Technological Forecasting & Social Change 79 (2012) 1099–1110

Contents lists available at SciVerse ScienceDirect

Technological Forecasting & Social Change

Identifying and visualizing technology evolution: A case study of smart grid technology

Ssu-Han Chen [a], Mu-Hsuan Huang [b], Dar-Zen Chen [a,*]

[a] Department of Mechanical Engineering and Institute of Industrial Engineering, National Taiwan University, No. 1, Sec. 4, Roosevelt Road, 10617 Taipei, Taiwan
[b] Department of Library and Information Science, National Taiwan University, No. 1, Sec. 4, Roosevelt Road, 10617 Taipei, Taiwan

ARTICLE INFO

Article history:
Received 5 May 2011
Received in revised form 4 November 2011
Accepted 27 December 2011
Available online 23 January 2012

Keywords:
Technology evolution
Timeline plot
Community detection
Patent citation network analysis
Smart grid

ABSTRACT

This paper attempts to illustrate the technology evolution for describing the emergence, development, or demise of a technology field. The basic idea is to divide a technology field into tight-knit communities over time and track their inter-year continuity. Then the evolving trajectories are presented through visualizing the timeline plot where each community is drawn as a function of its size, average age, and time. Analyzing a set of patents related to smart grid, we found that this technology consists of several trajectories. Among them, the subjects of network management and e-commerce are relatively young and active. The power system recently has emerged owing to the joining of integration and management concepts. As aging subjects, wireless communication system receives more attention than wired one does. The proposed timeline plot gives insights into evolving trajectories, from which the structure of the technology could be investigated and certain emerging subjects might be figured out. Such understandings are essential information to experts who endeavor to profile technology development and keep up with current trends.

1. Introduction

Technology management (TM) is one of the fastest growing research topics. It is a set of management disciplines that allows organizations to make a better use of their technological capital to maintain their competitive edge. Typical functions used in technology management include technological strategy, planning, forecasting, life cycle management, and road-mapping. Among them, life cycle management, a method to monitor and to track the technology evolution, is regarded as the third-most important function of TM for large companies [1]. It guides organizations to invest, develop, and venture in the desired direction. Thanks to the increasing availability of information resources, supercomputing power, storage capacity, and advanced visualization tools [2], management can be carried out much more rapidly and robustly than in the past.

Social network is a social structure composed of vertices which are tied by specific type of interdependency. Its main goals are to monitor and interpret patterns of social ties among vertices, to visualize how the vertices in a group relate to each other, and to reveal structural characteristics of social groups [3]. This graph-based method is widely applied on various implications due to many kinds of ties between various of vertices. Research in some academic fields has shown that social networks operate on distinct levels and objects, from individuals up to the level of nations; from non-biology to biology, to which researchers succeed in solving their problems. In the field of bibliometrics, the construction of social network is commonly used to trace relationships amongst disciplines, institutions, fields, documents and authors. It has been used to assess the impact of specific articles, authors, and journals; to classify documents in one or more disciplines; to identify interrelationships between authors from different institutions; and to study the homogeneity or heterogeneity or collaboration patterns among different analyzed units. Bibliometrics has gained its importance in understanding the past intellectual knowledge accumulation or interaction and even in forecasting

* Corresponding author. Tel.: +886 2 2366 2723; fax: +886 2 2369 2178.
E-mail addresses: ssuhanchen@ntu.edu.tw (S.-H. Chen), mhhuang@ntu.edu.tw (M.-H. Huang), dzchen@ntu.edu.tw (D.-Z. Chen).

0040-1625/$ – see front matter
doi:10.1016/j.techfore.2011.12.011

its future development, which would be useful for analysts in the process of decision making, science policy, and management [4,5].

To look insight into the technology structure, in general, the analysis results of a bulk of documents are usually presented in a visualized graph called a map. The properties and behaviors of a map can be viewed from graph-, community-, or vertex-level perspectives. Interests are especially highlighted in understanding underlying community-level structures where documents share a passion about a particular topic, and interact to expand their field expertise [6,7]. Early research, however, extracted technology communities and captured their static properties as a snapshot, ignoring the fact that most real-world communities are of a dynamic nature. Recent research has, by and large, trended to use dynamics to explore how technology communities evolve over time [8]. With a window sliding through a sequence of time points, the addition and deletion of documents and citation relationships will cause previous documents of a community to be removed while new ones are gradually added. The critical issue after extracting communities at each snapshot is then to use temporal information to track which technology community at a time has evolved into which community in the next snapshot. Following these links through time allows researchers to create what are referred to as technological community strings. These strings can represent the complex survival, birth, death, branching, or merging patterns of a community transition [9]. It is commonly believed that monitoring technology evolution can be useful for effectively categorizing and tracking the changes of trajectory of the technology dynamics.

In this paper, we describe here a framework for modeling community evolution that successfully addresses the requirements of the dynamic environment and presents the evolving patent citation network. Smart grid technology is an emerging field for next-generation energy delivery and measurement; it serves as an example in this study. The proposed method starts by converting an evolving network into a sequence of static but overlapping snapshots. Next, communities of each snapshot are obtained separately using the Girvan–Newman (GN) algorithm [10], *i.e.* which are called rolling clustering procedure [11]. Two communities across consecutive time points are linked by a string if they share most of their contents. Those strings were then characterized as certain types of transition patterns, enabling new and interesting insights into the characterization of the dynamic behavior of the evolving networks [12]. Finally, evolving community trajectories are visualized by a timeline plot where communities are plotted as a function of size, average age, and time. It would be more evident and convenient to characterize technology evolution if a timeline plot could be visualized [13]. The benefits of visualization have recently been of much interest for knowledge discovery and tracking. The evolution of tight-knit communities within a particular technology helps one to make useful inferences about its development profile. It can further offer an intellectual basis for understanding of the "selection environment" [14] such that research and development (R&D) managers or policy makers will find profitable innovations to undertake.

The rest of this paper is organized as follows. In Section 2, we review related research. In Section 3, we describe and explain our research methodology of this study. In Section 4, we depict the experimental environment and present the results. The concluding remarks and further suggestions are discussed in Section 5.

