both cocitation and co-word have such correlations), while HDP has none. In this sense, HDP appears to be able to differentiate topics better than the other two methods.

2. HDP has a close similarity relationship with both cocitation and co-word methods. However, bundled lines between cocitation and HDP are denser than bundled lines between co-word and HDP. This means that HDP and cocitation cover more overlapping topics than HDP and co-word. In other words, the important topics detected by one method are more likely to be found by another method, and vice versa. Thus, HDP is more similar to the cocitation method than the co-word method. This observation indicates that the result generated by HDP is more similar to the result generated by cocitation.

3. The relationship between cocitation analysis and co-word analysis was close, but the similarity is relatively weaker. This phenomenon might result from possible vocabulary changes within this area. But HDP maintains a relatively stable relationship with both of these two methods. From this perspective, we could conclude that HDP is not so sensitive to the change of vocabulary in a field, but maintains the semantic feature better.

Figure 7 shows that the results generated by cocitation, co-word, and HDP are distinct but similar, and this is because they are based on very different principles to generate topics. However, even though they represent different approaches, their results are closely related. A topic detected by one method may be similar to a topic detected by another method. Here, related topics are the same underlying topic detected by different methods. If one method can find all the topics detected by other methods, we consider this method to be good at topic detection. We measure this property in terms of the coverage of a method. The coverage of each method is shown in Table 1. Because each method generates different numbers of topics, the function is not necessarily symmetric, that is, Coverage (A, B) may not be the same as Coverage (B, A).

Table 1 shows that HDP has the highest coverage compared with both cocitation method and co-word method. This means that HDP tends to find more topics than others, and further indicates that HDP is likely to be a comprehensive tool in topic detection.

Besides the general comparison, one may be more interested in how the important topics relate to each other. We
count the number of times each important topic detected by one method connects to important topics detected by other methods and call the number the Coverage. The denominator of Formula 5 in this situation is the same for all the three methods, so we do not need to divide it. In addition, because one topic may link to several other topics, the total number of the linkage might be more than five. The final result is shown in Figure 8, where CW represents co-word and CC represents cocitation.

Figure 8 shows that the HDP has a higher coverage than both cocitation and co-word methods. Thus we concluded that HDP is better at detecting the important topics.

In short, HDP performs better in topic detection than the other two methods in terms of the coverage of both overall topic discovery and important topic discovery.

Detecting different types of topics. As mentioned earlier, we identified three types of topics: topics detected by all three methods, topics detected by two of them, and topics detected by only one method. We refer to the first kind of topic as “significant topics,” the second as “ordinary topics,” and the third as “exclusive topics.” We may have two explanations for exclusive topics. Exclusive topics may represent a method’s ability to detect topics that other methods may fail to detect. Alternatively, they may also be explained as a kind of noise. In different situations, we shall take into

<table>
<thead>
<tr>
<th></th>
<th>Cocitation</th>
<th>Co-word</th>
<th>HDP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cocitation</td>
<td>–</td>
<td>0.486</td>
<td>0.544</td>
</tr>
<tr>
<td>Co-word</td>
<td>0.564</td>
<td>–</td>
<td>0.521</td>
</tr>
<tr>
<td>HDP</td>
<td>0.636</td>
<td>0.573</td>
<td>–</td>
</tr>
</tbody>
</table>
account different interpretations of exclusive topics. In contrast, significant topics can be seen as a good indicator of the capability of a topic detection method. The more significant topics a method can find, the better this method is likely to be, because significant topics reflect the consensus of all three methods. The proportion of each type of topic generated by each method can be represented by the area graph presented in Figure 9. Figure 9 (top) represents the proportions of methods on significant topics, Figure 9 (middle) represents the ordinary topic, and Figure 9 (bottom) represents the exclusive topic.

As shown in Figure 9, generally speaking, HDP and cocitation are better than co-word in terms of detecting significant topics. Even though cocitation performs better than HDP before 2000, HDP’s ability to detect significant topics begins to exceed cocitation’s after 2000. Especially after 2008, HDP has significantly surpassed cocitation.

HDP could not compete with cocitation or co-word analysis in ordinary topic detecting. It is obvious that both cocitation and co-word are stronger than HDP in this area. Cocitation and co-word performed similarly in this category.

For exclusive topics, HDP performs a little better than the other two methods in this field, and comparatively, co-word performs a little better than cocitation.

In summary, HDP is good at detecting both significant topics and exclusive topics. Cocitation does well in detecting significant and ordinary topics, but not the exclusive ones. Co-word does well in both ordinary topic detection and exclusive topic detection.

Within the significant topics, we can find another type of topic—“critical topic,” which is detected by all three kinds of methods, and each of the methods ranks it as important topic (ranks in top five). Figure 10 depicts a schematic diagram of such topics. This kind of topic is evidently of great significance in the subject area.

