PNAS RESEARCH ARTICLE | APPLIED MATHEMATICS OPEN ACCESS

# The evolution of Traditional Chinese Medicine as recombinant inventions

Min Wu[a], Tieju Ma[b,c,1], Yan Zhu[d,1], Hongtao Ren[a], and Lu Fu[e]

Affiliations are included on p. 8.

Edited by David Weitz, Harvard University, Cambridge, MA; received January 16, 2024; accepted October 8, 2024

**It has been widely recognized that technologies evolve with recombinant inventions. However, it remains unknown whether technologies developed using different approaches would exhibit different features during evolution. In particular, would technologies developed mainly based on accumulated experience in practices—formulas of traditional Chinese medicine (TCM) are typical examples of such technologies—have similar evolution features found in modern technologies? This study applied network science to explore the evolution of TCM from the perspective of recombinant inventions based on 59,063 TCM formulas documented over the last two thousand years, with each formula being a combination of components that are mostly herbs. Our results show that similar to modern technological systems, the TCM component networks maintained the core–periphery structures during evolution, and the (weighted) degrees of components followed heavy-tailed distributions. Moreover, simple tuples, which are frequently used combinations of TCM components, serve as building blocks for complex ones. A significant difference with modern technological systems is that the TCM core components were quite stable, while substitutions of core components are frequently observed in modern technological systems, leading to new technological trajectories. TCM comprises ancient knowledge and wisdom. This research provides insight into how it will be like in the future and what is important for its future.**

Traditional Chinese Medicine | network science | recombinant invention | evolution

## Significance

Traditional Chinese medicine (TCM) has been used for thousands of years and still being developed. How TCM formulas were developed and evolved somehow remain mysterious. This study tries to reveal part of the mystery by analyzing the evolution of TCM as recombinant inventions quantitively, based on 59,063 TCM formulas documented over the last two thousand years. A significant finding is that TCM core components were quite stable even in thousands of years, while the substitutions of core components are frequently observed in modern technological systems. The result of this study provides insight into how new TCM formulas could be developed and what are important for future development of TCM.

It has been widely recognized that the evolution of technological systems originates from the recombination of existing and newly discovered ones (1–3), with some earlier combinations serving as building blocks for more complex ones (4). Technological systems would evolve into specific structures with constant recombination (5).

Network science is generally used to explore these structures (6). The evolution and structure of technological systems has been studied mostly based on patent data (2, 7–9), scientific publications (10, 11), and software systems (12–14), which are well structured and documented. It has been found that similar to many other complex systems (15, 16), the degrees of components in most technological systems follow heavy-tailed distributions, which means only a small number of components, called "hubs," interact with other components, while most components only interact with a small number of other components (16–18). During the evolution of complex systems, the "hubs" tend to attract more new edges to them (19, 20). It has also been found that most technological systems have core–periphery structures with some hub components tightly coupled with each other, being core components, with other components being peripheries (21–23).

It remains unknown whether technologies developed mainly based on experience accumulated in practices—traditional Chinese medicine (TCM) formulas are typical examples of such technologies—would have similar evolution features found in modern technologies, because usually there were no well-structured historical data available for studying their evolution. This study fills this research gap by exploring the evolution of TCM, which is one of the few technological systems that has evolved for thousands of years and is still being developed and used. TCM is thought a treasure of ancient human's knowledge and wisdom. How it has evolved and most importantly how it will evolve in the future? Answers to these questions not only reveal some mysteries about TCM but also provide insight into how it will be like in the future and what kind of changes are important for its future.

Thousands of formulas were developed during the evolution of TCM. A TCM formula comprising several components can be considered as a technological invention resulting from a recombinant search. The components are mainly different herbs but also include some animal parts and special stones. TCM comprises thousands of formulas and components and can be viewed as a complex technological system.

Author contributions: T.M. designed research; M.W., T.M., and L.F. performed research; H.R. contributed new reagents/analytic tools; M.W., Y.Z., and H.R. analyzed data; Y.Z. developed the dataset; L.F. provided explanations of the TCM core components and examples of the building-out of tuples; and M.W. and T.M. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.



[1]To whom correspondence may be addressed. Email: tjma@sjtu.edu.cn or zhuy@mail.cintcm.ac.cn.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2400812121/-/DCSupplemental.

Published November 7, 2024.