2. Literature review

2.1. Bibliometric analysis on technology evolution

Bibliometrics has been used to analyze academic publications or patent citations. One of the essential issues among current bibliometric studies would be how to implement a convincing tool to identify the progresses of technologies and have been undertaken over the years. Dynamics is the crucial element on depicting the evolvement of technology across a certain time span. It is also valuable for researchers or enterprises to analyze the content of the documents and identify dominating topics so as to keep update with the recently technology development. Important prior arts regarding technology evolution have been presented as follows:

Chen et al. [13] constructed a static map by integrating principal component analysis and pathfinder network for documents so that viewers could focus on the citation changes in communities annually. Boyack et al. [15] used a force-directed placement algorithm to place similar documents together which were clustered to form mountains on a terrain. If the full time data was limited to consecutive time periods, the growth, reduction, and shifts among the terrains can be analyzed. However, the dynamics of a network is not well understood owing to the pursuit of a stable reference framework for tracking. Viewers can only observe the growth and reduction of communities rather than the phenomena of branching or merging. Morris et al. [16] suggested the self-organizational map (SOM) to identify interesting communities on the map manually. The map comprised a horizontal set of groups graphed along the time axis. However, the work is not actually dynamic in nature. It still does not reveal the condition of branching or merging. Morris et al. [17] provided a timeline plot where documents were plotted on the *x*-axis in terms of their issue date and their cluster leaf position on the dendrogram on the *y*-axis. In their method, technology branching can potentially be examined through the arrows indicating information flow. However, the dilemma about network dynamics is still not adequately addressed. Huang et al. [18] adopted the multi-level SOM to generate content maps of several time periods. In their method, users can observe some general trends by comparing the dominating regions of maps in different periods. However, it is a difficult task to ascertain the topic evolution based on a series of maps. Blei and Lafferty [19] developed a dynamic topic model and used the posterior analysis to capture different scientific themes for the trends of word usage inspection. However, this method does not reveal the technology branching or merging. With the efforts of Small [20], Small and Upham [9], and Kandylas et al. [11], the sliding window was introduced for modeling the dynamic environment of a citation network. After a clustering was carried out in a snapshot, dynamic change was engaged by sliding the window over time, and then strings were linked between communities in successive snapshots if they shared common documents. This concept is quite attractive except for the ordination of the community evolving trajectories. Recently, Shibata et al.

[21,22] employed Girvan–Newman (GN) clustering algorithm on largest connected component (LCC) of the network in each snapshot, then handled dynamic change in a manner similar to Small [20]. They provided a plot of topic evolution as well as a novel plot where communities are located in terms of their average age against the time. Average age can effectively point out the technology which is under the newly breakthrough. Distinct from previous studies described above, their works ought not to specify the number of clusters due to the GN algorithm is a parameter-free algorithm. Unfortunately, their plots are not well integrated. And some emerging communities may not link to the LCC and are thus excluded. Finally, only arrival rather than departure of documents occurred because an incremental window was used. Based on the research cited above, some prerequisites for modeling evolving communities can be defined:

(1) A 2D visualization plot is preferable for human eyes.
(2) Dynamics of arrival and departure of documents at each snapshot can be simulated by creating a sliding window.
(3) Only the inherent structure of the entire citation network should be analyzed to determine how it is partitioned by a natural clustering algorithm rather than by human judgment.
(4) Patterns of community transition should be thoroughly considered.
(5) Ordinations of community evolving trajectories should be meaningful.

With these prerequisites in mind, we describe here a framework in subsequent section for modeling community evolution that successfully addresses the requirements of the dynamic environment and presents the evolving patent citation network.

2.2. Girvan–Newman algorithm

Some actual networks, such as the citation networks, present modular structure in which these networks are formed by communities of vertices. Ties found between vertices inside the same community are more common; ties between vertices of different communities are less common [23]. Community detection in citation networks is a subject worthy of careful study because documents belonging to the same community are more likely to share properties and dynamics in common. As mentioned before, many methods had been applied to detect communities [15–22]. Among them, the single-linkage clustering is subject to chaining effects. The complete-linkage clustering could result in a number of singletons. The SOFM clustering or dynamic topic model requires “a priori” based on the analyst's perception of the optimal partition. Hence the Girvan–Newman (GN) clustering [10] was used in this study to alleviate the problems concerned above for it is a model to cluster networks with common characteristics based on modularity. Besides, GN clustering has two major strengths: first, it is a topological clustering method and therefore suitable for analyzing data and presenting network structure; second, it is parameter-free, which means that it does not involve human judgment to set a priori for the number of clusters. The Girvan–Newman (GN) algorithm is implemented as follows. Given a patent citation network which is composed of N vertices and M ties. This network can be represented by an $N \times N$ adjacency matrix **A** with elements

$$\mathrm{a}_{ij} = \begin{cases} 1 \text{ if } i \text{ and } j \text{ are connected.} \\ 0 \text{ otherwise.} \end{cases} \tag{1}$$

where $i, j = 1, 2, \ldots, N$. Betweennesses of ties are first calculated. The tie that has the highest betweenness is removed from the network. The betweennesses of all ties on the remaining network are then recalculated. The modularity (Q), a measurement to represent the goodness community division, is evaluated by

$$\mathrm{Q} = \frac{1}{2M} \sum_{ij} \left(a_{ij} - \frac{\mathrm{n}_i \mathrm{n}_j}{2M} \right) \delta\left(\mathrm{p}_i, \mathrm{p}_j\right) \tag{2}$$

where $\delta(\mathrm{p}_i, \mathrm{p}_j)$ is 1 if patents i and j fall in the same community and 0 otherwise; n_i is the number of references to patent i. This process is repeated until ties are totally removed. After the iterations stop, the succession of split networks competes with each other in terms of its own modularity. The result with highest value of the modularity is adopted because it has the best split structure [10]. In general, a good partition of a network into clusters means there are many within-cluster ties and minimal between-cluster ties [24]. In the case of patent analysis, patents related to the same subject are typically arranged in the same community, facilitating subsequent procedure of topic identification.

2.3. Natural language processing

After the clustering procedure was carried out, one important step to help analysts interpret the results is to generate a thematic topic for each community. This could be a manual task of summarization [25]. For each community, the topic can be detected by identifying the characteristic terms using natural language processing (NLP) [21]. First of all, the document titles and abstracts are collected for each community, called a corpus. Second, the keywords of the corpus are extracted through a cleaning process by lower case conversion, multiple whitespace stripping, and stop-word elimination, and then a keyword hierarchy is generated. Finally, these terms are weighted using a weight scheme to measure the importance of each term in a certain community. Such automatic procedure paves the way for identifying the topic of a community as well as understanding its evolution dynamically.

2.4. Community transition

A community transition at a given time is a change experienced by a community that has been discovered at an earlier time [26]. Community transitions may survive by recruiting or losing members; two or more communities may merge or a community may split; old communities may die and new ones may be born [27]. Many solutions have been discussed. Palla et al. [27] proposed a joint graph method to do matching. Such a method can reveal only growing, merging or unchanging transitions because any community from successive snapshots is contained in exactly one community in the joint graph. Aynaud and Guillaume [28] adopted the Kuhn–Munkres (KM) algorithm to do intelligible matching. However, the KM matching is a one-to-one association; that is, there are no branching transitions, no merging, and no birth of new communities. Spiliopoulou et al. [26] matched communities across two snapshots: if the percentage of intersection is significantly high, different kinds of transitions could therefore be identified through the introduction to many rules. In this paper, the idea of matching intersection percentage of communities is adopted due to a many-to-many association where the transition patterns could be thoroughly considered.

