

ITGInsight — Discovering and Visualizing Science, Technology and Innovation Information for Generating Competitive Technological Intelligence

Xuefeng Wang¹, Shuo Zhang¹ and Yuqin Liu^{2*}

¹ School of Management and Economics, Beijing Institute of Technology, Beijing, 100081,
China

² School of Journalism and Publication, Beijing Institute of Graphic Communication, Beijing,
102600, China

Abstract. Nowadays, most organizations are facing the challenge of tracking the latest technological developments and identifying technology opportunities or threats of the competitive environment. In this context, intelligence analysis methods have been widely used, and lots of technology intelligence techniques have been embedded into general purpose tools to support the need for extracting valuable information from textual data. However, there was no single tool powerful and flexible enough to incorporate all the key elements (data retrieval, pre-processing, normalization, analysis, visualization, and interpretation) in analysis process or have only in limited form. Therefore, obtaining such intelligence awareness, especially from textual data, remains one difficulty. In this paper, addressing concerns of competitive technological intelligence and remedying the shortcomings mentioned before, ITGInsight has been developed. It presents four key features that are remarkable in respect to other software tools: (a) powerful data preprocessing module; (b) flexible user-defined analysis module; (c) gorgeous data visualization module; and (d) professional automatic interpretation module. Finally, an empirical study for synthetic biology is used to describe ITGInsight in deeper detail. The experiment results show that it is a powerful tool for generating effective competitive technological intelligence, such as profiling

science & technology domain, mapping research front relationships, and discerning overall trends, providing more insightful information to intelligent consumers especially in non/few-expert supported environment.

Keywords: science and technology text mining, multi-dimensional statistics, research front, research trend, competitive technological intelligence.

1 Introduction

Nowadays, organizations are facing the challenge of keeping pace with the latest technological developments and identifying technology opportunities or threats of the competitive environment, especially given the exponential growth of accessible information. Further, facing such technique revolution wave tide of the fast fierce development, timely and relevant information on new technology and its industrialization in both internal and external environments will not only have a direct impact on the competitiveness and development of organizations (Liao, Sun, and Wang, 2003), but also become one of the important forces to boost the economy growth, as well as the main index to decide the integrative competition of a country or territory. With intelligence analysis methods being applied to the industrial or commercial setting, the technical competitiveness of industries and organizations and the ability of rapid response to information improved. However, it is crucial to acknowledge that many organizations are still facing the challenge of interpreting the acquired information, which can be an effective source of intelligence if well analyzed, and therefore detailed guidelines on how to analyze data, generating different types of intelligence information to meet specific needs of intelligence consumers are in urgent need.

According to Porter and Cunningham, CTI also have tremendous innovation potential and involves a wide range of Science, Technology and Innovation (ST&I) information, which address real-world concerns for diverse targets and emphases – e.g. profiling technology domain, mapping research front relationships, discerning overall trends and detecting “who is doing what, when and where” (Porter and Cunningham, 2005). The “what” question is far more challenge, whereas it can be enriched by summarizing the experiences and knowledge of domain experts or extracting topical content, especially the noun phrases/domain terms from related text (Newman, Porter, Newman, Trumbach, and Bolan, 2014). In the process of intelligence activity analysis,

apart from the typical qualitative methods (e.g., Delphi, expert interview, scenario planning), quantitative methods offer an appealing alternative to expert opinion. However, some quantitative methods (e.g., bibliometric, patent analysis) use only simple bibliometric indicators (Milanez et al., 2014), which cannot reflect technology changes in granular details. Therefore, more researchers are making efforts to adopt advanced quantitative methods to solve the problem, such as morphological analysis (Lee et al., 2007; Yoon & Park, 2005; Yoon et al., 2008), TRIZ (Yoon & Kim, 2011; Zhang, Zhou, Porter, & Gomila, 2014), conjoint analysis (Xin et al., 2010; Yoon & Park, 2007), technology roadmapping (Choi et al., 2013; Huang et al., 2014; Lee et al., 2008; Lee et al., 2009; Zhang et al., 2013; Zhang, Zhou, Porter, Gomila, et al., 2014) and text mining, which could enhance the capability of CTI, but also rely heavily on expert opinions. With the rapid development of information and the fragmentation of technology domains, domain experts may become less reliable (Shibata et al., 2008). In some cases, experts' biases and insufficient knowledge may even create inaccurate information to the results. In this context, tech mining techniques emerged, providing information from textual data by combining bibliometrics and text mining techniques. On the basis of this technology, many scholars have carried out research. By using text-mining tools and bibliometric indices, Zhu and Porter (2002) addressed capability to exploit huge volumes of available information, ways to do so very quickly, and informative representations via partially automated processes to generate helpful knowledge from text quickly and graphically. These analytical findings can be tailored to the needs of particular technology managers. Joung and Kim (2017) proposed technical keyword-based analysis of patents to monitor emerging technologies and applied a keyword-based model in contents-based patent analysis. A case study of mechanisms of electron transfer in electrochemical glucose biosensors is given to demonstrate how the proposed method can monitor emerging technologies. Above research results show that the output of tech mining to good effect. It could help us to understand detailed contents and knowledge flows from technology information at some extent (Huang et al., 2016).

