https://doi.org/10.51408/issi2025_036

Exploring Scientist's Research Trajectories within a Field with Main Path Analysis

Chung-Huei Kuan[1], Szu-Chia Lo[2]

[1]*maxkuan@mail.ntust.edu.tw*
National Taiwan University of Science and Technology, Graduate Institute of Patent, Taipei (Taiwan, ROC)

[2]*szuchialo@ntu.edu.tw*
National Taiwan University, Department of Library and Information Science, Taipei (Taiwan, ROC)

Abstract

Main Path Analysis (MPA) is commonly applied to citation networks constructed from papers or patents within a research or technological field to reveal representative knowledge-diffusion trajectories for that field. These trajectories, known as Main Paths (MPs), reflect the overall knowledge development and evolution within the field.

Rather than examining field-level trajectories, this study introduces a novel approach to explore individual scientists' research trajectories within the field. These individual-level trajectories enable analysts to trace the lineage of a scientist's work, understand its origins, and uncover its influence on subsequent research. Additionally, these individual-level trajectories can be contrasted with field-level trajectories to examine their interactions, providing further insights into a scientist's contributions relative to the field's mainstream development.

This approach relies on a previously overlooked path search algorithm in MPA, referred to as key-node search, to generate MPs that capture distinct knowledge flows centered around a scientist's works. A case study based on patents in the field of Evolutionary Computation, using an official artificial intelligence patent dataset, demonstrates both a macro-view and a micro-view of the proposed individual-level MPs.

Introduction

Hummon and Doreian (1989) developed the so-called Main Path Analysis (MPA), which aims to uncover "the mainstream of literature of a clearly delineated area of scientific research" from the citation network of a specific research area. Since its inception, MPA has become a widely recognized method, leading to a proliferation of studies employing it. Its popularity can be attributed to its conceptual simplicity, further bolstered by its availability in the popular network analysis tool Pajek (Batagelj & Mrvar, 1998; De Nooy, Mrvar, & Batagelj, 2018).

MPA is typically employed to analyze networks of mutually citing documents, such as scientific papers or patent publications associated with a specific field of study. In such networks, documents are represented as nodes, while their citations are represented as arcs, denoting pathways for the flow of knowledge from cited documents to citing ones. By applying MPA, one or more series of connected arcs—referred to as main paths (MPs)—are derived from the network and identified as representative trajectories of knowledge development within the field.

Rather than focusing solely on the MPs for a field, or field-level MPs, this study explores whether the same approach can be applied to individual researchers or scientists within the field. Specifically, it investigates whether the most representative trajectories passing through the papers or patents associated with a researcher or scientist can be identified as individual-level MPs. To the best of the authors' knowledge, no prior study has explored this endeavor of uncovering individual-level trajectories. Therefore, this study aims to fill that gap.

Uncovering individual-level MPs offers several benefits. First, these MPs can illuminate how a researcher's or scientist's works evolve, particularly how they are influenced by or contribute to other works in the field. Second, these MPs can be compared against those of other researchers or scientists to explore how their trajectories interrelate. Their respective MPs may run parallel, diverge, or converge at certain points, revealing overlaps or intersections in their research efforts. Third, these MPs can also be contrasted with field-level MPs to examine their interactions, offering deeper insights into a researcher's or scientist's contributions relative to the field's mainstream development.

Literature Review

Overview of MPA

To derive MPs from a citation network, MPA primarily involves two key components. First, weights are assigned to the network's arcs to reflect their significance in knowledge diffusion (Hummon & Doreian, 1989). Once the arc weights are assigned, MPA performs a path search on the weighted network to identify chains of connected arcs that extend from the network's sources to its sinks, which are then identified as MPs.

There are various weight assignment and path search algorithms in MPA. The most widely used weight assignment algorithms—namely SPC, SPLC, and SPNP—are collectively known as SPX algorithms. For an in-depth description of these algorithms, refer to Kuan (2020).

The most popular path search algorithms, such as those available in Pajek, can be broadly categorized into global and local searches, each with a number of similar variants listed in Table 1. The approach introduced in this study is based on a path search algorithm called the key-node search (Kuan, 2024; Kuan & Liao, 2024), which also has global and local variants (more details are provided later).

Table 1. Categorization of common path search algorithms.

Category	***Variants***	***Related parameters***
Global searches	Global standard	
	Global key-route	arcs having the topmost *N* weights as key routes
	Global key-node	a designated set of key nodes
Local searches	Local forward	a tolerance value between 0 and 1
	Local backward	a tolerance value between 0 and 1
	Local key-route	1) a tolerance value between 0 and 1 2) arcs having the topmost *N* weights as key routes
	Local key-node	1) a tolerance value between 0 and 1 2) a designated set of key nodes

Global searches identify MPs by selecting paths having the highest path weights (i.e., the sum of all arc weights along a path) between two sets of nodes. In contrast, local searches construct the MPs incrementally, progressing step by step from one set of nodes to another.