3. Methodology

For a long time, patents have been acknowledged to have rich information for assessing R&D management and technique diffusion. Patent citation analysis is a widely used method for conducting advanced analysis of technological change. The reason is that a technology will be expanded, enhanced, or enriched as new patents are issued to technology researchers or developers, whose knowledge is built on original patents [29]. So patent citation analysis over time may capture useful information about the evolution of a technology as it follows its growing pattern [30]. Patent documents related to smart grid technology were collected and then arranged in increasing order by issue date. Analysis of this dataset follows four steps. First, the length and sliding step of the window were determined to divide the whole dataset into a sequence of overlapping snapshots. Second, a patent citation matrix for each snapshot was independently constructed and then clustered into communities. Third, strings were formed across communities between consecutive time points for qualified pairs. A timeline plot of the technology evolution was finally developed. The detailed process of the proposed method is described and explained below.

3.1. Sliding window

The proposed framework not only aggregates the patents but also discards them after a defined period. It provides a dynamic view of the patent citation network by splitting it into time slices that are defined by a window [31]. Generally, the length of the window depends on the time lag between a set of current patents and their prior art. Such an interval can measure how long the granted patents impacted on the technology or received of scientific information [32]. The technology cycle time (TCT) of a specific domain is one of the well-known indicators used to measure the pace of technological progress or change [33]. Let T_i be the issue year of patent *i*. Then the formula for calculating TCT is

$$\mathrm{TCT} = \left[\frac{\sum_{i=1}^{N} \mathrm{Median}_{j=1}^{\mathrm{r}_i} \left| \mathrm{T}_j - \mathrm{T}_i \right|}{N} + 0.5 \right] \tag{3}$$

where patent *j* cites patent *i* for $j=1, 2, \ldots, \mathrm{r}_i$ where r_i is the number of patents cite the patent *i*. *N* denotes the number of patents. The $|X|$ is the absolute value of *X*. The notation of the enclosing square brackets, $[Y]$, indicates the Gaussian function. The value *Y* will be rounded to the nearest integer. The TCT represents the average value of median age gaps between the subject patent and other cited patents within a technology domain. The TCT varies for different technologies: a shorter TCT value reflects faster-moving technologies, while a longer TCT value indicates slower-moving technologies [4].

Furthermore, we decided that the defined window should be in an overlapping mode, *i.e.* the next window partially overlaps with the preceding one as for a moving average [34]. The rationales of using shorter time periods and temporal overlapping is to ensure some consistency in community composition over time [35], to allow new communities to emerge and existing communities to merge, split or die away [11], to show evolution [36], to identify new members as a sign of an emerging trend, and to provide a mean for detecting the shift of research focus [37]. Following the rule of thumb of Moody et al. [38], a sliding window that overlaps in incremental shifts of a fifth of window length was considered most appropriate. Moody et al. [38] suggested that such sliding window step width could best capture the fluidity of interaction patterns and simultaneously reduce the fluctuation across snapshots so that some continuity is preserved. Thus, we reused the patents and citations involved in the last $(\mathrm{TCT}-[\mathrm{TCT}/5+0.5])$ years of the prior window for the next one. It is noteworthy that if the value of TCT of a technology field is less than or equal to 2 years, the criterion of Moody et al. [38] does not work. Scaling up the resolution into half yearly, quarterly, or even monthly are potential solutions. A series of large-scale networks would be necessary to analyze such an extremely rapidly progressing environment.

3.2. Community detection and identification

In patent citation network of this paper, vertices represent individual patents and ties between two vertices represent direct citations. Longitudinal data were used to create a linear window-by-window relationship among patents, thus enabling a time series analysis. If networks are created on a specific cut-off year, then the patents issued from year $(\mathrm{y}-\mathrm{TCT}+1)$ to year y are adopted. A small proportion of isolated patents did not contain references and were therefore excluded. In network analysis, communities are detected using the GN algorithm. After a clustering procedure is carried out in one time period, two measurements, based on size and average age, are evaluated for the communities. The first index, size, reflects how influential the scope of a topic is in the current technology environment. The larger the cluster size, the more attention (and likely funding) it has received. The size of each community can be normalized by dividing by the total number of patents in the network. A dominant community threshold $\alpha \in [0, 1]$ is set to filter out weak communities having relatively few patents involved. Another index, the average age, reflects the age of a community. In a rapidly emerging community, patents related to that community are recent and in many cases zero. In contrast, an old community usually indicates that the patents belong to a stagnating topic or are out of contact with the main stream.

Furthermore, technology topics are detected by identifying the characteristic terms for each community using NLP. After patents' titles and abstracts of each community are aggregated as the corpus, keyword extraction process is executed through the following stages: first, conversion of lower case, and then stripping of multiple whitespace; next, elimination of stop-words (such as 'a', 'been', 'the' and so on), and finally, generation of 4-level keyword hierarchy. In such hierarchy, a keyword at a higher level is semantically more general than a keyword at a lower level. The phrase was truncated at the fourth level for the term frequency was hardly more than one in the fifth level or more. In each community, if a specific key phrase is found, the corresponding vector field is assigned by the times of its occurrence. Then a weighting scheme, named *tf-icf* (term frequency-inverse community frequency), is adopted here to measure the frequency and distinctiveness of each term in a certain community [21]. The *tf-icf* value of term *k* in community *s* is given by

$$\mathrm{tf-icf}_{k,s} = \mathrm{tf}_{k,s} \times \log\left(\frac{S}{\mathrm{cf}_k}\right) \tag{4}$$

where $tf_{k,s}$ is the number of occurrences of term *k* in community *s*, cf_k is the number of communities containing term *k*, and *S* is the total number of communities generated by the GN algorithm. In this paper, the terms associated with the top *tf-icf* values in each community are regarded as its characteristic terms.

3.3. Community transitions

As what have been stated in Section 2.4, the idea of Spiliopoulou et al. [26] is followed by this study because it takes into account the transition patterns thoroughly. Let ζ_y be a set of communities discovered at year *y* and $C \in \zeta_y$ and $C' \in \zeta_{y+[TCT/5+0.5]}$ be two communities of two successive snapshots that we wish to compare. The overlap of C with C′ is

$$\mathrm{overlap}\left(C, C'\right) = \frac{\left|C \cap C'\right|}{|C|} \tag{5}$$

where |C| is the number of patents in a community. A high overlap occurs when two communities have many patents in common. Further, a strong string threshold $\beta \in [0, 1]$ is set to determine whether C′ is a match to C. The strong string threshold captures the tolerance to member fluctuation. Note that if β is restricted to the interval [0.5, 1], a community is at most a match (*i.e.* branching transitions will never happen). With the proper setting of the β value, those strings can decide complex community transitions, comprising survival, birth, death, branching, or merging patterns.

3.4. Technology evolution plot

Technology evolution is visualized in a timeline plot in which technological communities are drawn as a function of their size and average age against time. The communities are plotted two-dimensionally according to the analytical time points of the sliding window and average age. The number of patents in each cluster is represented by the size of a circle. Refer to Fig. 1, the drafting procedure can be done by filtering out the weak communities, linking the community strings among consecutive time points, sweeping off insignificant strings, and then identifying trajectories. In such timeline plot, a technology trajectory is an isolated connected component and it is composed of at least two successive year fronts linked by string which shows the evolution of a technology front over time. These trajectories are all marked in color in order to present more clearly. Finally, a pie chart is superimposed on the circle where the arc length of each sector is proportional to the quantity of inheritable, non-inheritable, and immediate patents. An inheritable patent indicates a pre-existing patent inherited from one or more dominant communities at a previous time point; a non-inheritable patent indicates a pre-existing patent without involving an inherent relationship; an immediate patent indicates a newly joining patent that belongs to a dominant community.

