

---

## Research and Applications

# Trends in anesthesiology research: a machine learning approach to theme discovery and summarization

Alexander Rusanov,<sup>1</sup> Riccardo Miotto,<sup>2</sup> and Chunhua Weng<sup>3</sup>

<sup>1</sup>Department of Anesthesiology, Columbia University, New York, New York, USA, <sup>2</sup>Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, New York, USA and <sup>3</sup>Department of Biomedical Informatics, Columbia University, New York, New York, USA

Corresponding Author: Chunhua Weng, PhD, Department of Biomedical Informatics, 622 West 168th Street, PH-20, New York, NY 10032, USA (cw2384@columbia.edu)

Received 28 December 2017; Revised 18 March 2018; Editorial Decision 22 March 2018; Accepted 23 August 2018

### ABSTRACT

**Objectives:** Traditionally, summarization of research themes and trends within a given discipline was accomplished by manual review of scientific works in the field. However, with the ushering in of the age of “big data,” new methods for discovery of such information become necessary as traditional techniques become increasingly difficult to apply due to the exponential growth of document repositories. Our objectives are to develop a pipeline for unsupervised theme extraction and summarization of thematic trends in document repositories, and to test it by applying it to a specific domain.

**Methods:** To that end, we detail a pipeline, which utilizes machine learning and natural language processing for unsupervised theme extraction, and a novel method for summarization of thematic trends, and network mapping for visualization of thematic relations. We then apply this pipeline to a collection of anesthesiology abstracts.

**Results:** We demonstrate how this pipeline enables discovery of major themes and temporal trends in anesthesiology research and facilitates document classification and corpus exploration.

**Discussion:** The relation of prevalent topics and extracted trends to recent events in both anesthesiology, and healthcare in general, demonstrates the pipeline’s utility. Furthermore, the agreement between the unsupervised thematic grouping and human-assigned classification validates the pipeline’s accuracy and demonstrates another potential use.

**Conclusion:** The described pipeline enables summarization and exploration of large document repositories, facilitates classification, aids in trend identification. A more robust and user-friendly interface will facilitate the expansion of this methodology to other domains. This will be the focus of future work for our group.

**Key words:** natural language processing, machine learning, data mining

---

## BACKGROUND AND SIGNIFICANCE

In anesthesiology, as in many scientific fields, published works, such as journal articles or meeting abstracts, serve a historical record of academic work in the field. While each document stands alone as a record of a particular scientific investigation, a collection of such documents, as a whole, contains covert information about prevalent

research themes, or topics. Uncovering these themes facilitates the discovery of topical patterns and historical trends in the field, which leads to a deeper appreciation of shifts in research focus and aids in the prediction of future research directions.

Conventionally, experts would review documents in a repository, then summarize the major themes and trends. While this approach works well for smaller collections, it becomes more and

more difficult as the amount of information to be reviewed increases. In anesthesiology, as in most established scientific disciplines, the amount of available information is increasing at an exponential rate, a phenomenon which has been attributed to both the increasing rate at which scientific literature is published and to the growth in publications in less traditional venues such as online open access journals, conference proceedings, etc.<sup>1</sup> Given the current size of anesthesiology collections, traditional methods for theme discovery and summarization are now impractical—the “information overload” problem.<sup>2,3</sup> The growth in size of repositories also creates new challenges in organizing them and in finding specific documents of interest within them—the “needle in haystack” problem.<sup>4,5</sup>

Fortunately, computational methods for coping with these problems have arisen from the fields of text mining, which is devoted to deriving high level abstractions, such as trends, patterns, and relationships, from collections of textual data,<sup>6</sup> and information retrieval, which seeks to facilitate the finding of relevant information in large collections.<sup>7</sup> One such method is topic modeling, a text mining approach that relies on statistical inferences to automatically discover themes, or topics, within large collections of text documents thus aiding in their summarization.<sup>8,9</sup> The discovered thematic relationships can then also be leveraged to aid in information organization and retrieval.<sup>10,11</sup>

## OBJECTIVES

Our first objective was to build upon these methods to develop a pipeline for unsupervised theme extraction from scientific abstracts and a novel automated method for summarization of thematic trends. Our second objective was to applying this pipeline to a large collection of anesthesiology abstracts in order to gain an understanding of the general themes and trends in anesthesiology research by considering a large collection of anesthesiology research documents as a whole. Such an understanding is of interest not only to individual investigators but also to departments, institutions, and funding agencies, where it can be used to guide research and recruitment agendas and to optimize the allocation of limited financial resources. Furthermore, by uncovering the topical structure of a repository, relationships and interconnections between individual documents are illuminated, and areas of convergence and divergence in research interests are identified. This has the potential to be used as a powerful tool for identifying others with shared interests and fostering collaboration.

## MATERIALS AND METHODS

First, we briefly review the topic modeling technique, and then describe the specifics of its application. Next, we use the resulting topic model to make sense of prevalent themes and trends in anesthesia research. Finally, we generate a thematic-similarity-based network visualization of the document collection and demonstrate how it can aid in information retrieval and collection exploration.

### What is topic modeling?

Topic modeling is a machine learning method for discovering themes—or topic within a corpus of documents. In this study, we chose to use Latent Dirichlet Allocation (LDA),<sup>12</sup> one of several existing topic modeling implementations.<sup>8</sup> LDA was chosen not only because it is one of the best studied and widely used algorithms

but also because it has been shown to be effective at identifying semantically relevant topics in science texts<sup>13</sup> and to outperform many newer algorithms on this task.<sup>14</sup>

## Implementation

We applied our pipeline to the American Society of Anesthesiologist Annual Meeting Abstract Archive (<http://www.asaabstracts.com/strands/asaabstracts/abstractArchive.htm>), henceforth referred to as the Archive. We chose the Archive because we were interested in research trends in the entire broad field of anesthesiology and felt that a repository of work presented at the largest national anesthesiology meeting would represent the broadest and most generalized collection of anesthesia-related research. We also chose to work with abstracts, as opposed to entire manuscripts, both because of their currency (scientific results typically presented in abstract form before their submission, acceptance, and publication in journals), and because we felt the need to be more representative of the collective scientific work of the anesthesiology community (while each abstract represents some academic work, only a fraction of these get published as manuscripts).