More specifically, global standard (GS) search (Liu & Lu, 2012) selects MPs from the paths between the network's sources and sinks. Local forward (LF) search (Hummon & Doreian, 1989) begins at the sources and progressively selects the highest weighted outgoing arcs, moving to subsequent nodes until a sink is reached. Conversely, local backward (LB) search (Hummon & Doreian, 1989) starts at the sinks and traces backward through the highest weighted incident arcs until a source is reached. When conducting local search, a tolerance value can be set to include arcs within a specified range of the highest weight for tracing (De Nooy, Mrvar, & Batagelj, 2018). For instance, a local search with the tolerance value of 0.10 would trace incident or outgoing arcs with weights that are at least 90% of the highest weight among them.

Key-route searches (Liu & Lu, 2012) develop MPs by starting with a set of highest-weighted arcs, referred to as key routes. For a given key route (i, j), the global key-route (GKR) search performs the GS searches between the sources and the arc's start node i, and between the arc's end nodes j and the sinks, to identify one or more global paths preceding and succeeding the key route (i, j), respectively. These global paths are then concatenated with the key route (i, j) to form its GKR MPs. Similarly, the local key-route (LKR) search employs LB and LF searches, instead of GS searches, to derive the preceding and succeeding paths for the key route (i, j).

Key-route searches involve a parameter N, which specifies arcs with the topmost N weights to be used as key routes. For instance, a key-route 1 search initiates MP development from the arc with the highest weight, while a key-route 10 search includes arcs with weights up to the 10^{th} highest. In key-route N searches, the number of key routes may exceed N if weights are tied.

Key individuals along the MPs

There is a wealth of research involving the application of MPA to uncover a field's field-level MPs. Among these studies, some have also focused on identifying significant individuals, especially firms, within the field. These studies generally follow a common approach: they first derive the field-level MPs and then identify individuals whose works appear on these MPs. Such individuals are considered key contributors, as their works are integral to the most representative trajectories of the field's evolution.

Recent studies provide several examples. For instance, Su, Chen, Chang, and Lai (2019) employed MPA to trace the dominant knowledge flow in the field of blockchain technology and identified owners of the patents on the MPs as key players for the field. The study then analyzed the patent families of these key players to investigate their strategic intent in managing their patent portfolios.

Cho, Liu, and Ho (2021) applied MPA to patents related to autonomous driving to uncover the technology development trajectory for the field. The study identified assignees whose patents appeared on the trajectory as key players. Additionally, based on the different phases along the development trajectory and the associated key assignees within each phase, the study categorized these key players into groups such as "technology developers," "technology integrators," and "technology implementers." A similar methodology was adopted by Chen and Cho (2023) to analyze trends and identify key players in the field of Low Earth Orbit (LEO) satellite technology using patents.

Watanabe and Takagi (2021) used MPA to examine how technology has evolved within the field of computer graphic processing systems. The study developed MPs for the field at 5-year intervals and observed the appearance and disappearance of firms owning patents on the MPs over time. The authors noted that these patterns of firm appearances and disappearances align with the historical evolution of the field.

The above studies have several limitations. Firstly, as only individuals with works along the MPs are considered, those without any works on the MPs are overlooked. Additionally, field-level MPs may fail to capture other relevant or even key individuals, as Verspagen (2007) empirically demonstrated that MPA is highly selective at the firm level, with many active individuals in the field not present on the MPs. Furthermore, the identified individuals may have additional works beyond those located on the field-level MPs, which may be overloooked under this approach.

MPs from specific nodes

The key-node search algorithms employed in this study is similar to the key-route algorithms, with the key distinction being that they begin MP development from a set of analyst-designated key nodes, rather than a number of top-weighted key routes determined for the analyst. More details on this approach will be provided in the Methodology section.

This study has identified several prior works with methodologies akin to the approach adopted here. Unlike traditional MPA, which typically develops field-level MPs by searching the citation network from sources to sinks, or vice versa (except for the key-route searches described earlier), these studies first analytically identified a number of key documents. They then developed field-level MPs starting from the nodes of these key documents.

Park and Magee (2017, 2019) introduced a modified MPA that develops field-level MPs from designated nodes. The authors first identified patents with high knowledge persistence—a measure of the extent to which knowledge remains in the patents or contributes to later patents based on their structural positions in the patent citation network. They then developed field-level MPs exclusively from the nodes of these so-called high-persistence patents (HPPs), tracing forward to the sinks and backward to the sources. Feng and Magee (2020) followed a similar approach in analyzing patents from four domains of electric vehicles. They derived MPs for each domain from a number of HPPs and identified the assignees of these HPPs as key players for each domain.

Unlike the above studies, which analytically selected key nodes from the citation network, this study manually designates nodes representing the works of a specific researcher or scientist as key nodes. The resulting key-node MPs are therefore referred to as the researcher's or scientist's individual-level MPs. As these MPs are constructed from the field's citation network, rather than using the researcher's or scientist's works in isolation, they reflect how their works evolve within the broader context of the field to which they belong.