TCM formulas were invented mostly based on inventors' accumulated knowledge and experience. Inventors' knowledge could be explicit knowledge from existing TCM books but mostly comprises tacit knowledge learned from supervisors and individual practices. In addition to trial-and-error, the ancient Chinese also developed theories for inventing TCM formulas, with the theory of Yin-Yang and Five Elements being the most popular. The ancient Chinese also developed principles for combinations of different components, the most dominant of which is called the principle of Monarch, Minister, Assistant, and Guide. When inventing TCM formulas, an inventor may be influenced by such theories and principles, which is in accordance with the argument that the combinatorial potential of components in recombinant inventions depends on revealed beliefs rather than the state of nature (8, 24, 25).

The development of ancient TCM was influenced by international trade and cultural changes. For example, TCM absorbed components (especially spice materials) from Middle East Asia and Europe, which were imported into ancient China through the Silk Road (26). The introduction of Buddhism in ancient China also introduced changes in the philosophy underlying the development of TCM (27).

Significant efforts have been dedicated to qualitatively study the history and evolution of TCM. These studies have summarized and discussed the main practices, theories, and development of TCM and distinguished ancient Chinese doctors from different historical periods, as well as their inheritance relationships (28, 29).

In contrast to these existing studies, our research attempts to reveal the evolution of TCM by quantitatively analyzing all documented (ancient) TCM formulas by applying network science, treating TCM as a complex system, and viewing the development of TCM from the perspective of recombinant inventions. Such an analysis is impossible because of the lack of well-structured data on the large number of ancient TCM formulas. With the significant efforts of Cui et al. (30), a well-structured dataset of TCM formulas has been made available, which enables the analysis of TCM as a complex technological system over a long-time horizon.

The rationale for applying network science is that it is considered a useful tool and has been widely used to study complex systems with a large number of components and interactions/edges among them. Moreover, a significant number of indicators have been developed based on network science to identify important components, structures, and evolutions of complex systems (for example, 14, 15, 31, 32). TCM includes thousands of formulas that are formed by different combinations of thousands of components (mainly herbs); thus, applying network science is appropriate for our study, because our aim is to explore the evolution of structures and important components of TCM. We construct TCM component networks for different periods with components (mainly herbs) as nodes and the co-occurrence frequency of a pair of TCM components in different formulas as their weighted edges (more details on the construction of the TCM component network can be found in *SI Appendix*, Figs. S1 and S2), which is the most frequently used method to define edges when using network science to analyze patent classifications, keywords in the literature, etc. (for example, 25, 33–35). Considering the heterogeneity in the intensity of edges could be important in understanding technological systems (36), all the analysis in the main text takes weights of edges between TCM components into account. We also conduct analysis ignoring the weights of edges, which is reported in *SI Appendix*.

This study aims to explore whether the evolution of TCM shares similar features found in modern technological systems, which could provide insights into how systems such as TCM, which is the result of the accumulation of knowledge and wisdom of humans for thousands of years, would evolve in the future, and how to keep such systems vital in the modern age.

## Results

**Dynamics of Number of TCM Components and Formulas.** The Fig. 1 shows the dynamics of the numbers of components (mostly herbs) and formulas starting from AD 1, with time intervals of 100 years and different dynasties of China segmented with dashed lines. It can be seen that the cumulative number of TCM formulas as well as that of components used in TCM formulas exhibited an increasing trend. However, not all components are always used in documented TCM formulas in different periods. We call these components used in a period as active components in that period. Fig. 1 reveals that the number of active components increased significantly from AD 1 to 1,200, it remained relative stable from years 1200 to 1800, and then an obvious decreasing trend from years 1800 to 2016 is observed, denoting that a significant number of components were abandoned during this period. An increasing trend of disappearing of components after year 1,000 is also observed, which means that some previously used components became obsolete, because they might have been considered as not useful with the accumulation of experience. The lower degree a component is in a former period, the higher the probability of it being abandoned in the later periods during the evolution of TCM (*SI Appendix*, Fig. S5).

**Dynamics of Edges.** Fig. 2 illustrates ratios of different types of new edges contributing to the expansion of the TCM component network, from which we can see that the reinforcement of existing edges contributes heavily to the expansion of the TCM component networks, the triadic closure contributes second, followed by the new edges between existing and new components. If we ignore the weights of edges of the TCM component network, the triadic closure contributes heavily (*SI Appendix*, Fig. S6).

We also conduct regression analysis on the preferential attachment (*SI Appendix*, Table S1 and S2 for the weighted and unweighted TCM component networks, respectively). In the regression considering the weights, the reinforcement of existing edges is treated as new edges. The regressions show that preferential attachment is an important mechanism for the growth of the TCM component network.

In a short summary, preferential attachment (including reinforcement of existing edges and new edges) and triadic closure works together in the expansion of the TCM component network.