Application of the proposed pipeline to meeting abstracts is also advantageous because scientific meeting archives also tend to not grow too rapidly in document number. Given the space and time constraints of physically presenting abstracts at a meeting, the number of new abstracts remains somewhat constant from year-to-year (see Results section) and the collection grows linearly. At the current rate of growth, the Archive will not reach 100K documents for 40–50 more years and will not reach 500K for more than 300 years. Anesthesiology may not even exist as a unified field by then, and if it does, other methods for data storage, retrieval, and evaluation will surely replace our current ones. It is impossible to predict what technology will exist that far in the future. Though at their current size the meeting abstract archives are getting too large for manual human curation, the proposed machine learning method should suffice for many years to come.

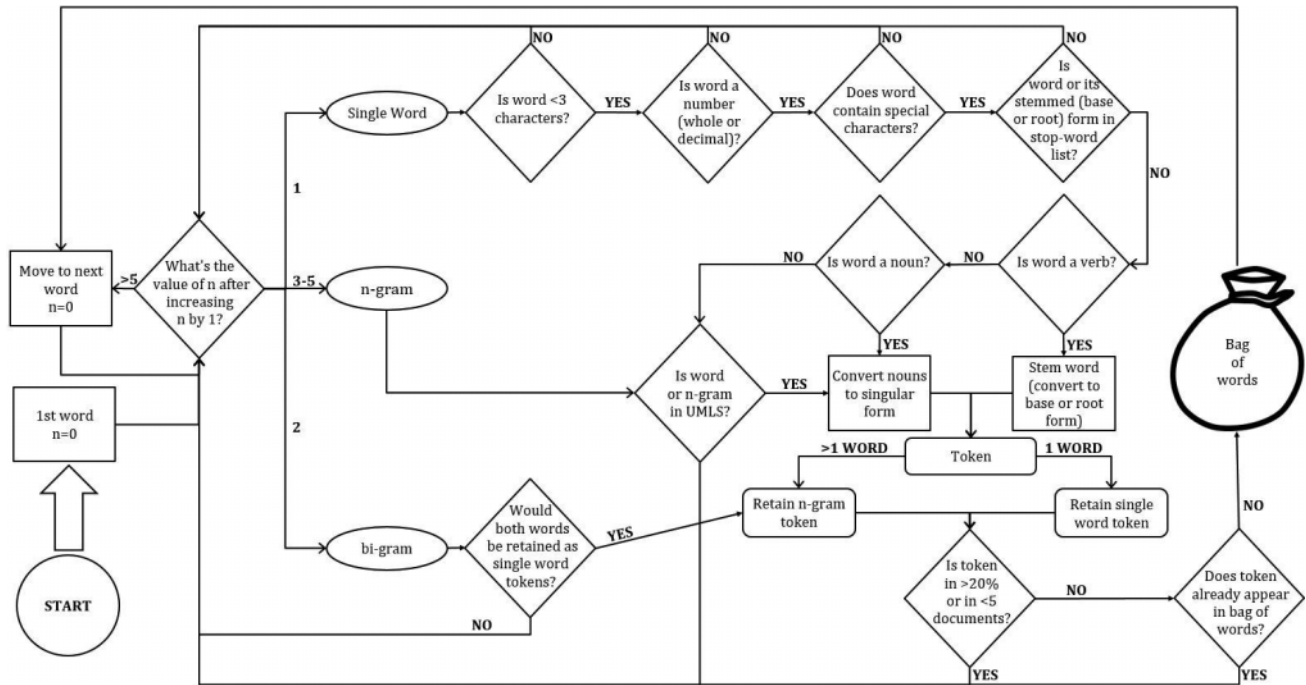
The implemented pipeline consists of the following 8 steps: (1) text retrieval from the Archive, (2) text processing, (3) topic model generation, (4) topic naming, (5) frequency and popularity ranking, (6) statistical analysis, (7) similarity measurement, and (8) network visualization.

### Text retrieval

Abstracts in the Archive are organized by meeting year (years 2000–2013 were available at the time of website access) and further organized by the category in which they were presented. Every year, each abstract presented at the meeting was assigned a category. Category assignment was initiated by abstract authors and then approved by meeting organizers. Using a custom-built web-crawler, we extracted the text and its associated metadata (Year, Category, Abstract Title, Authors, and Abstract Number) for each of the 22 262 available abstract.

### Text processing

Basic text processing techniques<sup>15</sup> were applied to the text and title of each abstract to extract tokens. Using the Natural Language Toolkit<sup>16</sup> (NLTK), the text was first split into sentences and then into words. Special characters were removed, and each remaining word was annotated with a part-of-speech (POS) tagger, defined in the NLTK, to identify the grammatical role of each word.<sup>17</sup> Each word was then considered independently as a single word, and as



**Figure 1.** A flowchart representation of the text processing pipeline. Starting with the first word in the first document, tokens are generated based on a series of rules as outlined. Tokens are then filtered for inclusion in the bag-of-words, as illustrated. This process is repeated for every word in every document. UMLS: Unified Medical Language System.

the start of  $n$ -grams (continuous sequence of  $n$  words), where  $n$  ranged between 2 and 5.

A bag-of-words (BOW) consisting of single word (1-gram) and multiword ( $n$ -gram) tokens representing the entire collection was generated using the logic outlined in Figure 1. For single words, those  $<3$  characters in length, those consisting only of numbers, and those containing characters other than letters and numbers were removed. Next, all remaining words were reduced to their base (root) form (stemmed), using the NLTK Porter Stemmer. Stop-words are frequently occurring words that are used to construct ideas but have no semantic meaning on their own.<sup>18</sup> A standard list of English stop-words and additional custom corpus-specific stop-word list (eg “introduction,” “methods,” “results,” “anesthesiology,” and “protocol”) were used. Words, and stemmed words, contained in the stop-word lists were removed. From the remaining set of words, all those contained in the Unified Medical Language System (UMLS)<sup>19</sup> Metathesaurus and all verbs (in their stemmed form) and nouns (in their singular form obtained using the NLTK Lemmatizer) were retained as tokens.

For all  $n$ -grams, if the entire  $n$ -gram string was a recognizable UMLS concept, the string was kept as a token after reducing all nouns contained within it to their singular form. Additionally, for bigrams, if both words were retained as single word tokens, then the bigram string was retained as a separate token. All unique tokens for all abstracts were combined to create the BOW. Last, rarely (appearing in  $<5$  documents) and frequently occurring (appearing in  $>20\%$  of documents) tokens were removed from the BOW.