Therefore, in this paper, addressing concerns of CTI, remedying the shortcomings mentioned before, ITGInsight has been developed. It is compatible with multiple data sources (e.g. Web of Science, Derwent Innovation, Google, Twitter) and incorporates various techniques, algorithms, and measures for all the steps in intelligence analysis, from data preprocessing to discovering to visualizing. Most importantly, it provides automatic interpretation function for analytical results to assist intelligence consumers

to write report under different technology domains in non/few-expert supported environment. Thus, users only need to add details into the report according to their different demand. To some extent, it can mitigate the drawbacks and make full use of various methods strengths, generating effective CTI, e.g., discovering significant clues about technology front/evolution trend. To the best of our knowledge, even though related work have been booming in academia, ITGInsight, especially, the automatic interpretation function makes substantial contribution for CTI discovering.

The remainder of this paper consists of four sections. We first make a brief literature review to introduce some related works. Then, we apply our software to synthetic biology, providing an overview of ITGInsight's functionality for generating and interpreting CTI, and elaborate on the technical implementation of specific parts of the program. Finally, we present remarks and directions for further study.

2 Related Work

Nowadays, many technology intelligence methods have been embedded into general purpose software to support the need for extracting valuable information from literatures or the other textual data. But for some organizations, it remains difficult to select appropriate data sources and software, because the development of them have been suggested by the need of a specific domain, and much science and technology performed is not documented, thus their applications are limited. Sometimes, researchers even have to use more than one software tools to perform a deeper analysis. Furthermore, most of the technology intelligence software do not have systematic analyses functions and there has been little related research which could give instructions about the methods and techniques embedded in them. Given the problems mentioned above, detailed guidelines on how to analyze data, generating different types of intelligence information to meet specific needs of intelligence consumers are in urgent need. Lee, Mortara, Kerr, Phaal, and Probert (2012) and Liu, Wang, and Lei (2015) investigated the possibilities of conducting various data – mining techniques on literatures, and exploratory analyses such as trend analysis, portfolio analysis or if-then analysis have also been conducted by Lee et al. (2012). They divided the technology intelligence software into four categories: search-only tools, search and summarisation tools, advanced text analysis tools, and advanced literature management tools. In Cobo, Lopez-Herrera,

Herrera-Viedma, and Herrera (2011), they presented an analysis of the features, advantages, and drawbacks of these tools. As a result, they concluded that there was no single tool powerful and flexible enough to incorporate all the key elements (data retrieval, preprocessing, normalization, analysis, visualization, and interpretation) in analysis process (Cobo et al., 2011). That is why we designed ITGInsight, which presents four key features that other software tools either do not have or have only in limited form:

- *Powerful data preprocessing module*: ITGInsight is compatible with multiple data sources and implements a wide range of data preprocessing tools such as natural language processing, duplicate record detection, entity disambiguation (normalization), and network extraction.
- *Flexibly user-defined analysis module*: ITGInsight incorporates various methods, algorithms, and measures for all the steps in intelligence analysis and supports combination application of them. In this way, users can carry out the research analysis according to their specific need.
- *Gorgeous data visualization module*: ITGInsight can easily process tens of thousands of data, and display maps that contain more than 20,000 items. ITGInsight also has functionality for zooming, scrolling, searching, and supporting user defined graph types which facilitates the detailed examination of large maps.
- *Professional automatic interpretation module*: ITGInsight can interpret the analytical results automatically under different technology domains on the basis of a widely used framework. Thus, it can provide general solution patterns and standardized decision support process to intelligence consumers especially in non/few-expert supported environment.

3 ITGInsight

In this section, we choose synthetic biology as our target technology field and provide an overview of ITGinsight's functionality for generating CTI¹, especially the ability from data preprocessing to visualization to interpretation of the results.

¹ For a more extensive discussion of the functionality of ITGInsight, we refer to the ITGInsight manual, which is available at http://cn.itginsight.com/Files/download/itginsight_manual.pdf.