4. Results and discussions

We appreciate the claims of Eisenhardt [39] and Yin [40] that case studies can involve either single or multiple cases. In order to demonstrate the feasibility of the research methodology and realize a given widely concerned technology deeply, a case, smart grid technology, was chosen in this paper. Smart grid technology has raised great expectations concerning environmental protection, energy saving, and reduced carbon emissions. The rapid development of smart grid technology is attracting both inventors and funding. It aims to deliver and monitor electricity consumption using multi-directional technologies that allocate and meter

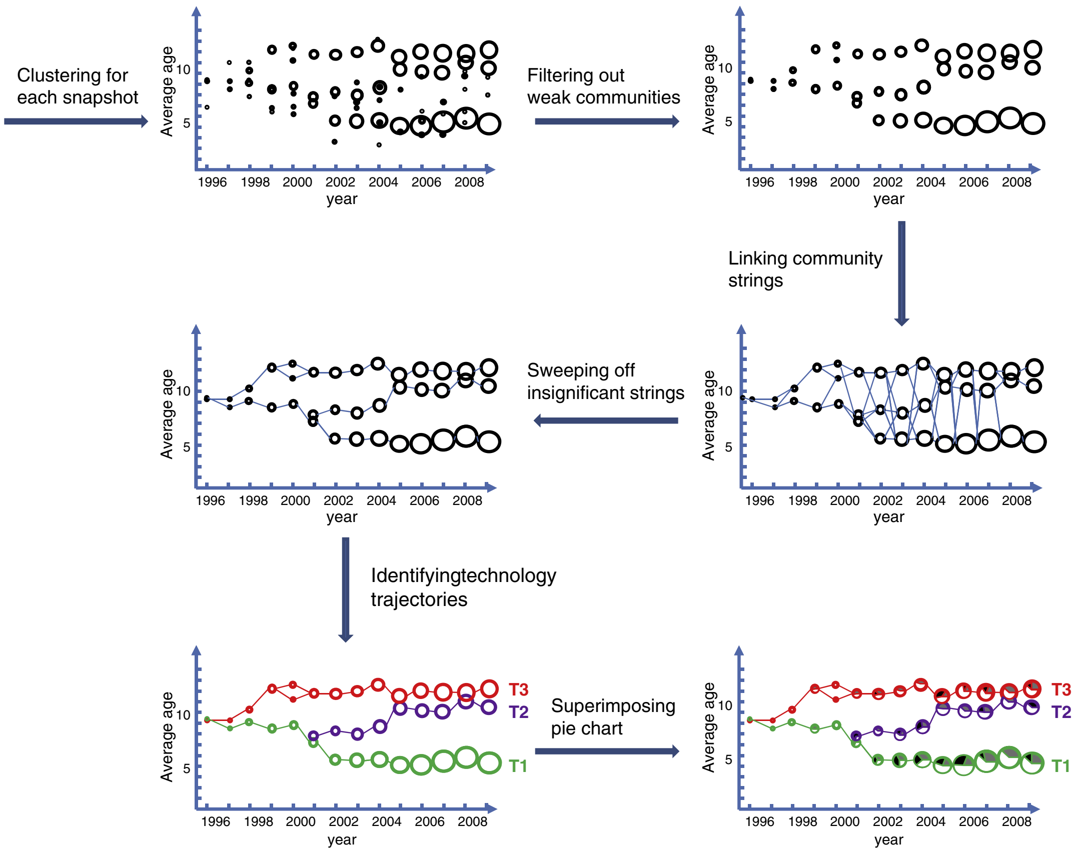

Fig. 1. The drafting procedure of the timeline plot.

power flows dynamically to ensure efficiency, savings, and reliability [41]. Having always been utilized in R & D project management to assess competitive position and to avoid infringement [42], patents were treated as the document source for analysis in this paper. The United States Patent Classification (USPC) categories are used to indicate different patent technology fields. Our focused technology is represented by partial of current USPC main class 307, 340, 370, 455, 700, 702, and 709. For the dates range from 1995 to 2009, 15,948 patent documents were retrieved from the database of the United States Patent & Trademark Office (USPTO). Not only attribute data (*i.e.* issue date, title, and abstract), but relational data (*i.e.* reference list), for the selected patents were also recorded. As shown in Fig. 2, the number of patents started to grow dramatically after 1997 and the rate of growth is increasing. Currently, there are more than 2000 patents annually on this subject matter.

After collecting the patent documentation, the next step was the analysis. This was done using a self-programming toolkit under the 'R' environment with the igraph, tm, RWeka, stringr, and plotrix packages (http://cran.r-project.org/). To decide the length of the sliding window, we calculated the average time lag of the patent inventions upon which a new invention was based: this yielded a TCT value of 5. It implicates that the smart grid technology is a fast-moving technology [4]. Using the criterion of a fifth of window length established the step width of a sliding window as 1 year. The 5-year period window is not too big to obscure the dynamics of community evolution; it is not too small either that the patents and the citations in each period were too scarce to be segmented appropriately [11]. After the length and step size of window was specified, all patents and citations that occurred in this window were aggregated into a patent citation network for each time point. Related patents were assembled as communities through a GN clustering operation, identifying dominant communities for subsequent size and average age calculation and topic detection. The problem of identifying dominant communities is transformed into generating an α value to filter out the communities with relatively small size. We stacked the normalized size of communities of all the networks and produced a histogram. According to our experience, the shape of histogram usually follows an exponential distribution as shown in Fig. 3. The shape of the distribution is primarily related to the GN algorithm which potentially provides many extremely weak communities as well as few strong ones. Rosin thresholding from image processing [43] was used in our method to automatically determine a corner point of the exponential histogram for preserving dominant communities [44]. In this study, the corner point of histogram is marked by a dotted line as shown in Fig. 3, which means that α value was set to 0.036. Only the communities with normalized size larger than α would be discussed. Moreover, to track community continuity, we linked these objects over time if their associated inter-year community overlap was larger than the threshold β. To do so, we stacked all the inter-year community overlaps to produce a histogram. According to our experiences, the shape of histogram usually follows a bimodal