**The Structure of TCM Component Networks.** We construct TCM component networks for different periods using the accumulated TCM formulas and components documented at the end of different periods. Fig. 3*A* illustrates the weighted TCM component network (in the form of a Maximum Spanning Tree) constructed based on all documented 59,063 TCM formulas, showing that this network has a core–periphery structure, with some components linked closely in the center and other components scattered away from the center. *SI Appendix*, Fig. S11 further shows that only a small number of components were frequently used together.

To explore whether the TCM component networks in different periods have a core–periphery structure, we use an indicator called the nested overlap and decreasing fill (NODF) and compare the NODF of TCM component network with those of randomly generated null models (*Materials and Methods* and *SI Appendix*, Text for more information about NODF). As shown in Fig. 3*B*, the weighted NODFs of the TCM component networks and the null model 1 (which preserving the degree sequence of the corresponding

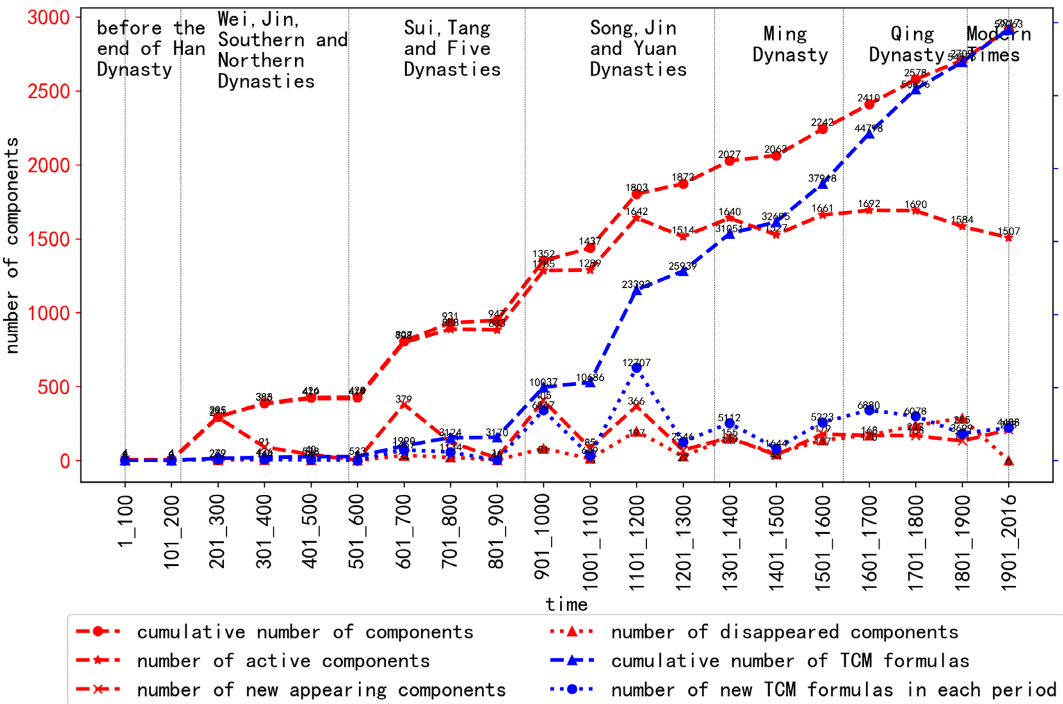


**Fig. 1.** Dynamics of number of TCM components and formulas. The observation period starts from AD 1, with time intervals as 100 y (with the exception that the last period is of 116 y from years 1901 to 2016) and different dynasties of China are segmented with dashed vertical lines. The dashed red line with dot marks denotes the cumulative number of TCM components documented by the end of each period. The dashed red line with star marks denotes the number of active TCM components, that is, the components documented in each period. The dashed red line with X marks denotes the number of new appearing components in each period. The dotted red line with triangle marks denotes the number of disappeared TCM components in each period, that is, components that not be documented after the period. The dashed blue line with triangle marks denotes the cumulative number of TCM formulas. The dotted blue line with dot marks denotes the number of new documented TCM formulas in each period.

TCM component network) exhibit increasing trend and are significantly larger than that of the null model 2 (which only preserving the numbers of nodes and edges). Fig. 3 *C–E* illustrate that a TCM component network as well as the null model 1 have higher nestedness (and thus a higher NODF) than their corresponding null model 2. In a short word, the TCM component networks maintained the core–periphery structure during TCM evolution, which could also be explained by null models taking the degree distribution into account. In addition, this finding also holds for the unweighted TCM component network (*SI Appendix*, Fig. S12).