In addition to the above-mentioned studies, several works have also explored path development from designated nodes. Ho, Saw, Lu, and Liu (2014) developed a method called "branch paths" to address the risk that minor technologies may be overshadowed by more prominent technologies and thus omitted from the field-level MPs. This method identifies a set of documents related to these minor technologies and traces paths from these designated documents both forward and backward until they encounter a node on the field-level MPs. Liu, Lu, and Ho (2019) referred to this method as the "designated-document approach" and suggested that it could reveal the relationship between these designated documents and the field-level MPs.

While these works also develop paths from specific nodes, their aim is to supplement field-level MPs rather than derive MPs from the perspective of individual researchers or scientists.

Methodology

Key-node search

As mentioned earlier, the key-node search includes global and local variants, similar to the global key-route (GKR) and local key-route (LKR) searches, as summarized in Table 1. The primary distinction is that key-node MPs are derived from specific nodes that are manually designated as key nodes by the analyst. In contrast, in the key-route search, the analyst cannot specify individual arcs as key routes but can only control the parameter *N*.

As illustrated in Figure 1, for a designated key node *k* (the white node), the key-node search identifies the representative preceding and succeeding paths (depicted in solid lines) between the sources (dark nodes to the left) and the key node *k*, and between *k* and the sinks (dark nodes to the right). These paths may pass through intermediate nodes (gray nodes). In global key-node (GKN) search, the representative preceding and succeeding paths are derived using global standard (GS) search, whereas in the local key-node (LKN) search, they are determined using local backward (LB) and local forward (LF) searches, respectively. These representative preceding and succeeding paths are then cascaded to form the MPs for the key node *k*. Finally, the MPs for all key nodes are aggregated to form the overall key-node MPs.

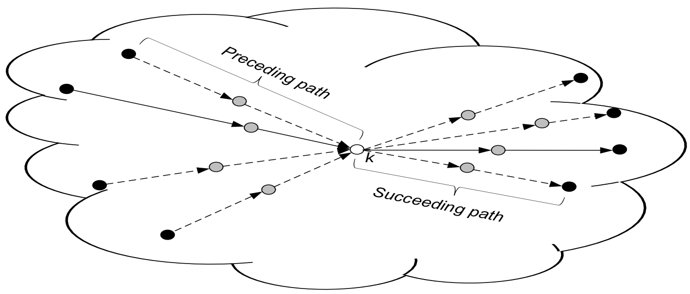

Figure 1. MPs by Key-node search.

In other words, the key-node search constructs MPs by initiating the development of significant paths both preceding and succeeding the designated key nodes. By assigning nodes that represent a researcher's or scientist's works as key nodes, the resulting key-node—or researcher's or scientist's individual-level—MPs reveal, on one hand, the works within the broader field that have most influenced the researcher's or scientist's works and, on the other hand, the representative subsequent developments stemming from the researcher's or scientist's works, also within the broader context of the field.

The individual-level MPs, therefore, offer deeper insights into the research evolution of researchers or scientists than simply aligning their works chronologically. Furthermore, individual-level MPs facilitate a more nuanced understanding of the interrelationships among the researcher's or scientist's works. For instance, some works may appear on separate paths within the individual-level MPs, suggesting that they stem from distinct developmental trajectories in the researcher's or scientist's intellectual endeavors. Conversely, instances where multiple works appear on the same path indicate a continuation of research efforts, signifying progressive knowledge expansion within a single thematic or methodological direction.

Applications of key-node search

Like the common path searches mentioned earlier, the key-node search described above is also available in Pajek, making it readily accessible to analysts. However, perhaps due to its introduction only after 2018—where it is obscurely labeled as "through vertices in cluster"—this path search has seen little application in the literature. To promote awareness of this method and to better reflect its characteristics and similarity to the key-route search, the term "key-node search" has been coined.

Despite its simplicity, the key-node search has the potential to enhance MPA in ways that other common path searches do not. Based on the few related studies available, the following are two examples of its potential applications.

One challenge in MPA is the lack of a quantitative measure to evaluate how well MPs accurately capture and reflect overall knowledge development within a field. To address this, Kuan and Liao (2024) proposed that the representativeness of MPs is limited to the portions of the network that are reachable from or to the MPs, referring to these portions as the MPs' coverage. The study further suggested that the proportion of documents falling within the MPs' coverage can serve as a quantitative measure of their representativeness.

To uncover MPs' coverage, the study applied the LKN search with a tolerance value of 1, using all nodes on the MPs as key nodes. This approach allowed the LKN search to trace all incident and outgoing arcs for each MP node, thereby encompassing the portions of the network that were reachable from or to the MP nodes. When a significant portion of the network fell outside the MPs' coverage, reflecting a low representativeness for the MPs, the study proposed a method to identify auxiliary MPs from this out-of-coverage portion. This portion was also determined using the LKN search with a tolerance value of 1, where the key nodes included those lying outside the MPs' coverage.