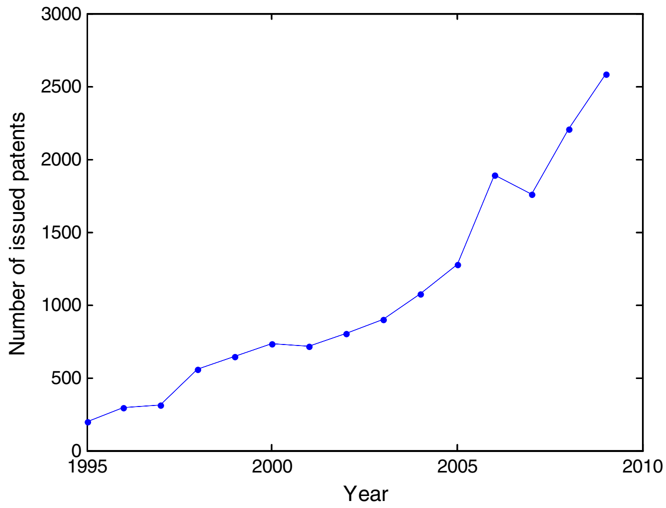

Fig. 2. Trend of smart grid technology according to issued patents.

distribution as shown in Fig. 4. The shape of this distribution indicates a crisp continuity of inter-year communities. Otsu thresholding originating from image processing [45] was used in our method to automatically determine a demarcation of the bimodal histogram for removing the weak strings. In this study, the demarcation of histogram is marked by a dotted line as shown in Fig. 4, which means the β value was set to 0. 411. Only the strings with the overlap larger than β would be drawn. Essentially, what we did above was to chain a series of communities with similar topic and make manifest continuity while presenting an evolving trajectory of communities. Using the information of dominant communities and strong strings, the evolution plot of smart grid technology over the 11 periods from 1999 to 2009 is shown in Fig. 5. The plot is chronological from left to right. The size of the pie chart is proportional to the number of patents in the community and the arc length of each sector is proportional to the quantity of inheritable (white), non-inheritable (gray), and immediate patents (black). Since a single characteristic term may represent only one aspect of technology topics, a set of the characteristic terms are more capable to point out the precise view of them. In other words, the topics can be regarded as the assembly of individual characteristic terms [46]. Thus, the top ten characteristic terms of each community in a trajectory were collected. To break down root topics that potentially contribute to a particular technology domain, characteristic terms were aggregated into subjects manually over time. Users would gain insight into what happened historically by cross-referencing the timeline plot and the corresponding subject evolution diagram of a trajectory.

As shown in Fig. 5, smart grid technology is composed of six evolving trajectories, marked T1 to T6, which are arranged by the order of average age of the terminal communities. Inspecting the top ten characteristic terms of the communities of each trajectory, smart grid system configuration is clearly a combination of the individual components. It basically encompasses e-commerce (T1), network management (T2), wireless communication system (T3), pulse width modulation (pwm) and reactive power

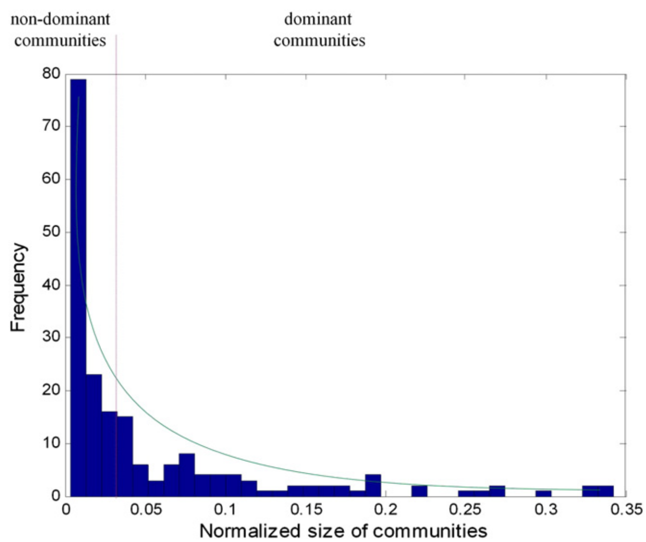

Fig. 3. Histogram of annually normalized size of communities.

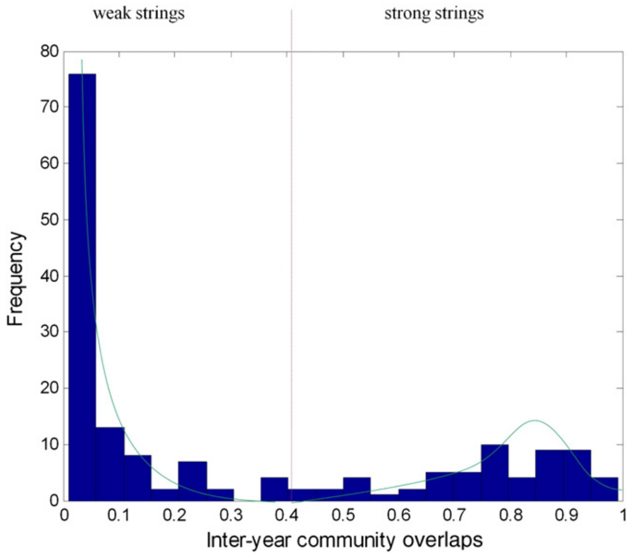

Fig. 4. Histogram of inter-year community overlaps.

compensation (T4), power system (T5), and modem and wired communication system (T6). A smart grid system operates in several steps. First, the energy consumption data are measured with meters. The wired/wireless communication system transfers data from the meters to the servers. The central system manages all the customer transactions and releases real-time information regarding market price and promotion plans on a website. Energy trading or exchanging is also possible if customers have surplus energy [47]. Additionally, as shown in Figs. 6–10, we summarized characteristic terms of each trajectory in a subject evolution diagram over time manually for further analyzing. Diagram for the trajectory T4 was excluded because its prolonged period is too short to make sense.

In Fig. 5, trajectory T1 represents e-commerce in smart grid technology, which provides a web portal for the market where customers can buy (sell) electricity from (to) different companies through different plans. In Figs. 5 and 6, the exchange, catalog, matching, auction, payment, and purchase systems have been developing for the past decade in order to enrich the functionality, private, or security of a trading and transactions platform. Then as the diversity of commodities increased, commodity management and e-commerce received more attentions, hailing the advent of adverting system. However, T1 community's age rose dramatically since 2002 as it came to maturity with fairly few related patents joining in T1 nearly vanished in 2005. Since 2005, the ideas of dynamic pricing developed significantly to support demand response to compensate intermittent production, activating e-commerce in smart grid technology. Recently, a graphical interface for displaying information in an electronic trading environment has been an intellectual subject of much interest. And the issue of purchasing security is being contemplated once again.

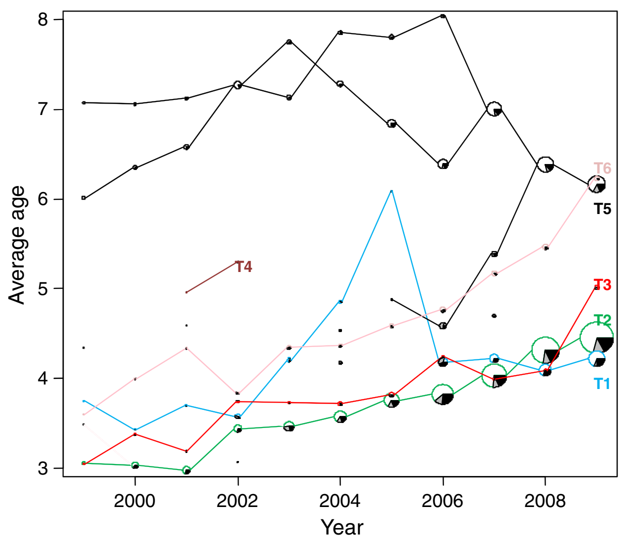

Fig. 5. Technology evolution plot of smart grid technology (□: inheritable patent; ■: non-inheritable patent; ■: immediate patent).