3.1 Data

In this research, we choose synthetic biology as our target technology field and we applied the consolidated search strategy proposed by Shapira, Kwon, and Youtie (2017) to publications recorded in Web of Science for the period 2000 – 2020 in Science Citation Index Expanded. The gross worldwide number of records obtained by applying this search strategy (on 9th May 2020) was 12,725. ITGInsight was then used for record cleaning. After removing duplicate records, including the papers published with the same title, abstract and authors as articles, our synthetic biology publication dataset comprises 12,525 publication records.

3.2 Natural Language Process

For intelligence researchers, data preprocessing is essential for the accuracy and quality of the results in intelligence analysis, because the data acquired by them is often totally unstructured or semi – structured and have some mistakes, which may affect the further analysis. ITGInsight provides automatic duplicate record deletion and entity disambiguation (normalization) function, etc. In addition, natural language processing techniques embedded in ITGInsight could help extract a set of subject words from any structured or unstructured text (e.g., the title and abstract of the publications, Web text data) for generating more intelligence information. In most cases, subject words retrieved in this way are large and “noisy”, making them difficult to understand. Therefore, ITGInsight also provides further processing function, generating better domain terms for achieving CTI. It includes four main steps as described in Table 1:

Table 1. Overview of Natural Language Processing

<i>Step</i>	<i>Description</i>
Word segmentation process	
1	Raw dataset for 12,525 publications – apply Hidden Markov Model for word segmentation
2	Data cleaning – remove common/meaningless words (e.g., a/an, the, what, detailed description, some time, method), extreme words (e.g., occurrence in only one record, word length less than 2)
3	Fuzzy matching – combine words with similar structures based on pattern commonality, such as stemming and text similarity)
4	Sequencing – sort the words according to their C-value, forming the preliminary results
Expert knowledge	

5 Consolidation and modification (preliminary results) – words that indicate the same meaning (especially refer to the same technology) will be merged to improve the integration level, and some weakly correlated words will be removed after consulting the domain experts, forming “Thesaurus Dictionary²”

Domain terms extraction

6 Raw dataset for 12,525 publications – apply Hidden Markov Model and “Thesaurus Dictionary” for domain terms extraction

7 Screening – Obtain the top 30 domain terms according to the C-value (PC-value/TF-IDF/ Frequency).

In Table 2, we selected the top 30 word segmentation outputs and domain terms according to the sequence of C-Value.

Table 2. The top 30 domain terms

No	Word segmentation output	Domain term	No	Word segmentation output	Domain term
1	gene expression	gene expression	16	recombinant protein	recombinant protein
2	metabolic engineering	metabolic engineering	17	cell free system	cell free system
3	artificial cell	artificial gene synthesis	18	saccharomyce cerevisiae	secondary metabolite production
4	mammalian cell	synthetic cell	19	transcription factor	synthetic biology method
5	escherichia coli	synthetic promoter	20	synthetic biology approach	saccharomyce cerevisiae
6	synthetic promoter	mammalian cell	21	gene regulatory network	transcription factor
7	essential gene	essential gene	22	cpg odn	cpg odn
8	synthetic gene	escherichia coli	23	cell free protein synthesis	gene regulatory network
9	degree c	gene circuit	24	system metabolic engineering	cell free protein synthesis

² ITGInsight supports various user-defined dictionaries, and they can be used to intervene the natural language processing without directly modifying the original data.

10	synthetic gene network	synthetic genetic network	25	c. glutamicum	system metabolic engineering
11	biological system	synthetic gene circuit	26	synthetic gene circuit	c. glutamicum
12	nucleic acid	biological system	27	expression level	genetic code
13	system biology	natural product	28	genetic code	biosynthetic pathway
14	genetic circuit	nucleic acid	29	negative auto regulation	membrane protein
15	natural product	system biology	30	biosynthetic pathway	dna sequence

Note. “Synthetic biology” is not included in Table 2.

From Table 2, we can find that there are some differences between the experimental results of “Word segmentation” and “Domain term” (marked as grey). For example, there some meaningless words (e.g., expression level, degree c) in word segmentation outputs. In addition, for the terms of “artificial cell”, “synthetic gene network”, “genetic circuit” and “synthetic biology approach”, we realize entity disambiguation (normalization) in “Domain terms extraction” by using “Thesaurus Dictionary”.

3.3 Finding Research Fronts

Realizing the research fronts in specific fields can not only understand the current development status and future trends, but also provide CTI information to intelligence consumers. In addition, it is also helpful for government to formulate relevant technical policies. We take “bibliographic coupling” analysis as an example to conduct our experiment. Further, we use LinLog layout algorithm³ for clustering analysis, and obtained six prominent clusters as shown in Figure 1.