**Topic model generation**

This BOW was then used as input for the LDA model. An adaptation of the inference algorithm described by Hoffman et al<sup>20</sup> was used to generate topics (see <https://radimrehurek.com/gensim/>

models/ldamodel.html). Determination of the “correct” number of topics covering the corpus for probabilistic models remains an open problem. Decreasing the number of topics decreases the resolution of the derived topics with the extreme case being one topic representing the entire corpus—ie an “Anesthesiology Research” topic. Conversely, increasing the number of topics results in many topics of poor quality (see below). Though some possible quantitative solutions have been proposed,<sup>21,22</sup> using quantitative methods to judge topic models is not always meaningful.<sup>8</sup> Topics from models deemed superior based on quantitative metrics, such as perplexity and held out likelihood, have been shown to be semantically less meaningful to human users than topics from models that perform poorer on these metrics.<sup>14</sup> We thus chose to rely on a mixture of qualitative and quantitative methods. We used a quantitative measure of held out perplexity to evaluate models with varying numbers of topics. Models were trained on 80% of the data set and tested on the remaining 20%. Perplexity was noted to decrease with increasing number of topics in a manner similar to what has been previously reported,<sup>12</sup> with the rate of decrease being rapid at first and eventually reaching an inflection point where it slows, approaching zero (a nearly flat line). We used the inflection point (observed at 133 topics) as an approximation of the optimal number of topics. We then performed qualitative assessments of several models around this approximation. A 100-topic model was ultimately chosen.

**Topic naming**

A topic is deemed to be cohesive if the highest probability tokens within the topic relate to a similar theme or subject matter. Although LDA generated topics have been demonstrated to be generally cohesive,<sup>14</sup> not all topics generated by the model correspond to a single coherent semantic concept.<sup>23</sup> Our analysis focuses on those topics deemed cohesive enough to be named.

Though some attempts at automated naming have been made,<sup>24,25</sup> assigning a semantically meaningful, or interpretable, name for these topics is a labor-intensive task that is still best performed by humans. For general knowledge corpora, this can be done on a large scale using crowd-sourcing; however, the task becomes more difficult for specialty texts where domain knowledge is necessary.<sup>26</sup> The topics inferred in this study were named by the lead author (A.R.), who, as an anesthesiologist, is familiar with the specialty texts being analyzed.

### Frequency and popularity ranking

For every year in the analysis, the frequency with which each of the 100 topics occurred was determined. A topic's frequency was calculated as the percentage of abstracts, in a given year, where the topic was 1 of the top 10 most probable topics. Top 10 was empirically chosen as the cut-off based on inspection of results produced by using top 1, top 3, top 5, top 10, and top 20.

The frequency metric is useful for determining whether a particular topic becomes more or less mentioned over time. However, it does not describe how a topic's popularity relative to other topics changes. As an example, a topic whose usage frequency increases may actually become relatively less popular if the frequency of other topics becomes even more frequent. In order to discern trends in topic popularity, for each year all topics were sorted in order of decreasing frequency and a rank was assigned to every topic with 1 corresponding to the most frequent topic for that year and 100 to the least frequent.

### Statistical analysis

Temporal trends in both frequency and rank were assessed using linear regression—a parametric test which defines the significance of a relationship between 2 variables, year ( $X$ ) and frequency or rank ( $Y$ ) in this case, related by a line described by the equation  $Y = \beta_0 + \beta_1 X$ . Each of the 100 topics was treated as a predictor variable, and a regression line was fitted for each. We were interested in determining whether the predictor value (year) for each topic had an effect on changes in the response variable (either frequency or rank). More simply, we wanted to identify topics which exhibited a significant linear relationship between time and frequency or popularity of the topic. In a linear regression analysis, the  $P$ -value of the regression coefficient (also known as slope, or slope coefficient) tests the null hypothesis that  $H_0: \beta_1 = 0$ . With  $P$ -values smaller than the significance level the null hypothesis is rejected, and the alternate hypothesis ( $H_A: \beta_1 \neq 0$ ) is accepted. We used 0.01 as the significance level for identifying topics with a nonzero regression coefficient ( $\beta_1$ ). Only topics with a significant trend in both frequency and rank (ie regression coefficient  $P$ -value  $< .01$  in both the frequency and rank regression models) were retained for further analysis. In addition to the presence or absence of a significant trend, via significance testing, linear regression provides the direction (indicated by the sign of  $\beta_1$ ) as well as the magnitude, or rate of change or strength, of the trend (indicated by the absolute value of  $\beta_1$ ).<sup>27</sup>

### Similarity measurement

For each document, the similarities between its own and each of the other document's topical probability distributions were calculated using the cosine similarity metric. The cosine similarity metric has been previously used for comparison of documents<sup>7</sup> and has also been used by others to compare similarity between LDA-generated

topics.<sup>28–30</sup> The metric returns a value between 0 and 1, with 1 corresponding to identical distributions.

### Network visualization

To visualize the similarity network, document pairs with similarity lower than 0.9 were excluded. This threshold was chosen for demonstration purposes as it prevented overcrowding of the resultant network and thus facilitated visualization of the network for print. This is one of the limitations of a static representation of such a network in print, as opposed to dynamic exploration where the threshold can be varied. As noted by Smith et al,<sup>30</sup> thresholds should be independently determined for each data set to optimize semantic relevance of the similarity associations, while minimizing the presence of irrelevant links.

Pairs of nodes with similarity  $\geq 0.9$  were imported into Cytoscape<sup>31</sup> as an edges table. An undirected network was generated where nodes (circles) represent individual abstracts connected by edges (lines) based on similarity. A force-directed layout was applied. Nodes were labeled with abstract number and year and colored based on their category. Edge thickness was set to correspond to similarity (ie more similar nodes are connected by thicker edges).

## RESULTS

On average 1590 abstracts were presented each year. The highest number presented in a single year was 2141 in 2007, and the lowest was 1295 in 2013. Text processing resulted in 61 388 unique tokens to be used as the BOW representation for the LDA model. These tokens appear in the abstracts for 4 160 532 times, 187 tokens per abstract on average.