We find that the weighted degree (i.e., strength) of TCM components in different periods followed a heavy-tailed distribution, fitting both the power-law and lognormal functions well (37) (*SI Appendix*, Figs. S7 and S8), which implies the heterogeneity of nodes' strength of the TCM component network. And this also holds for the unweighted degree (*SI Appendix*, Figs. S9 and S10). This heavy-tailed distribution is consistent with our regression results on *preferential attachment* (*SI Appendix*, Tables S1 and S2).

**TCM Core Components and their Evolution.** We identify the TCM core components in different periods using the continuous model (38) introduced in the *Materials and Methods* (also *SI Appendix*, Text and *SI Appendix*, Fig. S15). The core components identified using this method are linked to many peripheral components but are also themselves densely linked. The set of core components identified from all 59,063 TCM formulas is called the final core components set (FCCS), the set of core components identified with cumulative TCM formulas documented by the end of a period is called the period core component set (PCCS). Following the idea in ref. 32, we use the proportion of components in the FCCS that can be traced back to each PCCS to indicate the stability of TCM core components, and a value higher than 0.5 would indicate the TCM core components are stable.

Fig. 4*A* shows the number of core components in each PCCS as well as the stability of the core components. We find that there is overlap between core components and hubs in different periods, which is consistent with that preferential attachment predicts that hubs would be stable over time (*SI Appendix*, Text and *SI Appendix*, Fig. S16). Fig. 4*B* illustrates the dynamics of the core components, that is, the periods of a component entering the TCM system, becoming a core component, or becoming a noncore component. Fig. 4*C* shows the co-occurrence of core components in three periods, from which the high cohesion of components in the FCCS can be seen. From Fig. 4, it can be seen that most of the components in the FCCS became core components in the early stage, and the TCM core components were stable during the evolution of TCM, with the stability increasing to 0.67 at approximately year 400, and further to 0.88 at approximately year 900, which is different from the modern technology cycle in which technological progress is generally engendered by the frequent substitution of core components (for example, 39–41).

Existing studies on substitutions of core components in modern technological systems are qualitatively analysis and presented in the form of historical stories. We identify the cores of some modern technological systems with the method and measure used in our study and compare the stabilities of their core components with that of the TCM by the first-order statistics, and the detailed results can be found in *SI Appendix*, Fig. S18 which further proves the TCM core components are quite stable comparing with those of these modern technological systems. Of course, the evolution of different technological system could differ substantially, our study does not exclude the possibility that some modern technological systems could have stable cores.

For robustness checks, we also apply other methods to identify TCM core components, such as the rich club method (42), continuous extension model method (43), and random walk method (44) (*SI Appendix*, Figs. S19–S21). *SI Appendix*, Table S3 presents