³ ITGInsight provides various clustering algorithms, such as LinLog, VOSmapping, TSNE.

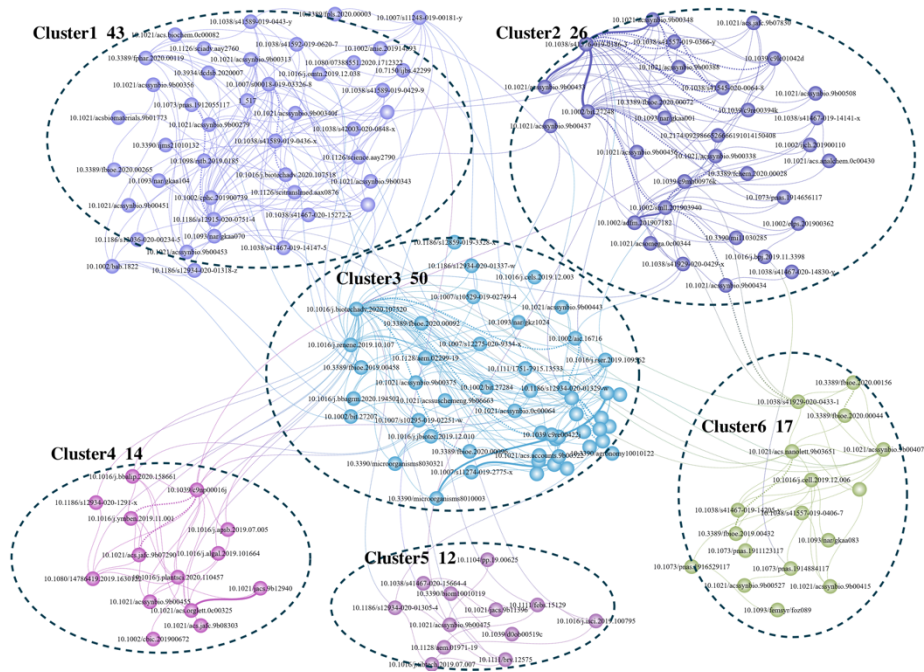


Fig. 1. A 396-node network of bibliographic coupling on synthetic biology (2000-2020) *Note.* The number at the corner of each circle is the number of the articles founded in the cluster

We take Cluster1 as an example and add details to it. Cluster 1 – Synthetic mammalian gene circuits. This cluster contains 43 core documents and the high frequency domain terms of this cluster include: mammalian cell [6], gene circuit [5], synthetic biology method [3], gene expression [3], mathematical model [3], synthetic gene circuit [3], phase separation [3], live cell [3], logic gate[3], etc. An important aspect is to associate domain term identifiers to research fronts, which could help researchers to identify and trace the key innovations and relations that significantly impact a technology’s development, complementing the initial results of bibliographic coupling clustering. To be specific, ITGInsight can display the analysis results about research fronts in seven different ways and provide automatic interpretation of them. We take the topographic map here as an example, in which we can gain more insights into the main advancements in a specific technology through the density and color depth of domain terms (We will discuss the technical implementation of it later on in this paper). In Figure 2, we find that “crispr system” has strong relationship with “novel drug target”, and this has been confirmed in some researched. Fione et al. (2019) performed genome-scale

reveals insights on two levels – first, some broad trends in the field and, second, a host of micro-level translations, and all the results could be provided by the automatic function of ITGInsight. Actually, there are too many specific evolutions to discuss individually. Hence, these next few sections will discuss the two main overarching trends and we add some micro-level examples to confirm them.