Figure 2 provides a visual representation of some of the generated topics as word-clouds, where the font size for each word (token) is proportional to its probability in the underlying distribution. As discussed previously, not all topics were deemed coherent enough to be named (ie bottom right topic in Figure 2). The top 20 tokens (along with their associated probabilities) for each of the 100 topics are provided in Supplementary Material Table S1.

### Topical trends

Of the 23 topics with a statistically significant increase in both frequency and popularity rank, 18 were named and are shown in Table 1.

Of the 29 topics with a statistically significant decrease, 20 were named and are shown in Table 2.

The topics are sorted by popularity rank in the most recent year (2013), with most popular topics appearing on top in Table 1 and least popular ones appearing on top in Table 2. The current “hot topics” in anesthesia research (Table 1) include “Multivariate Analytical Methods,” “Practice Management,” “Patient Safety and Quality Improvement,” “Information Systems,” “Devices and Algorithms,” and “Education and Critical Care Outcomes.”

For each topic in Tables 1 and 2 the rank and frequency at the start (year 2000) and end (year 2013) of the data collection period are shown. To facilitate readability, line charts depict the data for the intervening years, with the y-axis corresponding to rank or frequency, and the x-axis corresponding to time. The y-axis for frequencies has lower values at the bottom and higher values at the top, while the y-axis for ranks is reversed, with lower numbers (corresponding to higher ranks) appearing at the top, and higher numbers (lower ranks) appearing at the bottom. The x-axis for all line



**Figure 2.** Token-cloud representations of a selection of topics. The size of the token in each cloud corresponds to its probability for the topics. Names assigned to topics appear below each token-cloud. A topic which could not be named due to being too broad is shown in the bottom right.

charts increases from year 2000 on the left to 2013 on the right. To facilitate visualization of details in the line charts, the y-axis maximum and minimum values for each chart are set to the maximum and minimum values for rank or frequency for the corresponding topic over the entire 14-year time period. The slope of the linear regression line ( $\beta_1$ ) for both frequency and rank represents the strength of the temporal trend. In the interest of clarity, only the magnitude of the trend (absolute value), and not its direction, is shown, as the direction is already demonstrated by the corresponding line chart. By revealing the year-to-year variations for each topic, the line charts provide detail a linear trend alone would fail to capture. As an example, they would allow the reader to note if a topic with a strong increasing trend decreased before increasing.





































Figure 3 is a scatterplot comparing the frequency trend magnitude ( $\beta_1$ ) to the rank trend magnitude for each topic in Table 1, while Figure 4 is a similar plot for the topics in Table 2. In each of the 2 plots, the mean frequency  $\beta_1$  for all topics in that plot is represented by a vertical line (mean = 0.30 for Figure 3 and 0.24 for Figure 4). Topics to the right of the line show a stronger trend (increase or decrease) in frequency than the mean trend for all increased (Figure 3) or decreased (Figure 4) topics. That is, among all the topics with increasing (Figure 3) or decreasing (Figure 4) frequency, those on the right side are topics that increased or decreased the most. Similarly, the mean rank  $\beta_1$  (63.43 for Figure 3 and 64.47 for Figure 4) is represented by a horizontal line. Again, the topics above the line have a rank trend with a magnitude higher than the mean trend magnitude.

The plots are thus divided into quadrants. The topics in the top right quadrants changed the most in terms of both their absolute use in the texts (frequency) and their popularity (rank), or use compared to the use of other topics. The change in popularity for topics in these quadrants results from the change in their frequency of use. The topics in the bottom right quadrant changed more than the mean in their overall use (frequency), but not in their use compared to other topics (rank). These tend to be topics that were already popular and whose continued increased use kept them popular or those that were already unpopular and remain unpopular because of their decreased use. The top left quadrant is the opposite, with topics here having a higher than the mean change in popularity compared to other topics but a lower than the mean change in frequency of use. The changes in popularity of topics in this quadrant are attributable mostly to changes in frequency of use of other topics, and not to the changes in their own frequency of use. Finally, the bottom left quadrant shows topics which had lower than the mean changes in frequency and rank. These topics changed least in terms of both their use and their popularity.

### Topical similarity network

Figure 5 is a zoomed out overview of the most connected abstracts in the network visualization generated as described above. Visual inspection of the network reveals abstracts forming homogeneously colored groups, corresponding to the categories assigned to each abstract by meeting organizers. Furthermore, groups representing

**Table 1.** Topics with statistically significant increases in frequency and rank

Topic name	Rank				Frequency			
	In 2000	In 2013	Line chart	$\beta_1$	In 2000	In 2013	Line chart	$\beta_1$
Multivariate regression	11	1		36.20	15.69	34.67		0.82
Practice management	17	3		38.27	14.23	25.10		0.43
Patient safety/quality improvement	35	6		64.31	11.10	22.01		0.48
Information systems	29	7		54.38	12.04	21.47		0.41
Devices and algorithms	27	9		53.77	12.41	20.31		0.34
Education	33	10		56.60	11.61	19.54		0.38
Critical care outcomes	44	11		69.75	9.85	17.92		0.38
Pain	42	12		61.36	10.15	17.30		0.24
Multivariate regression	43	13		70.72	9.93	17.22		0.36
Obesity	79	17		79.64	5.99	14.90		0.38
Cardiac	24	17		50.65	13.07	14.90		0.11
Postoperative cognitive dysfunction	75	24		74.86	6.79	12.51		0.20
Critical care outcomes	61	27		73.90	7.96	11.35		0.17
Epidural complications	64	41		72.92	7.52	9.34		0.16
Simulation	91	52		75.06	4.45	8.11		0.19
Stress/inflammation	83	58		70.40	5.47	7.10		0.11
Airway management	96	62		70.03	4.09	7.03		0.13
Liver transplantation	84	62		68.93	5.40	7.03		0.12

categories dealing with similar subjects, or those that overlap in certain areas of research, are located close each other due to the many edges between their constituent documents (see red circle and black square in Figure 5).







































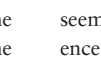
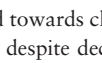
## DISCUSSION

Some of the “hot topics,” such as “Information Systems” and “Devices and Algorithms,” relate to recent advances in biomedical informatics,<sup>32</sup> and their increased use in medicine spurred in part by the Health Information Technology for Economic and Clinical Health (HITECH) Act<sup>33</sup> of 2009. Others such as “Multivariate

Analytical Methods,” are a consequence of the increased availability and use of the massive amounts of data resulting from the use of these technologies. Yet others such as “Patient Safety and Quality Improvement,” “Practice Management,” and “Critical Care Outcomes,” reflect the recent emphasis on provision of quality care that is patient centered and affordable. These were spurred by national (passage of Patient Protection and Affordable Care Act and establishment of the Patient Centered Outcomes Research Institute),<sup>34</sup> as well as specialty specific (establishment of the Anesthesia Quality Institute<sup>35</sup> in 2008) factors.