Fig. 6. Subject evolution diagram of T1 (e-commerce).

Trajectory T2 represents network management in smart grid technology. For long, managers increasingly question the need to retain application separation. The primary responsibility of network management is to develop a framework that includes protocols and model standards for information communication, collection, analysis, and management to achieve interoperability of smart grid devices and systems [48]. The network supporting smart grids will be very complex and will handle large volumes of information. In Figs. 5 and 7, the management aspects firstly focused on local area network (LAN), followed by outage management, and then distribution management. Generally, managing or integrating the whole system is always the most dominant and most active feature of smart grid technology. A great deal of recent concern has focused on wireless communication management.

Trajectory T3 represents the wireless communication system, whereas T6 represents the modem and wired communication system. Traditionally, wired communication systems offered a reliable way for data transmission. However, with rises in wages and costs of metals, the deployment costs for cables, trenching tunnels, and maintenance, costs have risen sharply. As the requirement of distribution automation continues to increase, wireless technologies have stepped up to the plate [49]. As shown in Fig. 5, it is obvious why trajectory T3 is uniformly younger than T6: wireless is more recent than wired communication in the past decade. We also observe that the slope of trajectory T6 is greater than that of T3. This could potentially be explained by the fact that R&D staffs have paid more attention to wireless technology. In Figs. 5 and 8, cellular network technologies are continuously evolving to meet the increasing demands for wireless services at mega-data level. Global system for mobile communications (GSM), general packet radio service (GPRS), universal mobile telecommunication system (UMTS), and wireless interoperability for microwave access (WiMAX) have been widely deployed cellular systems. In the following years, other unique wireless technologies such as ad-hoc network, coordinator-based wireless network, backbone network, and bluetooth are continuously developing to meet different communication requirements. Recently, the frequency-hopping spread spectrum comprises most of the immediate patents that are used to defend against the cyber threat of denial of service and to keep the grid secure. In Figs. 5 and 10, the wired communication systems begin with developing components for data transmission such as pulse-coded modulation (PCM) modem, receiver, line probing, loop carrier, modulator–demodulator, digital subscriber line (DSL), etc. so as to establish wired communication environments. In 2002, minimizing distortions in data communications through linear/non-linear equalizers stirred investors' appetite. Then some components designed for improving modem & DSL architecture were proposed during 2004–2007: the multiple x digital subscriber line (xDSL) modems and multi-carrier technology that provide wide selection of interfaces and flexible incorporating on different types of DSL service to obtain high-bandwidth and point-to-point function over existing infrastructure. The line-isolation circuitry was embedded with a modem to provide a modem interface that allows synchronous modem transmission protocols. Recently, voice over internet protocol (VoIP) is quite popular that uses a high-speed network connection to send signals of users' voice in the formation of packets via the network.

Trajectory T4 represents the pwm inverter and reactive power compensation technique, which is used to increase transmission efficiency through reduced losses, grid congestion, increased transfer capability, and enhanced grid reliability [50]. T4 with only two communities was found for 2001 and none was identified after 2002. Besides birth and death of communities in T4, we also observe merging and splitting. In light of the move toward the trajectory T5, it is generally called the power system and includes the components of electrical distribution, energy storage, power supplies, and their integrated management. It is of relatively vintage as one of smart grid subjects. From Figs. 5 and 9, T5 is a branching trajectory that consists of three sub-trajectories: the first sub-domain is power storage and supplies, which existed in 1999 with an average age of around 6, the second one is related to power distribution, which existed in 1999 with an average age of about 7, and the last one is remote control system, which emerged in 2005. The aspects of power management caused the first two sub-domains to merge in 2002. However, due to the unstable relationship between them in the communities, it then split at the next time point. The ideas about power factor control system joined in 2003 in the first sub-domain. It was, however, irresistible that the relative importance of the sub-trajectories of the first two sub-domains was in a reverse order in 2004 due to many diagnostic and monitoring systems conceptualized in the second sub-domain. Recently, controlling the flow of power to maintain reliable service and stable operation, aspects of power system integration and energy management, were added to the power system in 2008. Until now, both of the two aspects have inspired scientists a lot.

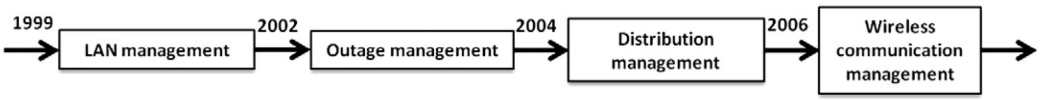

Fig. 7. Subject evolution diagram of T2 (network management).

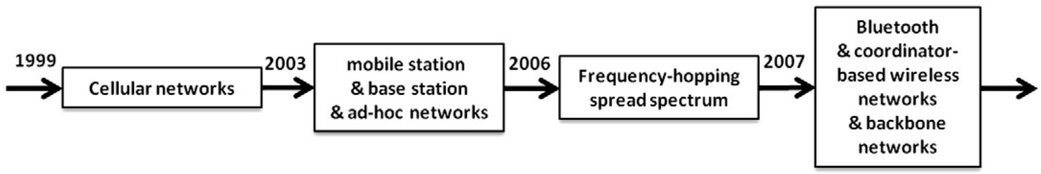

Fig. 8. Subject evolution diagram of T3 (wireless communication system).