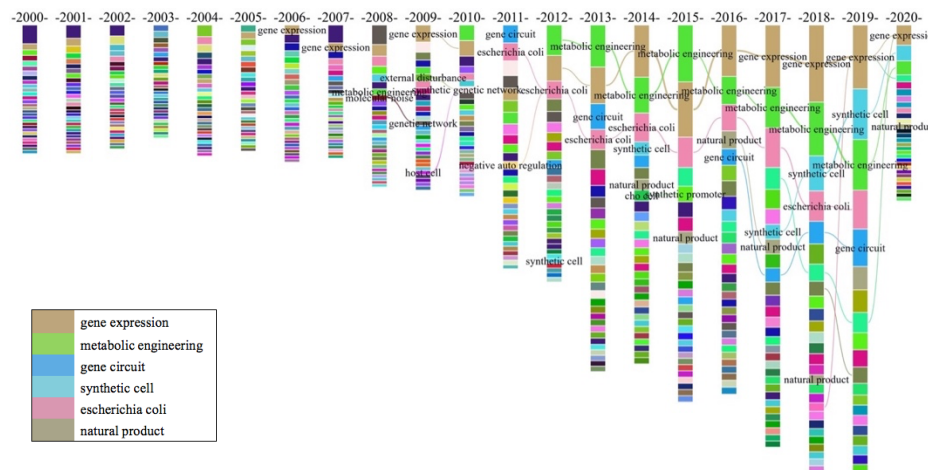


Fig. 3. The evolution of scientific research in synthetic biology

From the macro level, scholars pay more attention to the research topic of "gene expression" (correspond to Cluster 1), "metabolic engineering (correspond to Cluster 2⁴)", "gene circuit (correspond to Cluster 1)", "natural product (correspond to Cluster 6⁵)", "synthetic cell (correspond to Cluster 2)", "escherichia coli", and "saccharomyce cerevisiae", etc., because those domain terms gain more importance at late stage.

From the micro level, we listed two key points in the development of synthetic biology. In 2008, a strong connection between molecular noise and genetic network emerged. Molecular noises in gene networks come from intrinsic fluctuations, transmitted noise from upstream genes, and the global noise affecting all genes. Knowledge of molecular noise filtering in gene networks is crucial to understand the signal processing in gene networks and to design noise-tolerant gene circuits for synthetic biology (Chen, Chang, and Wang, 2008). Further, they find that biochemical regulatory networks suffer

⁴ Cluster 2 – Cell-free protein synthesis.

⁵ Cluster 6 – Genome mining

from process delays, internal parametrical perturbations as well as external disturbances due to the context of host cells. Then the filtering ability of attenuating additive external disturbances is estimated for time-delay biochemical regulatory networks (Chang and Chen, 2010). That is why external disturbance, synthetic genetic network, and host cell appear together in 2009. In 2011, a strong connection between “negative auto regulation” and “Escherichia coli” emerged as an evolution of these previous findings. Dianel et al. (2011) find gene regulation networks are made of recurring regulatory patterns, called network motifs. One of the most common network motifs is negative auto-regulation, in which a transcription factor represses its own production. Negative auto-regulation has several potential functions: it can shorten the response time (time to reach halfway to steady-state), stabilize expression against noise, and linearize the gene’s input – output response curve. This latter function of negative auto-regulation, which increases the range of input signals over which downstream genes respond, has been studied by theory and synthetic gene circuits. To address this, they studied the negative auto-regulation motif in the arabinose utilization system of escherichia coli, in which negative auto-regulation is part of a complex regulatory network.

4 Methods

4.1 Topic clustering process analysis

Technology theme map is one the most effective methods to carry out the technical analysis and technical route tracking. Topographic map, as a new means of technology theme map could help the researchers finding effective research hotspot or research front (Liu et al., 2017). The specific methods are outlined below.

Step1. Construction of co-occurrence matrix

According to the relationship strength between domain terms, we constructed co-occurrence matrix as follows :

$$\text{Corr}_{n \times n} = \begin{bmatrix}
 & \text{Domain term}_1 & \text{Domain term}_2 & \dots & \text{Domain term}_i & \dots & \text{Domain term}_n \\
 \text{Domain term}_1 & r_{11} & r_{12} & \dots & r_{1i} & \dots & r_{1n} \\
 \text{Domain term}_2 & r_{21} & r_{22} & \dots & r_{2i} & \dots & r_{2n} \\
 \dots & \dots & \dots & \dots & \dots & \dots & \dots \\
 \text{Domain term}_i & r_{i1} & r_{i2} & \dots & r_{ii} & \dots & r_{in} \\
 \dots & \dots & \dots & \dots & \dots & \dots & \dots \\
 \text{Domain term}_n & r_{n1} & r_{n2} & \dots & r_{ni} & \dots & r_{nn}
 \end{bmatrix} \quad (1)$$

In this research, relationship strength r_{ij} depends on the number of co-occurrences between domain terms. At the end of this step, we can obtain a set of nodes V (domain terms) and a set of edges E (the relationships between domain terms).