While the most recent popularity rankings identify current “hot topics,” examination of the changes in frequency and popularity over time, provide a glimpse into trends in research interests. Trends

**Table 2.** Topics with statistically significant decreases in frequency and rank

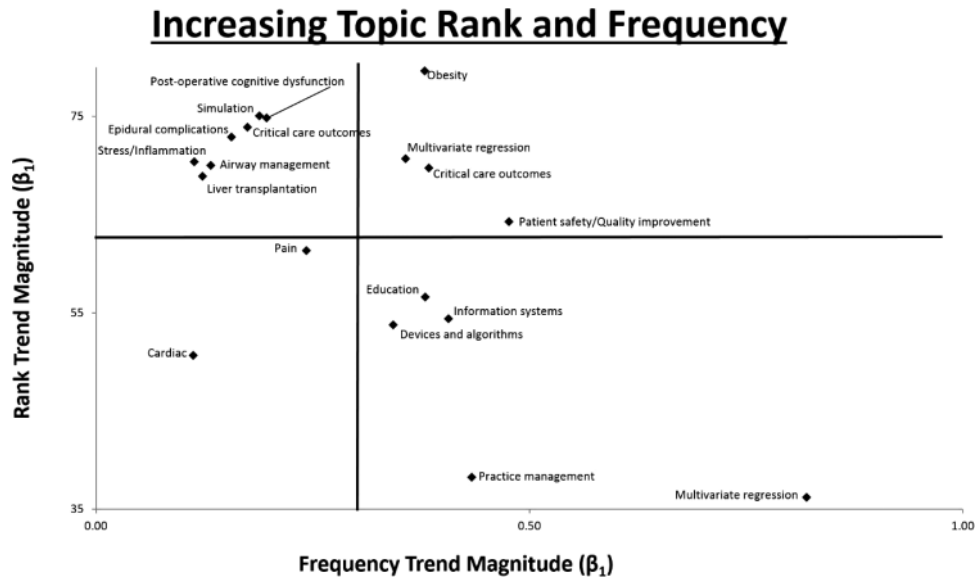
Topic name	Rank				Frequency			
	In 2000	In 2013	Line chart	$\beta_1$	In 2000	In 2013	Line chart	$\beta_1$
Membrane channels	59	95		64.53	8.25	4.02		0.14
Laparoscopic surgery	57	92		65.43	8.54	4.32		0.14
Muscle function	41	90		77.94	10.37	4.71		0.31
Receptors	36	88		73.08	10.95	5.02		0.25
Pulmonary hypertension	25	75		78.33	12.99	5.87		0.36
Membrane channels	12	75		78.69	14.82	5.87		0.43
Electrical currents	49	72		65.16	9.27	6.02		0.16
Cardiopulmonary resuscitation	49	65		57.16	9.27	6.95		0.11
Neuromuscular blockade	38	62		70.77	10.80	7.03		0.21
Cardiac function	32	57		67.19	11.68	7.18		0.22
Sedation	25	54		62.09	12.99	7.72		0.21
Hypotension/shock	12	53		71.50	14.82	7.88		0.28
Cerebral perfusion	10	43		65.10	15.77	9.19		0.27
Central nervous system	8	38		73.52	16.28	9.58		0.46
Epidural anesthesia	14	34		48.54	14.67	9.81		0.18
Inhaled anesthetic cerebral protection	20	34		64.76	13.87	9.81		0.25
Receptors	21	31		66.53	13.43	9.96		0.26
Infusions	4	19		57.38	20.22	13.98		0.31
Electroencephalography	6	16		43.52	18.98	14.98		0.20
Cellular molecular pathways	5	14		38.20	19.05	16.14		0.13

help detect “up-and-comers”—topics that, though not among the most popular, have shown significant increase in popularity. One such topic is “Obesity” which, at a rank of 17, is not among the most popular but is growing rapidly as evidenced by the higher than average rate of increase ( $\beta_1$ ) in both rank and frequency. The increased interest in this topic coincides with the increased prevalence of obesity in both anesthesia patients<sup>36</sup> and the general population.<sup>37</sup>

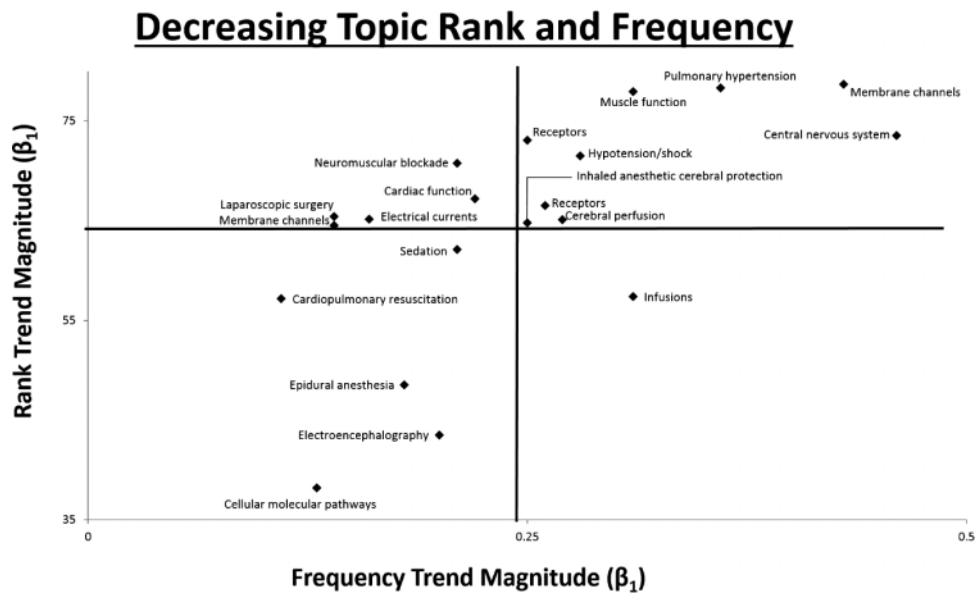
Conversely, trends also point out topics in which research interest has faded (Figure 4). These tended to be basic science topics such as “Membrane Channels,” which demonstrated the largest decline, as well as “Receptors,” “Cellular Molecular Pathways,” “Electrical Currents.” The general trend in anesthesiology research interest

seems to be a shift away from basic science and towards clinical science, operations, and epidemiology. However, despite declining interest, basic science research remains important with “Cellular Molecular Pathways” still the 3rd most prevalent topic in the entire corpus.