Kuan (2024) observed that MPA analysts often possess domain knowledge about the field under analysis, including seminal works crucial to its development. Rather than leaving this knowledge unused in the MPA process or restricting its use solely to document collection or validation of obtained MPs, the study proposed manually incorporating these seminal works into MPA using the key-node search to generate MPs that capture a distinct knowledge flow centered around these key documents.
The study further suggested observing key-document MPs alongside field-level MPs to simultaneously examine the focused knowledge flow through key documents and the overall knowledge flow of the field. This concurrent observation allows for an analysis of their interactions, providing additional insights into the field's development. To facilitate this process, the study proposed generating key-document and field-level MPs automatically and simultaneously, both using key-node searches.
While the key-node search may seem like just one of many path search options in MPA, Kuan (2024) formally verified that the field-level MPs generated by the popular path search algorithms listed in Table 1 can all be reproduced using the key-node search with appropriately selected key nodes—except for key-route MPs, which are subject to certain preconditions. This finding establishes the key-node search as a uniquely versatile method among the algorithms listed in Table 1.

Case study

Data set

To demonstrate the real-world application of the proposed approach, this study conducts a case study using the publicly available Artificial Intelligence (AI) Patent Dataset provided by the United States Patent and Trademark Office (USPTO). This dataset comprises 13,244,037 U.S. patent documents, including utility patents and pre-grant publications (PGPubs), spanning the years 1976 to 2020.
Each patent document is classified by the USPTO using a machine learning approach to predict its relevance to one of eight AI technology fields: machine learning (ML), natural language processing (NLP), computer vision (CV), speech (S), knowledge processing (KP), AI hardware (AIH), evolutionary computation (EC), and planning and control (P&C) (Giczy, Pairolero, & Toole, 2022).
This study selects patent documents predicted to belong to the field of Evolutionary Computation (EC), resulting in 48,999 patent documents covering 36,560 inventions. EC is chosen for its versatility, which makes it a foundational approach in modern AI research and applications, offering potentially diverse and complex knowledge flows for analysis.
EC draws inspiration from biological evolution to solve optimization and search problems. It encompasses a family of techniques, including genetic algorithms, genetic programming (applying a genetic algorithm to a population of computer programs), and differential evolution (generating new candidate solutions by combining the differences between randomly selected individuals in a population of candidate solutions), which simulate natural selection, mutation, crossover, and survival of the fittest to iteratively refine solutions (Bäck, Hammel, & Schwefel,

1997). EC is widely applied in machine learning, robotics, optimization, and complex system design due to its ability to efficiently explore large search spaces and adapt to dynamic environments.

As for the researcher or scientist whose research trajectory is to be observed, this study selects John R. Koza, a pioneer in the EC field (Mitchell & Taylor, 1999). He is known for his work in genetic programming, particularly in automated program generation, where computer programs are evolved to solve complex tasks. Mr. Koza is listed as the inventor on 14 U.S. patents, 12 of which are predicted to be EC-related in the AI Patent Dataset. This study considers these 12 patents to constitute Mr. Koza's body of research for analysis. These patents are listed in Table 2, arranged in ascending order of their application dates.

Table 2. Patents with John R. Koza as the sole inventor or one of the inventors.

#	*App. no.*	*App. date*	*Pub. no.*	*Pub. date*	*Title*
1	7196973	19880520	4935877	19900619	Non-linear genetic algorithms for solving problems
2	7584259	19900918	5148513	19920915	Non-linear genetic process for use with plural co-evolving populations
3	7787748	19911105	5136686	19920804	Non-linear genetic algorithms for solving problems by finding a fit composition of functions
4	7881507	19920511	5343554	19940830	Non-linear genetic process for data encoding and for solving problems using automatically defined functions
5	7899627	19920616	5390282	19950214	Process for problem solving using spontaneously emergent self-replicating and self-improving entities
6	8286134	19940804	5742738	19980421	Simultaneous evolution of the architecture of a multi-part program to solve a problem using architecture altering operations
7	8603648	19960220	5867397	19990202	Method and apparatus for automated design of complex structures using genetic programming
8	8813894	19970307	6058385	20000502	Simultaneous evolution of the architecture of a multi-part program while solving a problem using architecture altering operations
9	9290521	19990412	6532453	20030311	Genetic programming problem solver with automatically defined stores loops and recursions
10	9336373	19990617	6424959	20020723	Method and apparatus for automatic synthesis, placement and routing of complex structures
11	9393863	19990910	6564194	20030513	Method and apparatus for automatic synthesis controllers
12	10355443	20030130	7117186	20061003	Method and apparatus for automatic synthesis of controllers

As mentioned earlier, aligning Mr. Koza's patents as listed in Table 2 provides little insight into the evolution of his EC research. While a subjective examination of their document contents and prosecution histories may reveal how some patents are related to or directly derived from others, this alone does not objectively determine whether they follow a continuous line of research or originate from separate endeavors—let alone their relationship with other EC patents.