Step2. Coordinate calculation

In order to visualize domain terms, it is necessary to determine the position coordinates of them in plane. In this research, we take LinLog algorithm as an example to give a detailed description about computational process. LinLog is proposed by Noack (2004), in which the repulsion is based on edges repulsing each other instead of node repulsion and is defined as

$$U_{LinLog}(p) = \sum_{(u,v) \in E} \|P_u - P_v\| - \sum_{(u,v) \in V^{(2)}} \deg(u)\deg(v) \|P_u - P_v\| \quad (2)$$

Where $\|p_v - p_u\|$ represents the distance between node u and node v , p_v/p_u is a vector of node positions, $V^{(2)}$ is the set of all subsets of V which have exactly two elements, and $\deg(u)/\deg(v)$ is the node degree. To be specific, the first part of the subtraction in the above equation is a calculation of the attraction between adjacent nodes. The second part is the calculation for the repulsion between the edges. In this way, each node ends up having an influence on the drawing, which is proportional to the number of edges connected to them (the degree). This property of the node can thus be reflected on the graph being drawn by making the size of the node proportional to its degree. At the end of this step, we can obtain the position coordinates of each node (domain terms).

Step3. Plane pixel density function

After obtaining the position coordinates of each node, we use this method proposed by Liu et al. (2017) to determine the color of them in this step. The density function formula of is

$$Density(x, y) = \sum_{i=1}^n f(Number_i)_e - \alpha \left(\frac{\sqrt{(x-x_i)^2 + (y-y_i)^2}}{Distance} \right)^\beta, \alpha > 0, \beta > 0 \quad (3)$$

Where (x_i, y_i) , $i = 1 \dots N$ is the position coordinates of each node, the distance between two different nodes is depend on the average two-dimensional euclidean distance, and $f(Number_i)$ is the standardized value for the number of nodes. The most important part is that we use (x, y) to represent pixel points on the computer screen, after standardizing, it can be connected to RGB color code. By default, ITGInsight uses a blue – yellow – red color scheme. In this color scheme, blue corresponds with the highest item density and red corresponds with the lowest item density. Beyond that, Circle Layout (CR), Evolution Layout (EV), Reference Layout (RF), Weight Spring Layout (SP), No

Weight Spring Layout (UP), Kamada Kawai Layout (KK), Fruchterman Reingold Layout (FR), VOSmapping Layout (VS) and TSNE-Layout (TS) are also embedded in ITGInsight for nodes coordinates computing, and ITGInsight supports the simultaneous use of two or more layout algorithms to provide information from different perspective.

4.2 Topic evolution process analysis

Text mining techniques can reveal rich details on the technical information to provide a picture of how topics emerge as a subject of interest and how they develop over time. The topic evolution analysis function of ITGInsight could help the researchers to trace the emerging and evolving topics. The methods are outlined below. To avoid interference by meaningless terms, researchers could take the Thesaurus-dictionary generated in (NLP) and performed further word segmentation and part of speech tagging on the data set. With these results in hand, document collection is divided into time periods $D_t = \{d_{t1}, d_{t2}, d_{t3}, \dots, d_{tm}\}$ (d: document; m: document number; t: time). Each document corresponds to many terms, $d = \{w_1, w_2, w_3, \dots, w_n\}$ (w: term), and each term represents a single topic. The next step was to construct a co-word network for each time period based on term (topic) co-occurrence. More specifically, if Topic a and Topic b appeared in the same document, they were deemed to have a co-occurrence relationship. Further, the number of times a and b co-occurred in any document was considered a reflection of the strength of the connection and was weighted according to the number of co-occurrences. The differences between each co-word network for each time period reveals how topics have evolved. For instance, some topics emerge in a time period; some disappear. Some gain or lose importance; others merge or split. This paper identifies the following evolutionary relationships:

- **Type1 – Gain/Lose Importance:** the occurrences number of the same topic has been steadily increasing/dropping over the next several periods.
- **Type2 – Emergence:** a new topic is generated suddenly within a certain period;
Death: a topic does not appear again in research time.
- **Type3 – Topic Split:** a new topic is generated from an existing topic (In $T-1$ time period, they appear in the same document; In T time period, they appear in different documents).

- **Type4 – Topic Fusion:** an existing topic fuses with another existing topic (In $T-1$ time period, they appear in different documents; In T time period, they appear in the same document).

5 Conclusions

ITGInsight is an advanced text mining and visualization tool, mainly for generating competitive technological intelligence, ranging from basic statistics to more complicated analyses based on the selected data, that could be used by a wide range of users. It incorporates various techniques, methods, algorithms, and measures for all the steps in intelligence analysis, from data preprocessing to discovering to visualizing. Most importantly, it provides the function of automatic interpretation to assist intelligence consumers to write report under different technology domains on the basis of a widely used framework. Thus, users only need to add details into the report according to their different demand. To some extent, it can mitigate the drawbacks and make full use of various methods strengths, generating effective CTI. The experiment results show that it is a powerful tool for generating effective CTI, such as profiling science & technology domain, mapping research front relationships, and discerning overall trends, providing more insightful information to intelligent consumers especially in non/few-expert supported environment.