Our examination found the following eight critical topics:

1. Biological Terrorism (1996)
5. Al-Qaida and Suicide Terrorism (2004)
6. Nuclear and Chemical Terrorism (2011)
7. Influence of Terrorism (2011)
8. Islamist Terrorism and Olympic Game (2012)

The detailed information about these topics is shown in Table 2.

The topic of “biological terrorism” in 1996 results from the Oklahoma City bombing, which happened in 1995. When the Oklahoma City bombing happened, a lot of attention was paid to the risk of some larger-scaled attacks such as chemical and biological terrorism. Terrorism began to draw the public’s attention in 2001 because of the 9/11 event in that year, and from terms such as “America Mass,” “Destruction,” and “Terrorist Group,” we can infer that many articles published in this year are related to the 9/11 attacks.

In 2002, 1 year after the 9/11 attacks, people began to address whether there would be more chemical biological attacks and how would the government deal with them. So a lot of papers were about the “Federalism Preparation,” “Biological Weapons,” and “Medical Response.”

The keywords in 2003 and 2004, like “Sacrificing Civil Liberty” and “Global War” indicate that some articles published in these 2 years focused on how to fight terrorism.

In 2011, 10 years after the 9/11 attacks, many researchers continue to study the subject of terrorism. Two critical topics, topics #6 and #7, are detected in this year. Topic #6 is still about terrorism itself, characterized by terms such as “Nuclear Terrorism” and “Transnational Terrorism.”
FIG. 9. Proportion of each type of topic in each method. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]
#7 is about the influence of terrorism, especially the influence on economics.

Terrorism, which was the biggest threat to the London Olympic Games, was detected in 2012 with topics such as “Islamist Terrorism” and terrorism from the “Middle East.”

In order to understand the proportion of the critical topics detected by all three methods more intuitively, Figure 11 shows a direct view of the critic topic proportion change in each of the three methods. From Figure 11, we can see that HDP and cocitation analysis share very similar proportion trends in critical topic detection, whereas co-word follows a very different pattern in this part. Two of three methods share a similar change trend, and the third one is different. We may tentatively conclude that HDP and cocitation is more reliable than co-word in topic detecting.

**Topic Tracking**

**Comparison of cocitation, co-word, and HDP.** If we link the topics in 2 consecutive years by their similarities, we can see how a topic evolves from one year to another. We used the KL Divergence to measure the similarities between topics from 2 consecutive years and generated a time flow graph for topics generated by the three methods. Similar topics in consecutive years are connected by explicit lines (see Figure 5). Each color represents a year, and topics in each year are ranked as a proportion of the whole topic collection of that year. The most important topic is placed at the top of the rectangle, topics with the lowest importance scores are placed at the bottom. The thicknesses of the lines are proportional to the similarities between topics.

Generally speaking, topics generated by co-word and cocitation methods have fewer strong similarity relationships than topics generated by HDP, and the number of relationships decreases over time, while HDP maintains a relatively stable number of relationships. In this sense, we conclude that HDP is a more stable method for topic tracking.

Lines for the co-word and cocitation methods tend to gather at the bottom of the graph, while the line for HDP tends to appear in the middle of the graph or higher. This means HDP is more likely to find the relations between important topics than both co-word and cocitation methods. In practical applications, one tends to be more interested in relationships between important topics, so HDP is likely to be a more suitable tool.

As mentioned previously, we defined two indicators to compare the ability of these three in topic tracking: sensitivity and persistence. In the following we will use these two indicators to compare the three methods, and in addition we will check the three methods’ topic change pattern recognition capability.

To measure topic sensitivity, we chose the critical topics in Table 2 as our objects. We used the lead time, which is the time elapsed from the time a topic is first detected until the topic becomes a critical topic, as the indicator of sensitivity. A longer lead time indicates greater topic tracking.