EC citation network

This study constructs a citation network using EC patent documents and their backward citations. A few key points about this construction are as follows:

1. Node Representation: The nodes in the network are identified by their patent application numbers. This arrangement aggregates citations for an invention's patent and its corresponding pre-grant publications (PGPubs), providing a more comprehensive view of its citation relationships (Kuan, Chen, & Huang, 2020).
2. Citation Ordering: All citations are filtered so that the cited patent documents are always those filed earlier than their citing counterparts. This prevents cycles in the network and ensures that knowledge flows consistently from earlier-filed patents to those filed later.
3. Network Closure: The network is closed, meaning that only patent documents classified as EC are included—both cited and citing—by filtering out those outside the EC patent dataset. While this restriction is not mandatory, it is applied for simplicity in analysis.
4. Removal of Anomalies: Loops and duplicate arcs are removed from the citation network. These anomalies result from the aggregation mentioned in (1). For example, a loop occurs when a patent self-cites its own PGPubs, while duplicate arcs appear when a later patent simultaneously cites an earlier patent and its PGPub.

After applying the aforementioned processing steps, the final EC citation network consists of 46,261 arcs and 19,836 nodes, representing approximately 54% of the 36,560 EC inventions. In other words, roughly half of the EC inventions lack mutual citations and are therefore not part of the citation network, suggesting potential imprecision in the AI Patent Dataset. However, this study verifies that all 12 of Mr. Koza's EC patents are included in the citation network.

The citation network is distributed across 1,178 components (i.e., isolated sub-networks). The largest component includes 16,757 nodes, accounting for approximately 85% of the total nodes, whereas all other components are significantly smaller (the second-largest component contains only 27 nodes). The MPs produced in subsequent analyses will be derived entirely from this largest component, as MPs do not extend across disconnected components.

A macro-view based on a researcher's or scientist' entire set of works

To derive Mr. Koza's research trajectory, this study assigns SPNP weights to the arcs of the citation network (Kuan, 2020). Subsequently, the 12 nodes, each corresponding to one of Mr. Koza's patents listed in Table 2, are gathered into a Pajek cluster, and the GKN and LKN searches are applied to generate Mr. Koza's individual-level GKN and LKN MPs. For simplicity, the LKN search is conducted with a tolerance value of zero, meaning that only arcs with the topmost weight are traced.

The resulting GKN MPs include 54 nodes, while the LKN MPs include 69 nodes. Although the two sets of MPs differ—each containing some nodes absent from the other—both reflect a common theme in the evolution of Mr. Koza's research, as their interconnected structures share a consistent framework (as described below). Additionally, 50 out of the 54 nodes in the GKN MPs are also present in the LKN MPs. Therefore, for brevity, only the GKN MPs are presented in Figure 2.

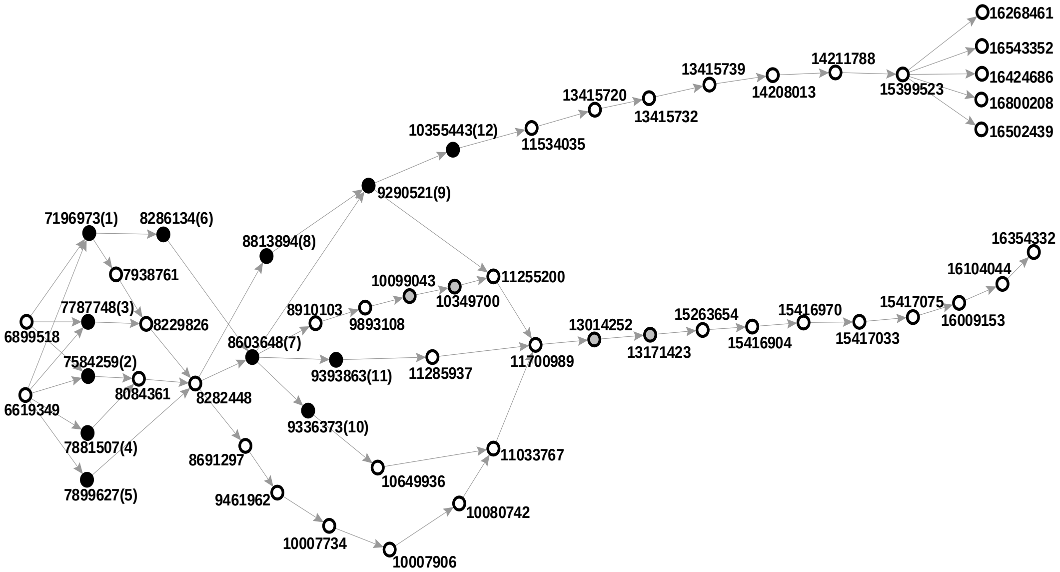

Figure 2. Mr. Koza's individual-level MPs by GKN search.