In the network visualization (Figure 5), the grouping of thematically similar abstracts is completely unsupervised. The agreement between this automated topic-based categorization and the manual categorization performed by humans (meeting organizers) who had no access to the topic model, not only further validates the model but also highlights the potential of topic models to be used as an aid to human curation of scientific works, especially as the number of such works continues to grow.



**Figure 3.** A scatterplot comparing the magnitude of the trend (absolute value of  $\beta_1$ ) in frequency, to the magnitude of the trend in rank for all topics with statistically significant increases in both frequency and rank. Solid black lines are placed at the mean  $\beta_1$  for all the plotted points (mean  $\beta_1 = 0.3$  for frequency and 63.43 for rank), dividing the plot into quadrants. Note that some generated topics, though composed of different mixtures of words, actually relate to the same semantic theme and were thus given the same name that is two of the generated topics were both named “Multivariate regression.”



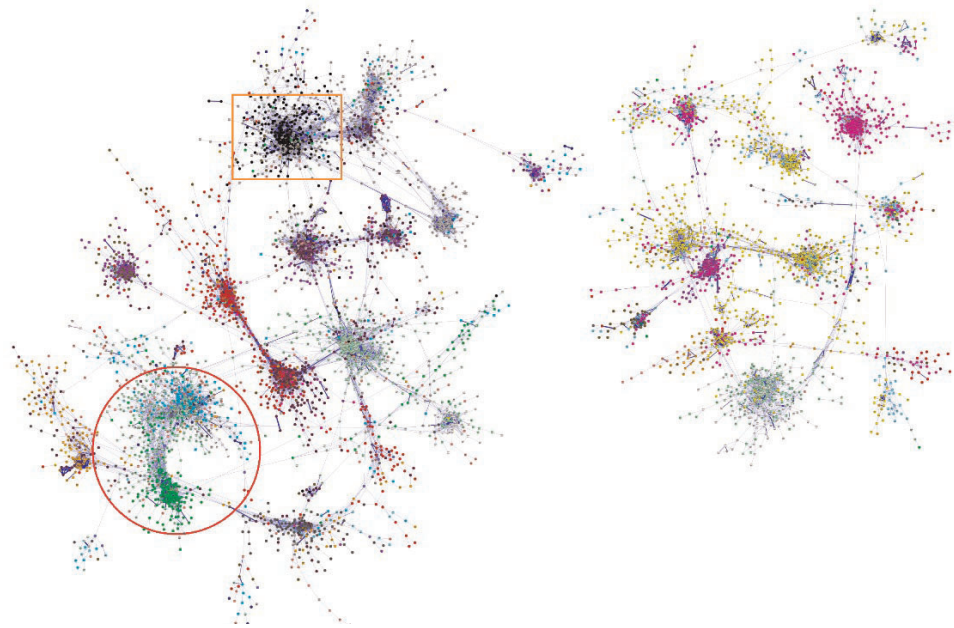
**Figure 4.** A scatterplot comparing the magnitude of the trend (absolute value of  $\beta_1$ ) in frequency, to the magnitude of the trend in rank for all topics with statistically significant decreases in both frequency and rank. Solid black lines are placed at the mean  $\beta_1$  for all the plotted points (mean  $\beta_1 = 0.24$  for frequency and 64.47 for rank), dividing the plot into quadrants. Note that some generated topics, though composed of different mixtures of words, actually relate to the same semantic theme and were thus given the same name, that is two of the generated topics were both named “Membrane channels,” 2 other topics were both named “Receptors.”

The network visualization is also useful in locating specific documents of interest amongst the many contained in such collections. Exploration of the network along the edges emanating from an abstract of interest, allows for the discovery of other thematically similar abstracts. Among these related abstracts, some may belong to a different category and/or year, making them difficult to identify using the Archive’s current organizational scheme.

#### Future work

By aiding in discovery of thematically related works, the network can become a tool for research idea generation and can foster collaboration among researchers. The latter would benefit from the inclusion of metadata such as authorship, institutional affiliations, funding sources, etc. in the network visualization. Thematic similarity between works of different researchers or institutions can be discovered by replacing abstract nodes with author and institution





**Figure 5.** A network visualization of the 2 biggest supergroups in the similarity network of the entire American Society of Anesthesiologists Abstract Archive. Nodes represent individual abstracts and are colored based on the categories to which they were assigned in the archive. Note the automatic grouping of abstracts by category and proximity of related groups. The red circle surrounds a predominantly green group (obstetric anesthesiology and perinatology) and a predominantly cyan group (local/regional anesthesia and acute pain). Local anesthetics are widely used in obstetric anesthesia practice as well as in regional anesthesia and in treatment of acute and chronic pain. Similar techniques that is (neuraxial anesthesia) are a widely used for obstetric anesthesia and pain management. Hence, the similarity between these categories. The orange square surrounds a group of mostly gray (patient safety) and mostly black (history and education) abstracts. These 2 categories used to be merged and only became distinct after 2006. Edges are only drawn between nodes with a similarity of  $>0.9$ . Edge thickness is proportional to similarity.

nodes.<sup>38</sup> This and other enhancements and tools for visualization and interactive real-time exploration of topic models remain an active area of research<sup>39–43</sup> and are a focus of future work for our group.

In addition to work on visualization, we aim to expand the underlying document collection by continually adding newly available abstracts and by including other sources. A continually updating topic model has the advantage of being current in a rapidly changing field.<sup>44</sup> Also, as new documents are added over time new topics may emerge and existing topics may evolve (ie their underlying token distributions can change) or disappear. While LDA assumes that a predefined number of static topics describe a collection with a fixed number of documents, models that treat topics in a more dynamic way<sup>45–48</sup> may offer new insights into topic evolution in an expanding collection.