Acknowledgments

We are grateful to many scholars and software enthusiasts who provide their valuable opinions and suggestions in the process of ITGInsight design and development. Users could download and install the latest version of ITGInsight from <http://en.itginsight.com/download/>.

Reference

1. Bacchus, W., Lang, M., El-Baba, M. D., Weber, W., Stelling, J., & Fussenegger, M. (2012). Synthetic two-way communication between mammalian cells. *Nature Biotechnology*, 30(10), 991. <https://doi.org/10.1038/nbt.2351>

2. Boyack, K. W., & Klavans, R. (2010). Co-citation analysis, bibliographic coupling, and direct citation: which citation approach represents the research front most accurately? *Journal of the American Society for Information Science & Technology*, 61(12), 2389-2404. <https://doi.org/10.1002/asi.v61:12>
3. Chang, Y. T., & Chen, B. S. (2010). A fuzzy approach for robust reference tracking control of nonlinear distributed parameter time-delayed systems and its biological application. *IEEE International Conference on Systems Man & Cybernetics*. <https://doi.org/10.1109/ICSMC.2010.5642363>
4. Chen, B. S., Chang, Y. T., & Wang, Y. C. (2008). Robust -stabilization design in gene networks under stochastic molecular noises: fuzzy-interpolation approach. *IEEE Transaction on Cybernetics*, 38(1), 25-42. <https://doi.org/10.1109/TSMCB.2007.906975>
5. Choi, S., Kim, H., Yoon, J., Kim, K., Lee, Y.J. (2013). An SAO-based text-mining approach for technology roadmapping using patent information. *R&D Management*, 43(1), 52-74. <https://doi.org/10.1111/j.1467-9310.2012.00702.x>
6. Cobo, M. J., Lopez-Herrera, A. G., Herrera-Viedma, E., & Herrera, F. (2011). Science mapping software tools: review, analysis, and cooperative study among tools. *Journal of the Association for Information Science & Technology*, 62(7), 1382-1402. <https://doi.org/10.1002/asi.21525>
7. Fiona, M. B. et al. (2019). Prioritization of cancer therapeutic targets using CRISPR-Cas9 screens. *Nature*, 568(7753):511-516. <https://doi.org/10.1038/s41586-019-1103-9>
8. Gaber, R., Lebar, T., Majerle, A., Ster, B., Dobnikar, A., Bencina, M., & Jeralam, R. (2014). Designable dna-binding domains enable construction of logic circuits in mammalian cells. *Nature Chemical Biology*, 10(3), 203-208. <https://doi.org/10.1038/nchembio.1433>
9. Huang, Y., Zhang, Y., Ma, J., Porter, A. L., & Guo, Y. (2016). Generating Competitive Technical Intelligence Using Topical Analysis, Patent Citation Analysis, and Term Clumping Analysis. *Anticipating Future Innovation Pathways Through Large Data Analysis* (p:153-172) https://doi.org/10.1007/978-3-319-39056-7_9
10. Joung, J., Kim, K. (2017). Monitoring emerging technologies for technology planning using technical keyword based analysis from patent data. *Technological Forecasting & Social Change*, 114(1), 281 -292. <https://doi.org/10.1016/j.techfore.2016.08.020>
11. Kerr, C. I. V., Mortara, L., Phaal, R., & Probert, D. R. (2006). A conceptual model for technology intelligence. *International Journal of Technology Intelligence and Planning*, 2(1), 73-93. <https://doi.org/10.1504/IJTIP.2006.010511>