In Figure 2, the nodes are labeled with their corresponding patent application numbers. The black nodes represent Mr. Koza's 12 patents, with their sequence numbers from Table 2 in parentheses attached to their labels. The four grey nodes denote those that are not present in the LKN MPs.

At first glance, Figure 2 reveals that, as an early pioneer in the EC field, Mr. Koza's works are concentrated in the early (or left) half of the trajectory. All of his works can be traced back to a common origin. Then, Mr. Koza's works initiate two distinct strands of subsequent development in the later (or right) half of the trajectory.

As mentioned earlier, the LKN MPs reveal an identical framework to that depicted in Figure 2, except that they include an additional source, an additional sink, and several extra nodes and branches in the denser left portion of the trajectory.

A closer examination of the patents in Figure 2 reveals that the early half of Mr. Koza's individual-level MPs follows a development trajectory centered on the evolution of computer programs based on genetic algorithms. Interestingly, in the later half, the trajectory transitions toward neural network-based product design and the training of machine learning models.

The common origin of all 12 of Mr. Koza's patents involves two prior patents, both of which are based on genetic algorithms:

1. Application No. 6619349 (corresponding to Patent No. 4697242) lists John Holland as one of the inventors, who is recognized as the father of genetic algorithms (Bäck, Hammel, & Schwefel, 1997). This patent describes an adaptive computing system consisting of a population of classifiers. The system employs a genetic algorithm to generate new classifiers, replacing less effective ones and enabling continuous learning and improvement.
2. Application No. 6899518 (corresponding to Patent No. 4821333) describes a method for image recognition, particularly focusing on applying mutation and crossover mechanisms to evolve sets of structuring elements within an image. The goal is to identify an optimal set of structuring elements that can effectively distinguish between image categories.

For brevity, this study examines four patents, selecting two from each strand of subsequent developments. In the lower right part of Figure 2, the knowledge flow from Mr. Koza's genetic programming work shifts into the training of machine learning models:

1. Application No. 15263654 (corresponding to Patent No. 10387801) focuses on training and assessing a machine-learned model to refine a large collection of documents (e.g., web pages from a search engine) into a shorter ordered list (akin to a partial order). The ranking is derived from multiple parameters that reflect relevance, similar to fitness values in evolutionary algorithms.
2. Application No. 16354332 (corresponding to Patent No. 11494691) also focuses on training and assessing a machine learning model but specifically optimizes the training process. This more advanced patent introduces a technique that utilizes the idle time while the machine learning model awaits actual outcomes from its previous action. During this waiting period, the system generates a set of predicted outcomes and uses at least a subset of them to train the model, producing multiple candidate models—thereby accelerating the training process.

In the upper right part of Figure 2, the knowledge flow from Mr. Koza's work shifts separately into the domain of product design utilizing neural networks:

1. Application No. 11534035 (corresponding to Patent No. 8423323) discloses a system and method for designing new products. A mapping relationship between consumer preferences and product attributes is modeled using neural networks. Interactive Evolutionary Computation (IEC) and genetic algorithms are integrated to optimize the model and search process, allowing designers to predict the acceptance of new products and identify highly desirable yet underrepresented areas in the market.
2. Application No. 15399523 (corresponding to Patent No. 10783429) integrates artificial neural networks and evolutionary computation to automate the analysis of large-scale user data and efficiently identify the most effective web design. At the core of the system is a neural network that maps user attributes to different dimensions and values of a web page. The neural network is represented as a

genome and optimized through evolutionary operations such as initialization, testing, competition, and reproduction.

Despite the abridged description above, one can still discern a lineage of evolving ideas through the patents preceding and succeeding Mr. Koza's patents.

A micro-view based on a single work from a researcher or scientist

The previous section provides a macro-level perspective on Mr. Koza's individual-level MPs, demonstrating the usefulness of these MPs in identifying both the most influential sources contributing to his research and the most prominent subsequent developments arising from it as a whole.

However, this macro-view has limitations, as it does not explicitly clarify how Mr. Koza's specific patents are related to one another, nor how they connect to the identified sources and subsequent developments. For example, considering the patent Application No. 9290521, the macro-view alone does not help analysts determine whether it is more closely related to the upper strand of development, as suggested in Figure 2.

Additionally, Figure 2 shows that five of Koza's patents (numbered 1 to 5) appear in parallel immediately after their two common prior sources. However, the macro-view again fails to inform analysts whether they are equally related to Application No. 9290521. In fact, as will be demonstrated later, Figure 2 may even be somewhat misleading in answering these questions.