5. Conclusions and suggestions for future work

To study technology structure and its evolution, we have borrowed from social network theory and applied bibliometric analysis to the issued patents of USPTO longitudinally. As detailed above, we described a framework for modeling community evolution of the patent citation network with 2D visualization, environmental dynamics, natural clustering, completed transition patterns, and rational trajectories ordination promised. The basic idea was that division of a technology domain into strongly connected communities over time and track of their inter-year continuities can help to realize the technology evolution. Visualization technology improves users' ability to comprehend the results of patent analysis more efficiently. The smart grid technology was chosen as a case study, which is an emergent technology domain in energy issue and to be performed a series of patent citation network analyses at several time points. First, we found that the smart grid technology encompasses several trajectories, including e-commerce, network management, wireless communication system, pwm and reactive power compensation, power system, and the modem and wired communication system. We observed that the combination of the individually components can be regarded as an epitome of the smart grid system configuration; it can be well mapped into service, control and connectivity, and energy planes of smart grid which were summarized by Darmois [48]. This characteristic is quite attractive for users because this visualization can impress him/her to seize the global structure of domain knowledge of smart grid technology more effectively. Second, the proposed methodology provides understanding and insight into technology development by cross-referencing a trajectory on the timeline plot and its corresponding subject evolution diagram. Such process potentially assists policymakers to decide which innovation most deserve to be invested. Among those trajectories, the pwm inverter and reactive power compensation technique are used to be popular but now have become a history topic. The network management is the most dominant subject because smart grids become complex progressively. Recently, it has focused on management of wireless communications for facilitating the integration of different metering protocols, and has added security items into the wireless module. The e-commerce remains the hottest subject that engages to a visual market display and the issue of purchasing security. Especially, the visualization techniques have reduced the large quantities of data into the easily understandable visual formats so that managers can efficiently operate a grid. For households or business establishments, visualization of power consumption is expected to feature with an effect of power-saving. The power system integration has recently emerged owing to the join of management in energy concepts. Remote power control and optimal power distribution planning are mainly concerned. As the aging subjects, the wireless communication system has relatively received more attention from R&D staffs than the wired one has done. Due to the wireless connectivity, quick, and affordable application development and deployment, it has become one of the critical components for infrastructures and has enabled the new solutions and revenue generating services. The proposed timeline plot and subject evolution diagrams could potentially aid the related companies in technology planning, road-mapping, and operations. Such understandings and insights are required to keep abreast of current trends, appropriate subdomains, and strategic timing of development and deployment.

The tie between vertices in this paper indicated a direct citation between patents. Our suggestion for further research is to explore whether such a relationship can be replaced by other types of citations (e.g. bibliography coupling, co-citation, or longitudinal citation), term similarities or their combinations. Using alternative natural clustering algorithms (e.g. Walktrap, spinglass, or label propagation algorithm) rather than the GN method should be also explored. These approaches may enable users to excavate other latent intelligence from different perspectives. Finally, it is evident that additional analytical results will be obtained if users further segment communities of a trajectory into sub-communities and track their evolution.

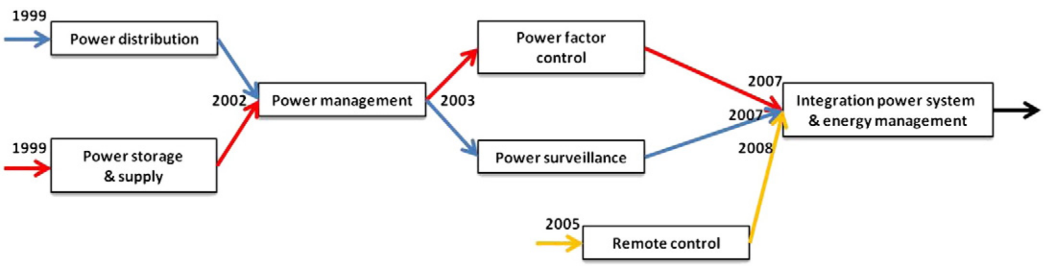

Fig. 9. Subject evolution diagram of T5 (power system). Note that the arrows in red represent the first sub-domain; those in blue represent the second sub-domain; those in yellow represent the third sub-domain; the rest represents their integration.

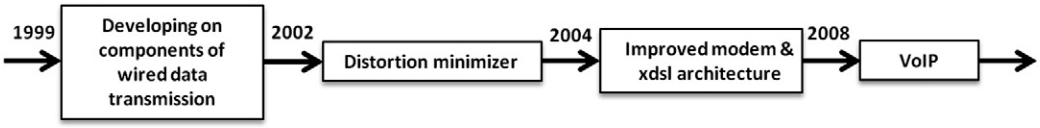

Fig. 10. Subject evolution diagram of T6 (modem and wired communication system).

Acknowledgements

This research is partially supported by the National Science Council, Taiwan, under contract No. NSC 100-3113-P-002-005.