12. Kempton, H. R., Goudy, L. E., Love, K. S., & Qi, L. S. (2020). Multiple input sensing and signal integration using a split cas12a system. *Molecular Cell*, 78(1).
<https://doi.org/10.1016/j.molcel.2020.01.016>
13. Kim, H., Bojar, D., & Fussenegger, M. (2019). A crispr/cas9-based central processing unit to program complex logic computation in human cells. *Proceedings of the National Academy of Sciences*.<https://doi.org/10.1073/pnas.1821740116>
14. Liu, Y. Q., Wang, X. F., and Lei, X. P. (2015). Design and Implementation of Academic Relation and Visualization System. *Library and Information Service*, 59 (8), 118-125.
<https://doi.org/10.13266/j.issn.0252-3116.2015.08.015>
15. Lee, C., Seol, H., Park, Y. (2007). Identifying New IT-Based Service Concepts Based on the Technological Strength: A Text Mining and Morphology Analysis Approach. *Fuzzy Systems and Knowledge Discovery (FSKD 2007)*. <https://doi.org/10.1016/j.techfore.2010.11.010>
16. Lee, S., Mortara, L., Kerr, C., Phaal, R., & Probert, D. (2012). Analysis of document-mining techniques and tools for technology intelligence: discovering knowledge from technical documents. *International Journal of Technology Management*, 60(1/2), 130-156.
<https://doi.org/10.1504/IJTM.2012.049102>
17. Liao, S. H., Sun, B. L., & Wang, R. Y. (2003). *A knowledge-based architecture for planning military intelligence, surveillance, and reconnaissance*. *Space Policy*, 19(3), 191-202.
[https://doi.org/10.1016/S0265-9646\(03\)00020-1](https://doi.org/10.1016/S0265-9646(03)00020-1)
18. Martin, V. J. J., Pitera, D. J., Withers, S. T., Newman, J. D., & Keasling, J. D. (2003). Engineering a mevalonate pathway in escherichia coli for production of terpenoids. *Nature Biotechnology*, 21(7),795-802. <https://doi.org/10.1016/j.jnocrsol.2004.08.013>
19. Milanez, D. H., De Faria, L. I. L., Do Amaral, R. M., Leiva, D. R., & Gregolin, José Angelo Rodrigues. (2014). Patents in nanotechnology: an analysis using macro-indicators and forecasting curves. *Scientometrics*, 101(2), 1097-1112. <https://doi.org/10.1007/s11192-014-1244-4>
20. Newman, N. C., Porter, A. L., Newman, D., Trumbach, C. C., & Bolan, S. D. (2014). Comparing methods to extract technical content for technological intelligence. *Journal of Engineering and Technology Management*, 32(4-6), 97-109. <https://doi.org/10.1016/j.jengtecman.2013.09.001>
21. Noack, A. (2004). Visual Clustering of Graphs with Nonuniform Degrees. *Proceedings of the 13th International Symposium on Graph Drawing (GD 2005, Limerick, Ireland, Sep. 12-14)*.

22. Porter, A. L., Cunningham, S. W. (2005). Tech mining: exploiting new technologies for competitive advantage[J]. *Information Processing & Management*, 41(5),1305-1306.
<https://doi.org/10.1002/0471698466.index>
23. Shapira, P., Kwon, S., & Youtie, J. (2017). Tracking the emergence of synthetic biology. *Scientometrics*, 112(3), 1439-1469. <https://doi.org/10.1007/s11192-017-2452-5>
24. Shibata, N., Kajikawa, Y., Takeda, Y., et al. (2009). Comparative study on methods of detecting research fronts using different types of citation. *Journal of the American Society for Information Science & Technology*, 60(3): 571580. <https://doi.org/10.1002/asi.20994>
25. Vincent, B., & Bernadette. (2013). Discipline-building in synthetic biology. *Studies in History & Philosophy of Biological & Biomedical Sciences*, 44(2), 122-129.
<https://doi.org/10.1016/j.shpsc.2013.03.007>
26. Yoon, J., Kim, K. (2011). An automated method for identifying TRIZ evolution trends from patents. *Expert Systems with Application*, 38(12),15540-15548.
<https://doi.org/10.1016/j.eswa.2011.06.005>
27. Yoon, B., Phaal, R., & Probert, D. (2008). Morphology analysis for technology roadmapping: application of text mining. *R&D Management*, 38(1), 51-68.
<https://doi.org/10.1111/j.1467-9310.2007.00493.x>
28. Zhang, Y., Zhou, X., Porter, A. L., Gomila, J. M. V. (2014). How to combine term clumping and technology roadmapping for newly emerging science & technology competitive intelligence: "problem & solution" pattern based semantic TRIZ tool and case study. *Scientometrics*, 101(2),1375-1389. <https://doi.org/10.1007/s11192-014-1262-2>
29. Zhang, Y., Zhou, X., Porter, A. L., Gomila, J. M. V., & Yan, A. (2014). Triple helix innovation in china's dye-sensitized solar cell industry: hybrid methods with semantic triz and technology roadmapping. *Scientometrics*, 99(1), 55-75. <https://doi.org/10.1007/s11192-013-1090-9>
30. Zhu, D., & Porter, A. L. (2002). Automated extraction and visualization of information for technological intelligence. *Technological Forecasting and Social Change*, 69(5), 495-506.
[https://doi.org/10.1016/S0040-1625\(01\)00157-3](https://doi.org/10.1016/S0040-1625(01)00157-3)