To overcome the shortcomings of the macro-view, this study proposes a micro-level perspective by conducting a GKN or LKN search on specific nodes representing the patents of interest, rather than designating all of Mr. Koza's patents as key nodes. To demonstrate the usefulness of this micro-view in supplementing the macro-view, this study performs a GKN and LKN search on a single key node, corresponding to Application No. 9290521.

Again, for brevity, only the resulting GKN key-node MPs are presented in Figure 3, as the differences between them and the LKN key-node MPs are minor. Similarly, in Figure 3, the black nodes represent Mr. Koza's patents (including 9290521), while the three gray nodes denote patents not present in the LKN key-node MPs.

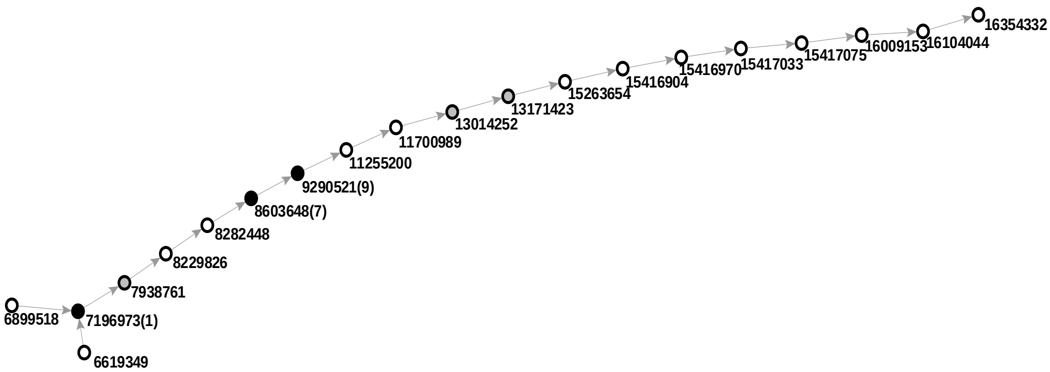

Figure 3. MPs from a single Mr. Koza's patent by GKN search.

Figure 3 reveals some unexpected observations. Firstly, two of Mr. Koza’s patents, Application Nos. 8603648 and 7196973, are more significantly related to 9290521 than the others in terms of knowledge diffusion, as they are aligned along the same knowledge-diffusion lineage.
There are also two other patents adjacent to 9290521 besides 8603648 in Figure 2—Application Nos. 10355443 and 8813894. However, the key-node search selects 8603648, including it in 9290521’s individual lineage.
As illustrated, this micro-view helps analysts differentiate the degree of relatedness between 9290521 and Mr. Koza’s other patents, as well as understand its lineage. The same approach can be individually applied to each of Mr. Koza’s patents.
Second, while 9290521 is structurally closer to the strand of subsequent development related to product design utilizing neural networks, its key-node MPs reveals that it is more closely aligned, in terms of knowledge diffusion, with the strand of subsequent development concerning the training of machine learning models.

Conclusion

This study contributes to the understanding of MPA by:

1. Promoting awareness of a previously overlooked path search algorithm in MPA, termed key-node search, which derives MPs extending both backward and forward from one or more key nodes.
2. Demonstrating the application of key-node search to capture researchers’ or scientists' individual-level MPs, reflecting their research trajectories within the broader context of the field to which they belong.

Using a case study, this study demonstrates both a macro-view and a micro-view of a researcher’s or scientist’s individual-level MPs. The macro-view designates all of the researcher’s or scientist’s works as key nodes, helping to identify both the most influential sources contributing to the researcher’s or scientist’s research and the most prominent subsequent developments arising from it as a whole.
The micro-view, on the other hand, designates one or a few of the researcher’s or scientist’s works as key nodes. This supplements the macro-view by differentiating the degree of relatedness between these works and the researcher’s or scientist’s other works. Additionally, the micro-view provides insights into the relationship between these specific works and the most prominent subsequent developments uncovered in the macro-view.
While the individual-level MPs uncovered in the case study appear reasonable, the greatest challenge to the proposed approach lies in verifying how accurately these MPs reflect a researcher’s or scientist’s research evolution and how trustworthy the identified contributing sources and subsequent developments are.
Currently, analysts can only rely on subjective evaluation, experts’ domain knowledge, or existing review articles and industry reports, if available. The issue of representativeness remains unresolved. However, this challenge is not unique to this study—it is a common limitation across all studies utilizing MPA.
There are several interesting extensions to this study. One such extension is that, instead of limiting the proposed approach to individual researchers or scientists, it could be applied to other types of "individuals," such as paper authors, research

institutes, or firms. The resulting individual-level MPs could then be interpreted as reflecting their research trajectories within the broader field.

Additionally, this study does not explore how one researcher's or scientist's individual-level MPs compare with those of another scientist or with the field-level MPs. Regarding the former, such an investigation could reveal how their research trajectories interact within the field—whether they run in parallel, converge, or diverge at certain points, among other patterns. Regarding the latter, examining interactions between individual-level and field-level MPs could uncover certain patterns, allowing researchers or scientists to be categorized based on their alignment with the field's mainstream development.