---

**TABLE 2.** Critical topics.

<table>
<thead>
<tr>
<th>Topic no.</th>
<th>Time</th>
<th>Representative words</th>
<th>Proportion (HDP)</th>
<th>Proportion (cocitation)</th>
<th>Proportion (co-word)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1996</td>
<td>Biological Terrorism, New Threat</td>
<td>7.29%</td>
<td>7.0%</td>
<td>8.0%</td>
</tr>
<tr>
<td>2</td>
<td>2001</td>
<td>America Mass, Destruction, Civil War, Terrorist Group, American Ground Strategy</td>
<td>9.59%</td>
<td>10.0%</td>
<td>11.0%</td>
</tr>
<tr>
<td>3</td>
<td>2002</td>
<td>Federalism Preparation, Medical Response, Biological Threat, Biological Weapons</td>
<td>6.04%</td>
<td>13.0%</td>
<td>14.0%</td>
</tr>
<tr>
<td>4</td>
<td>2003</td>
<td>Sacrificing Civil Liberty, War on Terrorism</td>
<td>13.2%</td>
<td>12.0%</td>
<td>7.0%</td>
</tr>
<tr>
<td>5</td>
<td>2004</td>
<td>AI-Qaida, Global War, Psychology, Suicide Terrorism</td>
<td>5.81%</td>
<td>9.0%</td>
<td>12.0%</td>
</tr>
<tr>
<td>6</td>
<td>2011</td>
<td>Nuclear Terrorism, Chemical, Transnational Terrorism</td>
<td>13.4%</td>
<td>14.4%</td>
<td>12.0%</td>
</tr>
<tr>
<td>7</td>
<td>2011</td>
<td>Influence, Economic Growth, Economic Consequence</td>
<td>7.40%</td>
<td>8.89%</td>
<td>20.7%</td>
</tr>
<tr>
<td>8</td>
<td>2012</td>
<td>Middle East, Democracy, Olympic Game, Islamist Terrorism, Radiation</td>
<td>9.62%</td>
<td>22%</td>
<td>9.5%</td>
</tr>
</tbody>
</table>

FIG. 10. Critical topic. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]
capability because it can help us detect topics long before they become widely recognized. So we track these topics in Table 2 to see the average lead time of every method, and the result is shown in Table 3.

As shown in Table 3, on average, HDP has the longest lead time of 1.88 years, co-word is in the second place (0.38 years), and cocitation has the shortest lead time (0.25 years). HDP can detect a critical topic almost 2 years ahead, whereas both co-word and cocitation can only detect less than half a year ahead. However, in comparison, co-word is slightly more sensitive than cocitation. This finding suggests that HDP is the most sensitive in topic tracking, and cocitation is the least sensitive one. This result is consistent with the common belief that cocitation has a significant lag when it comes to topic tracking.

The persistence of a topic is defined based on the time interval between the appearance and the disappearance of a topic. We counted the time interval for all the topics detected by all three methods, and show them in Figure 12. Even though co-word detected more topic links, most of them began in 1995 and ended in 1997, and this might result from the phenomenon that many topics in 1995 merge into one topic in 1996 and this topic in 1996 split to many other topics in 1997. However, our focus here is on the length of the topics links but not the number of links detected.

Figure 12 shows that in cocitation, no topic lasted longer than 5 years, but a topic may last as long as 8 years in HDP and co-word. Compared to co-word, HDP detects more topic links with length longer than 4 years. So HDP and co-word are better than cocitation, and HDP is better than co-word in topic tracking in terms of persistence.

What’s more, in Data and Method we defined three kinds of topic change patterns in Figure 6: one-to-one, one-to many (split), and many-to-one (merge). We examined all three methods for these topic change patterns and the results are shown in Table 4. From Table 4, we can see that all three methods are good at tracking the one-to-one relationship, especially the cocitation method; more than 60% of relationships it detects are one-to-one relationships. It seems that co-word detected more merge and split relationships, but most of them happened between 1995 and 1997. As mentioned above, a lot of topics in 1995 had merged into one topic in 1996, and the merged topic split into many other topics in 1997. If we remove the relationships starting in 1995 and ending in 1997, many fewer merge and split relationships are detected by co-word (the result is shown in the column of Co-Word (*) in Table 4). On the contrary, HDP is more stable and capable of detecting one-to-many and many-to-one relationship than the other two methods. So in a different situation we might

### Table 3. Sensitivity of topic tracking.

<table>
<thead>
<tr>
<th>Year</th>
<th>Detected by all three</th>
<th>Detected by HDP</th>
<th>Lead time (year)</th>
<th>Detected by co-word</th>
<th>Lead time (year)</th>
<th>Detected by cocitation</th>
<th>Lead time (year)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1996</td>
<td>1996</td>
<td>0</td>
<td>1995</td>
<td>1</td>
<td>1996</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>2001</td>
<td>1998</td>
<td>3</td>
<td>2000</td>
<td>1</td>
<td>2001</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>2002</td>
<td>2000</td>
<td>2</td>
<td>2002</td>
<td>0</td>
<td>2002</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>2003</td>
<td>2001</td>
<td>2</td>
<td>2003</td>
<td>0</td>
<td>2003</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>2004</td>
<td>2002</td>
<td>2</td>
<td>2004</td>
<td>0</td>
<td>2004</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>2011(#6)</td>
<td>2009</td>
<td>2</td>
<td>2010</td>
<td>1</td>
<td>2010</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2011(#7)</td>
<td>2010</td>
<td>1</td>
<td>2011</td>
<td>1</td>
<td>2010</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2012</td>
<td>2009</td>
<td>3</td>
<td>2012</td>
<td>0</td>
<td>2012</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td>1.88</td>
<td></td>
<td></td>
<td>0.38</td>
<td>0.25</td>
<td></td>
</tr>
</tbody>
</table>
need to choose different methods to track topics. If we want to track one-to-one relationships, cocitation is a preferred choice, but if we want to see how topics derive and merge, HDP is a good candidate.