Inclusion of more sources (ie abstracts from related meetings, journal articles, textbooks, etc.) brings the model closer to being a true representation of all related knowledge. Observing the movement of trends from abstracts presented at meetings, to papers published in journals, to textbooks, and ultimately to clinical practice (as evidenced by documentation in the electronic health record) would permit tracking of knowledge dissemination from bench to bedside.

Finally, we aim to test this pipeline in the “real world” by having users evaluate specific aspects of the model such as topic cohesiveness, as well as the overall utility of the model for summarization and exploration of large repositories of scientific literature and its utility and accuracy in predicting future directions based on past trends.

## CONCLUSION

The described pipeline enables summarization and exploration of large repositories of scientific publications, facilitates document

discovery, and classification, aids in trend identification. We demonstrate how its application to the Archive provides a unique perspective on prevalent themes and trends in anesthesiology research, and assists in exploration and identification of thematically related works for research idea generation and potential collaborator identification. Application of this pipeline to other domains will provide new insights and add value to their related document repositories. To facilitate, this a more user-friendly and robust interface for model generation, topic naming, and model exploration are needed. This will be our future work.

## FUNDING

Alexander Rusanov was funded by NIH T32 Training Grant (5T32 GM008464 23) to the Department of Anesthesiology at Columbia University (PI: Charles W Emala). Riccardo Miotto and Chunhua Weng were funded by NLM grant R01 LM009886 (PI: Chunhua Weng).

*Conflict of interest statement.* None declared.

## SUPPLEMENTARY MATERIAL

Supplementary material is available at *Journal of the American Medical Informatics Association* online.

## ACKNOWLEDGMENTS

We would like to thank Anil Yaman, MS, Department of Mathematics and Computer Science, Eindhoven University of Technology, Assen, The Netherlands for his help with the statistical analysis.