References

[1] H. Kropsu-Vehkapera, H. Haapasalo, J.P. Rusanen, Analysis of technology management functions in Finnish high tech companies, Open Manage. J. 2 (2009) 1–10.
[2] B. Hjorlan, H. Albrechtsen, Toward a new horizon in information science: domain analysis, J. Am. Soc. Inform. Sci. 46 (6) (1995) 400–425.
[3] W. De Nooy, A. Mrvar, V. Batagelj, Exploratory social network analysis with Pajek, Cambridge University Press, 2005.
[4] F. Narin, Technology indicators and corporate strategy, Rev. Bus. 14 (3) (1993) 19–23.
[5] T. Daim, P. Suntharasaj, Technology diffusion: forecasting with bibliometric analysis and Bass model, Foresight 11 (3) (2009) 45–55.
[6] E. Wenger, R. McDermott, W.M. Snyder, Cultivating communities of practice, Harvard Business School Press, 2002.
[7] M.E.J. Newman, Detecting community structure in networks, Eur. Phys. J. B 38 (2) (2004) 321–330.
[8] C. Lee, J. Jeon, Y. Park, Monitoring trends of technological changes based on the dynamic patent lattice: a modified formal concept analysis approach, Technol. Forecast. Soc. Change 78 (4) (2011) 690–702.
[9] H. Small, P. Upham, Citation structure of an emerging research area on the verge of application, Scientometrics 79 (2) (2009) 365–375.
[10] M. Girvan, M.E.J. Newman, Community structure in social and biological networks, Proc. Natl. Acad. Sci. U. S. A. 99 (12) (2002) 7821–7826.
[11] V. Kandylas, S.P. Upham, L.H. Ungar, Analyzing knowledge communities using foreground and background clusters, ACM Transactions on Knowledge Discovery from Data, vol. 4,(2), 2010, Article 7.
[12] K. Naimisha, Applications of social network analysis to community dynamics, Master thesis, Supercomputer Education and Research Centre, Indian Institute of Science, Bangalore, India, 2008.
[13] C. Chen, T. Cribbin, R. Macredie, S. Morar, Visualizing and tracking the growth of competing paradigms: two case studies, J. Am. Soc. Inform. Sci. Tec. 53 (8) (2002) 678–689.
[14] R.R. Nelson, S.G. Winter, In search of a useful theory of innovation, Res. Policy 6 (1) (1977) 36–76.
[15] K.W. Boyack, B.N. Wylie, G.S. Davidson, Domain visualization using VxInsight® for science and technology management, J. Am. Soc. Inform. Sci. Tec. 53 (9) (2002) 764–774.
[16] S. Morris, C. DeYong, Z. Wu, S. Salman, D. Yemenu, DIVA: a visualization system for exploring document databases for technology forecasting, Comput. Ind. Eng. 1 (43) (2002) 841–862.
[17] S. Morris, G. Yen, Z. Wu, B. Asnake, Time line visualization of research fronts, J. Am. Soc. Inform. Sci. Tec. 54 (5) (2003) 413–422.
[18] Z. Huang, H. Chen, A. Yip, G. Ng, F. Guo, Z.K. Chen, M.C. Roco, Longitudinal patent analysis for nanoscale science and engineering: country, institution and technology field, J. Nanopart. Res. 5 (3–4) (2003) 333–363.
[19] D.M. Blei, J.D. Lafferty, Dynamic topic models, Proceedings of the 23rd International Conference on Machine Learning, 2006.
[20] H. Small, Tracking and predicting growth areas in science, Scientometrics 68 (3) (2006) 595–610.
[21] N. Shibata, Y. Kajikawa, Y. Takeda, K. Matsushima, Detecting emerging research fronts based on topological measures in citation networks of scientific publications, Technovation 28 (11) (2008) 758–775.
[22] N. Shibata, Y. Kajikawa, I. Sakata, Extracting the commercialization gap between science and technology — case study of a solar cell, Technol. Forecast. Soc. Change 77 (7) (2010) 1147–1155.
[23] L.F. Costa, F.A. Rodrigues, G. Travieso, P.R. Villas Boas, Characterization of complex networks: a survey of measurements, Adv. Phys. 56 (1) (2007) 167–242.
[24] Y. Kajikawa, Y. Takeda, Citation network analysis of organic LEDs, Technol. Forecast. Soc. Change 76 (8) (2009) 1115–1123.
[25] Y.H. Tseng, C.J. Lin, Y.I. Lin, Text mining techniques for patent analysis, Inform. Process. Manag. 43 (5) (2007) 1216–1247.
[26] M. Spiliopoulou, I. Ntoutsi, Y. Theodoridis, R. Schult, MONIC: modeling and monitoring cluster transitions, International Conference on Knowledge Discovery and Data Mining, 2006, pp. 706–711.
[27] G. Palla, A.L. Barabási, T. Vicsek, Quantifying social group evolution, Nature 446 (2007) 664–667.
[28] T. Aynaud, J.L. Guillaume, Static community detection algorithms for evolving networks, Proceedings of the 8th International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks, 2010, pp. 513–519.
[29] C. Choi, Y. Park, Monitoring the organic structure of technology based on the patent development path, Technol. Forecast. Soc. Change 76 (6) (2009) 754–768.
[30] M.H. Fallah, E. Fishman, R.R. Reilly, Forward patent citations as predictive measures for diffusion of emerging technologies, Portland International Conference on Management of Engineering & Technology, 2009, pp. 420–427.
[31] T. Falkowski, J. Bartelheimer, Applying social network analysis methods to explore community dynamics, in: U. Serdült, V. Täube (Eds.), Applications of Social Network Analysis 2005, Wissenschaftlicher Verlag Berlin, 2008, pp. 189–212.
[32] W. Glänzel, R. Rousseau, A new representation of relative first-citation times, 11th International Conference on Science and Technology Indicators, Leiden, the Netherlands, 2010.
[33] A.A. Kayal, R.C. Waters, An empirical evaluation of the technology cycle time indicator as a measure of the pace of technological progress in superconductor technology, IEEE T. Eng. Manage. 46 (2) (2002) 127–131.
[34] P. Doreian, On the evolution of group and network structure, Soc. Networks 2 (3) (1980) 235–252.
[35] S. Upham, L. Rosenkopf, L.H. Ungar, Positioning knowledge: schools of thought and new knowledge creation, Scientometrics 83 (2) (2010) 555–581.
[36] R. Klavans, K.W. Boyack, Using global mapping to create more accurate document level maps of research fields, J. Am. Soc. Inf. Sci. Tec. 62 (1) (2011) 1–18.
[37] C. Chen, CiteSpace II: detecting and visualizing emerging trends and transient patterns in scientific literature, J. Am. Soc. Inf. Sci. Tec. 57 (3) (2006) 359–377.
[38] J. Moody, D. Farland, S. Bender-deMoll, Dynamic network visualization, Am. J. Sociol. 110 (4) (2005) 1206–1241.
[39] K.M. Eisenhardt, Building theories from case study research, Acad. Manage. Rev. 14 (4) (1989) 532–550.
[40] R.K. Yin, Case study research: design and methods, Sage Publications, Thousand Oaks, CA, 2009.
[41] M. Estefania, ATIS and the smart grid, Global Standards Collaboration, Personal CommunicationRetrieved 10 Dec 2010, from, http://gsc15.ccsa.org.cn/doc_list.php?doc_type=op2010.
[42] T.U. Daim, G. Rueda, H. Martin, P. Gerdsri, Forecasting emerging technologies: use of bibliometrics and patent analysis, Technol. Forecast. Soc. Change 73 (8) (2006) 981–1012.
[43] P.L. Rosin, Unimodal thresholding, Pattern Recogn. 34 (11) (2001) 2083–2096.

[44] D.B. Perng, S.H. Chen, Directional textures auto-inspection using discrete cosine transform, Int. J. Pro. Res. 49 (23) (2011) 7171–7187.
[45] N. Otsu, A threshold selection method from gray-level histograms, IEEE T. Syst. Man Cyb. 9 (1) (1979) 62–66.
[46] R.L. Ohniwa, A. Hibino, K. Takeyasu, Trends in research foci in life science fields over the last 30 years monitored by emerging topics, Scientometrics 85 (1) (2010) 111–127.
[47] P. Wang, J.Y. Huang, Y. Ding, P. Loh, L. Goel, Demand side load management of smart grids using intelligent trading/metering/billing system, Power and Energy Society General Meeting, 2010.
[48] E. Darmois, Smart grids and ICT standards, global standards collaboration Retrieved 10 Dec 2010, from, Personal Communication, 2010http://gsc15.ccsa.org.cn/doc_list.php?doc_type=op.
[49] M.H. van Doorn, Wireless technology for a secure smart grid Retrieved 15 Dec 2010, from, AltEnergy eMagazine, 2010http://www.altenergymag.com/emagazine.php?art_id=1590.
[50] R. O'Neill, Principles for efficient and reliable reactive power supply and consumption, Technology Report Docket Number AD05-1-000, Federal Energy Regulatory Commission, 2005.

Ssu-Han Chen received his Ph. D. degree in Industrial Engineering and Management, National Chiao Tung University, Taiwan, in 2010. Currently, he is a postdoctoral researcher at Intellectual Property Analysis & Innovative Design Laboratory, National Taiwan University, Taiwan. His current research interests include social network analysis, image processing and pattern recognition.

Dar-Zen Chen received his B.S. form National Taiwan University (NTU) and M.S. and Ph. D. in Department of Mechanical Engineering from University of Maryland, College Park, respectively. He served as an assistant Professor in Department of Mechanical Engineering, Cleveland State University in 1992. He is with Department of Mechanical Engineering and Institute of Industrial Engineering NTU, as a professor currently. His research interest includes intellectual property management, patentometrics, competitive analysis, robotics, kinematics and mechanism design.

Mu-Hsuan, Huang received her B.S., M.S. form NTU and Ph. D. in College of Library and Information Science from University of Maryland, College Park, respectively. She Joined Dept. of Library and Information Science NTU in 1992 and was the department Chair from 2001 to 2007. Her research is related to bibliometrics science and technology policy, intellectual property, and patent information. She is the principal investigator of "Performance Ranking of Scientific Papers for World Universities" project sponsored by Higher Education Evaluation & Accreditation Council of Taiwan, HEEACT).