In summary, HDP performs the best in topic tracking, while co-word is second best, and cocitation ranks last.

Tracking important topic threads. As shown previously, since HDP is good at topic tracking, we use HDP to find the significant topic threads in terrorism. In order to see a thread clearly, we retain the top 20 terms for each topic in HDP and redraw Figure 5 with the top five terms only in order to be consistent with the results of co-word and cocitation (Figure 13). The time period in Figure 13 is the same as in Figure 5.

As shown in Figure 13, a significant topic trend emerged in 2002 in the top part of the graph, and this trend remained visible until 2007. As we know that “9/11” happened in 2001, it is understandable to have such a trend in research from 2002. The topic of this thread changed from “September, the war on the terrorism” to “suicide terrorism, age of terrorism” from 2002 to 2007, and after 2007 this topic was divided to many other less dominant topics. We conclude from the trend that the event of “9/11” impacted the research of terrorism for 6 years, and now its effect has faded, and new topics emerge.

After 2007, topics in this area changed very quickly. Many dominant topics in one year were no longer important topics in the next year, and few topics dominated the subject matter for long. One good example is that, in the year 2010, the topic of “Politics and Arsenal” was the most important one, but it dropped to the fourth place in 2011, and disappeared altogether in 2012. A quick change in this graph means a quick switch of research focus in this area.

Conclusion

This paper compared the performance of three methods for topic detection and tracking: cocitation, co-word, and HDP. By testing them on the data set from the research area of “Terrorism,” the quality of topic detection was analyzed with hierarchical edge bundling visualization and a series of detailed comparisons. The quality of topic tracking was

<table>
<thead>
<tr>
<th></th>
<th>HDP</th>
<th>Co-word</th>
<th>Cocitation</th>
<th>Co-word(*)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 to 1</td>
<td>0.5933</td>
<td>0.4606</td>
<td>0.6172</td>
<td>0.8265</td>
</tr>
<tr>
<td>1 to many(split)</td>
<td>0.1866</td>
<td>0.2368</td>
<td>0.1953</td>
<td>0.0306</td>
</tr>
<tr>
<td>many to 1(merge)</td>
<td>0.2201</td>
<td>0.3026</td>
<td>0.1875</td>
<td>0.1429</td>
</tr>
</tbody>
</table>

analyzed with a time flow visualization. We found that HDP is better than the other two methods in terms of sensitivity and persistence.

Moreover, in topic detection the cocitation method performs better than the co-word method, but in topic tracking, the co-word method outperforms the cocitation method.

We detected eight critical topics and one thread of significant topics in terrorism research. The eight critical topics are: Biological Terrorism (1996), American Terrorism Strategy (2001), Biological Threat (2002), War on Terrorism (2003), Al-Qaeda and Suicide Terrorism (2004), Nuclear and Chemical Terrorism (2011), Influence of Terrorism (2011), and Islamist Terrorism and Olympic Game (2012). From the thread of topics, we found that “9/11” indeed had a profound impact on this research area, and its impact lasted for 6 years, but now its effect has faded and new topics have emerged and changed quickly.

However, some limitations in our current study need to be improved on in future research. First, we used only the title terms of a paper, and did not make use of information from the keywords and abstracts. Using title terms alone may miss some important semantic information. We will incorporate additional information to expand the scope of our approach. Second, we used the data from just one research area (terrorism), but did not test the methods on data from other fields. Future studies should address the extent to which the approach is applicable to a wider range of scientific domains or other types of documents, such as social media content. Third, in topic tracking, topics of high similarity are considered the same topic, but they may be distinct topics. Next, we will explore methods that can track a specific topic but may not rely on similarities determined by overlapping terms.

The data set used in this study can be reconstructed anyone who has access to the WoS. Researchers who are interested in studying the same topic would be able to obtain the data set with the same topic search query.

In conclusion, our result shows that HDP is a promising approach to topic detection and tracking when compared with traditional methods such as cocitation and co-word studies. Further research is needed to explore a broader range of data sources and disciplines.

Acknowledgments

This work is supported in part by an NSF IIP grant 1160960. We can make the data set available upon request.

References


Song, Z. (2010). Research on text categorization based on LDA. Master Degree Dissertation. Xi’an University of Technology, Xi’an, China.