## REFERENCES

1. Larsen PO, von Ins M. The rate of growth in scientific publication and the decline in coverage provided by Science Citation Index. *Scientometrics* 2010; 84 (3): 575–603.
2. Cohen AM, Hersh WR. A survey of current work in biomedical text mining. *Brief Bioinformatics* 2005; 6 (1): 57–71.
3. Zweigenbaum P, Demner-Fushman D, Yu H, et al. Frontiers of biomedical text mining: current progress. *Brief Bioinformatics* 2007; 8 (5): 358–75.
4. Conrick M. Looking for a needle in a haystack: searching the Internet for quality resources. *Contemp Nurse* 2002; 12 (1): 49–57.
5. Rodriguez-Esteban R. Biomedical text mining and its applications. *PLoS Comput Biol* 2009; 5 (12): e1000597.
6. Srivastava AN, Sahami M. *Text Mining: Classification, Clustering, and Applications*. Boca Raton, FL: CRC Press; 2009.
7. Singhal A. Modern information retrieval: a brief overview. *IEEE Data Eng Bull* 2001; 24 (4): 35–43.
8. Blei DM. Probabilistic topic models. *Commun ACM* 2012; 55 (4): 77–84.
9. Blei DM, Lafferty JD. Topic models. In: Srivastava AN, Sahami M, eds. *Text Mining: Classification, Clustering, and Applications*. Boca Raton, FL: CRC Press; 2009: 71–94.
10. Yi X, Allan J. Evaluating topic models for information retrieval. In: Amer-Yahia S, Manolescu I, Zhang Y, et al., eds. *Proceedings of the 17th ACM Conference on Information and Knowledge Management*. New York, NY: ACM Press; 2008: 1431–1432.
11. Yi X, Allan J. A comparative study of utilizing topic models for information retrieval. In: Boughanem M, Berut C, Mothe J, Soule-Dupuy C, eds. *Advances in Information Retrieval*. Berlin, Germany: Springer; 2009: 29–41.
12. Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. *J Mach Learn Res* 2003; 3 (4-5): 993–1022.
13. Griffiths TL, Steyvers M. Finding scientific topics. *Proc Natl Acad Sci USA* 2004; 101 (Supplement 1): 5228–35.
14. Chang J, Boyd-Graber JL, Gerrish S, et al. Reading tea leaves: how humans interpret topic models. In: Bengio Y, Schuurmans D, Lafferty J, et al., eds. *Advances in Neural Information Processing Systems*. vol. 22. La Jolla, CA: Neural Information Processing Systems; 2009: 288–296.
15. Manning CD, Raghavan P, Schütze H. *Introduction to Information Retrieval*. New York, NY: Cambridge University Press; 2008.
16. Bird S. NLTK: the natural language toolkit. In: Curran J, ed. *Proceedings of the COLING/ACL on Interactive Presentation Sessions*. Stroudsburg, PA: Association for Computational Linguistics; 2006: 69–72.
17. Tsuruoka Y, Tateishi Y, Kim J-D, et al. Developing a robust part-of-speech tagger for biomedical text. In: Bozanis P, Houstis EN, eds. *Advances in Informatics*. Berlin, Germany: Springer; 2005: 382–392.
18. Leskovec J, Rajaraman A, Ullman JD. *Data Mining: Mining of Massive Datasets*. 2nd ed. Cambridge, UK: Cambridge University Press; 2014.
19. Lindberg DAB, Humphreys BL, McCray AT. The unified medical language system. *Methods Inf Med* 1993; 32 (4): 281–91.
20. Hoffman MD, Blei DM, Bach FR. Online learning for latent Dirichlet allocation. In: Lafferty JD, Williams CKI, Shawe-Taylor J, et al., eds. *Advances in Neural Information Processing Systems* 23. La Jolla, CA: Neural Information Processing Systems; 2011: 856–864.
21. Arun R, Suresh V, Veni Madhavan CE, et al. On finding the natural number of topics with latent Dirichlet allocation: some observations. In: Zaki M, Yu J, Ravindran B, et al., eds. *Advances in Knowledge Discovery and Data Mining*. vol. 6118. Berlin, Germany: Springer; 2010: 391–402.
22. Teh YW, Jordan MI, Beal MJ, et al. Sharing clusters among related groups: hierarchical Dirichlet processes. In: Saul LK, Weiss Y, Bottou L, eds. *Advances in Neural Information Processing Systems*. vol. 17. Cambridge, MA: The MIT Press; 2005: 1385–1394.
23. Mimno D, Wallach HM, Talley E, et al. Optimizing semantic coherence in topic models. In: Merlo P, Barzilay R, Johnson M, eds. *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Stroudsburg, PA: Association for Computational Linguistics; 2011: 262–272.
24. Mei Q, Shen X, Zhai C. Automatic labeling of multinomial topic models. In: Berkhin P, Caruana R, Wu X, et al., eds. *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY: Association for Computing Machinery; 2007: 490–499.
25. Lau JH, Grieser K, Newman D, et al. Automatic labelling of topic models. In: Lin D, Matsumoto Y, Mihalcea R, eds. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1*. Stroudsburg, PA: Association for Computational Linguistics; 2011: 1536–1545.
26. Arnold CW, Oh A, Chen S, Speier W. Evaluating topic model interpretability from a primary care physician perspective. *Comput Methods Programs Biomed* 2016; 124: 67–75.
27. Draper NR, Smith H. *Applied Regression Analysis*. vol. 3. 3rd ed. New York, NY: Wiley; 1998.
28. Ramage D, Hall D, Nallapati R, et al. Labeled LDA: a supervised topic model for credit attribution in multi-labeled corpora. In: Mihalcea R, Koehn P, eds. *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*. Stroudsburg, PA: Association for Computational Linguistics; 2009: 248–256.
29. Liu Y, Niculescu-Mizil A, Gryc W. Topic-link lda: joint models of topic and author community. In: Bottou L, Littman M, eds. *Proceedings of the 26th Annual International Conference on Machine Learning*. New York, NY: Association for Computing Machinery; 2009: 665–672.
30. Smith A, Malik S, Shneiderman B. Visual analysis of topical evolution in unstructured text: design and evaluation of topic flow. In: Kazienko P, Chawla N, eds. *Applications of Social Media and Social Network Analysis*. Cham, Switzerland: Springer; 2015: 159–176.
31. Smoot ME, Ono K, Ruschinski J, et al. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* 2011; 27 (3): 431–2.
32. Shortliffe EH, Cimino JJ. *Biomedical Informatics: Computer Applications in Health Care and Biomedicine*. 4th ed. London, UK: Springer; 2014.
33. Buntin MB, Jain SH, Blumenthal D. Health information technology: laying the infrastructure for national health reform. *Health Aff* 2010; 29 (6): 1214–9.
34. Selby JV, Beal AC, Frank L. The patient-centered outcomes research institute (PCORI) national priorities for research and initial research agenda. *JAMA* 2012; 307 (15): 1583–4.
35. Dutton RP, DuKatz A. Quality improvement using automated data sources: the anesthesia quality institute. *Anesthesiol Clin* 2011; 29 (3): 439–54.
36. Hamlin RJ, Sprung J, Hofer RE, et al. Obesity trends in the surgical population at a large academic center: a comparison between 1989-1991 to 2006-2008 epochs. *Acta Chir Belg* 2013; 113 (6): 397–400.
37. Ng M, Fleming T, Robinson M, et al. Global, regional and national prevalence of overweight and obesity in children and adults 1980-2013: a systematic analysis. *Lancet* 2014; 384 (9945): 766–81.
38. Gretarsson B, O'Donovan J, Bostandjiev S, et al. TopicNets: visual analysis of large text corpora with topic modeling. *ACM Trans Intell Syst Technol* 2012; 3 (2): 1–26.
39. Chaney AJ-B, Blei DM. Visualizing topic models. In: Ellison NB, Shanahan JG, Tufekci Z, eds. *Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media*. Palo Alto, CA: AAAI Press; 2012: 419–422.
40. Hinneburg A, Preiss R, Schröder R. TopicExplorer: exploring document collections with topic models. In: Flach PA, Bie TD, Cristianini N, eds. *Machine Learning and Knowledge Discovery in Databases*. Berlin, Germany: Springer; 2012: 838–841.
41. Snyder J, Knowles R, Dredze M, et al. Topic models and metadata for visualizing text corpora. In: Dyer C, Higgins D, eds. *Human Language Technologies: The 2013 Annual Conference of the North American Chapter of the Association for Computational Linguistics*. Stroudsburg, PA: Association for Computational Linguistics; 2013: 5-9.
42. Chuang J, Manning CD, Heer J. Termite: visualization techniques for assessing textual topic models. In: Tortora G, Levialdi S, Tucci M, eds. *Proceedings of the International Working Conference on Advanced Visual*

- Interfaces*. New York, NY: Association for Computing Machinery; 2012: 74–77.
43. Smith A, Chuang J, Hu Y, *et al.* Concurrent visualization of relationships between words and topics in topic models. In: Chuang J, Green S, Hearst M, *et al.*, eds. *Workshop on Interactive Language Learning, Visualization, and Interfaces*. Stroudsburg, PA: Association for Computational Linguistics; 2014: 79–82.
  44. AlSumait L, Barbará D, Domeniconi C, On-line lda: adaptive topic models for mining text streams with applications to topic detection and tracking. In: Giannotti F, Gunopulos D, Turini F, *et al.*, eds. *Eighth International Conference on Data Mining*. Los Alamitos, CA: IEEE Computer Society; 2008: 3–12.
  45. Gohr A, Hinneburg A, Schult R, *et al.* Topic evolution in a stream of documents. In: Park H, Parthasarathy S, Liu H, *et al.*, eds. *Proceedings of the 2009 SIAM International Conference on Data Mining*. Philadelphia, PA: Society for Industrial and Applied Mathematics; 2009: 859–870.
  46. Blei DM, Lafferty JD. Dynamic topic models. In: Cohen W, Moore A, eds. *Proceedings of the 23rd International Conference on Machine Learning*. New York, NY: Association for Computing Machinery; 2006: 113–120.
  47. Wang C, Blei D, Heckerman D. Continuous time dynamic topic models. In: McAllester D, Myllymaki P, eds. *Proceedings of the Twenty-Fourth Conference on Uncertainty in Artificial Intelligence*. Corvallis, OR: AUAI Press; 2008: 579–586.
  48. Wang X, McCallum A, Topics over time: a non-Markov continuous-time model of topical trends. In: Ungar L, Craven M, Gunopulos D, *et al.*, eds. *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY: Association for Computing Machinery; 2006: 424–433.