Open science practices

The data and software used in the case study are both publicly and freely available. The AI Patent Dataset can be downloaded from USPTO's website (https://www.uspto.gov/ip-policy/economic-research/research-datasets/artificial-intelligence-patent-dataset). The software Pajek can be downloaded from its official website (http://mrvar.fdv.uni-lj.si/pajek/).

Acknowledgments

This work was financially supported by the Center for Research in Econometric Theory and Applications (Grant no. 113L900202), which is under The Featured Areas Research Center Program by the Ministry of Education (MOE) in Taiwan, and by the National Science and Technology Council (NTSC), Taiwan, under Grant No. 112-2221-E-011-116-MY2.

References

Back, T., Hammel, U., & Schwefel, H. P. (1997). Evolutionary computation: Comments on the history and current state. *IEEE transactions on Evolutionary Computation*, *1*(1), 3-17.

Batagelj, V., & Mrvar, A. (1998). Pajek - Program for large network analysis. *Connections*, *21*(2), 47-57.

Chen, P. H., & Cho, R. L. T. (2023). The Technological Trajectory of LEO Satellites: Perspectives from Main Path Analysis. *IEEE Transactions on Engineering Management*.

Cho, R. L. T., Liu, J. S., & Ho, M. H. C. (2021). The development of autonomous driving technology: perspectives from patent citation analysis. *Transport Reviews*, *41*(5), 685-711.

De Nooy, W., Mrvar, A., & Batagelj, V. (2018). *Exploratory social network analysis with Pajek: Revised and expanded edition for updated software* (Vol. 46). Cambridge University Press.

Feng, S., & Magee, C. L. (2020). Technological development of key domains in electric vehicles: Improvement rates, technology trajectories and key assignees. *Applied Energy*, *260*, 114264.

Giczy, A. V., Pairolero, N. A., & Toole, A. A. (2022). Identifying artificial intelligence (AI) invention: A novel AI patent dataset. *The Journal of Technology Transfer*, *47*(2), 476-505.

Han, F., Yoon, S., Raghavan, N., Yang, B., & Park, H. (2024). Technological trajectory in fuel cell technologies: A patent-based main path analysis. *International Journal of Hydrogen Energy*, *50*, 1347-1361.

Ho, J. C., Saw, E. C., Lu, L. Y., & Liu, J. S. (2014). Technological barriers and research trends in fuel cell technologies: A citation network analysis. *Technological Forecasting and Social Change*, *82*, 66-79.

Hummon, N. P., & Doreian, P. (1989). Connectivity in a citation network: The development of DNA theory. *Social Networks*, *11*(1), 39-63.

Kuan, C. H. (2020). Regarding weight assignment algorithms of main path analysis and the conversion of arc weights to node weights. *Scientometrics*, *124*(1), 775-782.

Kuan, C. H. (2024). Integrating prior field knowledge as key documents with main path analysis utilizing key-node path search. *Journal of Informetrics*, *18*(3), 101569.

Kuan, C. H., Chen, D. Z., & Huang, M. H. (2020). The overlooked citations: Investigating the impact of ignoring citations to published patent applications. *Journal of Informetrics*, *14*(1), 100997.

Kuan, C. H., & Liao, S. Y. (2024). Assessing main paths by uncovering their coverage with key-node path search. *Scientometrics*, *129*(11), 6629-6657.

Liu, J. S., & Lu, L. Y. (2012). An integrated approach for main path analysis: Development of the Hirsch index as an example. *Journal of the American Society for Information Science and Technology*, *63*(3), 528-542.

Liu, J. S., Lu, L. Y., & Ho, M. H. C. (2019). A few notes on main path analysis. *Scientometrics*, *119*, 379-391

Mitchell, M., & Taylor, C. E. (1999). Evolutionary computation: an overview. *Annual Review of Ecology and Systematics*, *30*(1), 593-616.

Park, H., & Magee, C. L. (2017). Tracing technological development trajectories: A genetic knowledge persistence-based main path approach. *PloS One*, *12*(1).

Park, H., & Magee, C. L. (2019). Quantitative identification of technological discontinuities. *IEEE Access*, *7*, 8135-8150.

Su, F. P., Chen, S. J., Chang, Y. H., & Lai, K. K. (2019). Construct a three-stage analysis model of integrated main path analysis and patent family-exploring the development of blockchain. In *Proceedings of the 2019 3rd International Conference on Software and e-Business* (pp. 151-156).

Verspagen, B. (2007). Mapping technological trajectories as patent citation networks: A study on the history of fuel cell research. *Advances in complex systems*, *10*(01), 93-115.

Watanabe, I., & Takagi, S. (2021). Technological trajectory analysis of patent citation networks: examining the technological evolution of computer graphic processing systems. *The Review of Socionetwork Strategies*, *15*, 1-25.