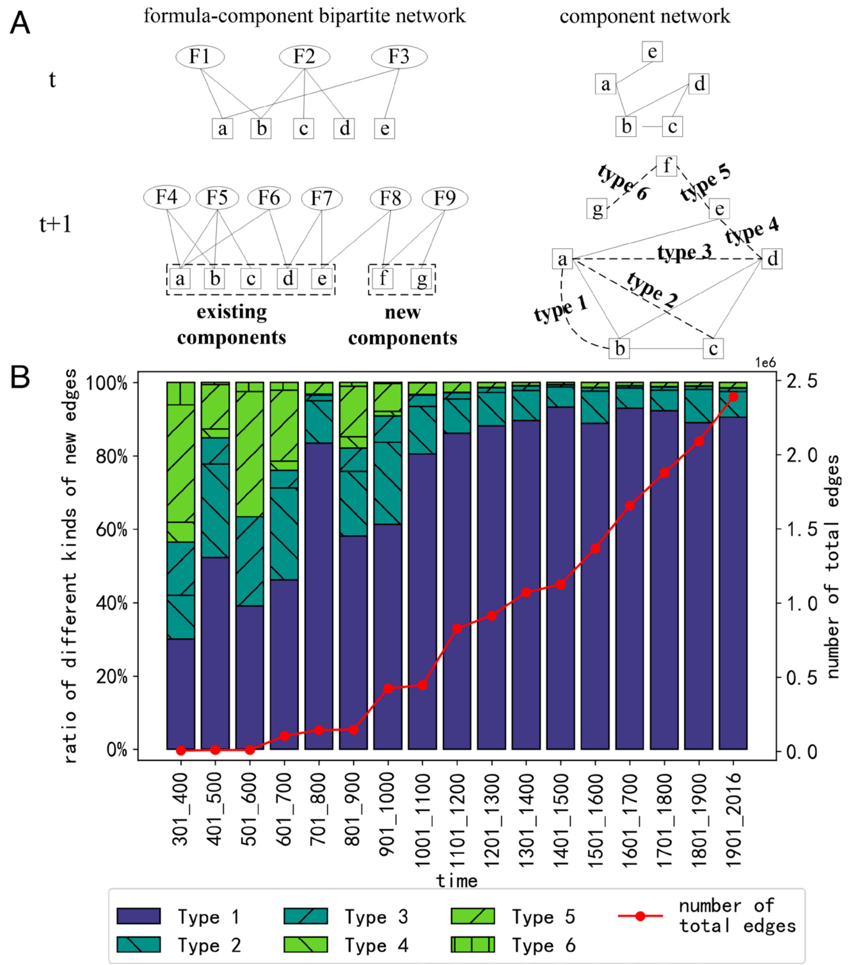


**Fig. 2.** Dynamics of edges during the expansion of weighted TCM component network. (*A*) illustrates different types of new edges contributing to the expansion of the TCM component network. The *Left* part illustrates the formula-component bipartite networks in period t and period t + 1 (with F1, ..., F9 representing formulas, and a, ..., g representing components), which are the basis of constructing the TCM component networks in the *Right* part. In the t + 1 component network, the new edge denoted with Type 1 (i.e., a and b) represents the reinforcement of an existing edge. Type 2 represents a new formula (i.e., F5 in the t + 1 formula-component bipartite network) including three components (i.e., a, b, and c) and forming a triadic closure by bringing a new edge (i.e., a–c) connecting two previous existing components with a common neighbor (i.e., component b). Type 3 denotes a new formula (i.e., F6) brings a new edge (i.e., a–d) connecting two previous existing components with a common neighbor (i.e., component b). From the perspective of the component network, Type 3 also forms a triadic closure, but different to Type 2, the three components (i.e., a, b, and d) are not in the same formula. Type 4 denotes a new formula (i.e., F7) brings a new edge (i.e., d and e) connecting two previous existing components without a common neighbor. Type 5 represents a new edge connecting an existing component (i.e., component e) with a new appearing component (i.e., component f) with a new formula (i.e., F8). Type 6 represents a new edge (i.e., e–f) connecting two new appearing components with a new formula (i.e., F9). (*B*) shows the ratio of different types of new edges during the expansion of the TCM component network (the *Left* axis) as well as the total number of edges in different periods (the *Right* axis). (*B*) started from period 301 to 400 as there are only two edges before year 201.

the core components of the FCCS identified with different methods as well as the most frequently used ones in modern Chinese medicine (45, 46). From *SI Appendix*, Figs. S19–S21, and Table S3, we can see that the core components identified with different methods are similar, and the top ten core components identified with the continuous model are similar to the most frequently used ones in modern Chinese medicine.

Although there are theories regarding the roles of TCM components, including Monarch, Minister, Assistant, and Guide, in practice, it is difficult to identify each component's role in all formulas, because they may play different roles in different formulas (47). For example, *licorice* is thought to play an assistant role in some formulas, but it also worked as a monarch herb in other formulas (*SI Appendix*, Text for concrete examples). Thus, our study does not specify whether an edge between two components means an assistant relation, a reinforced relation, or something else; it simply means that the two components are used together, and the weights of edges denote how frequently they are used together. However, we tested whether our findings would be quite different if we exclude some components that are generally considered assistants, and we found that our main findings still hold in such situations (*SI Appendix*, Fig. S23).

Recent research on TCM has mapped vital TCM substances to signaling molecules (48). The TCM functions and associated signaling molecules of the top 10 core components are reported in *SI Appendix*, Table S5.

**The Building-Out of Tuples.** Fig. 5 illustrates the building-out of tuples of different sizes. Fig. 5*A* provides an intuitive illustration of how large tuples are built from the small ones, from which it can be seen that a large tuple may be linked to several small ones from the merging of those small ones. A small tuple can function as the building block for various large tuples (which is also validated by the regression results in *SI Appendix*, Table S6). And many tuples can be traced back to more than one thousand years ago (*SI Appendix*, Fig. S24).

Fig. 5*B* shows that a large portion (above 70%) of the large tuples find their roots from the small ones. Fig. 5*C* illustrates a concrete example of building-out of a 6-tuple from simpler ones, from which it can be seen that although there could be several paths for its building-out, only one path can be traced back to a 2-tuple, and it is not always the

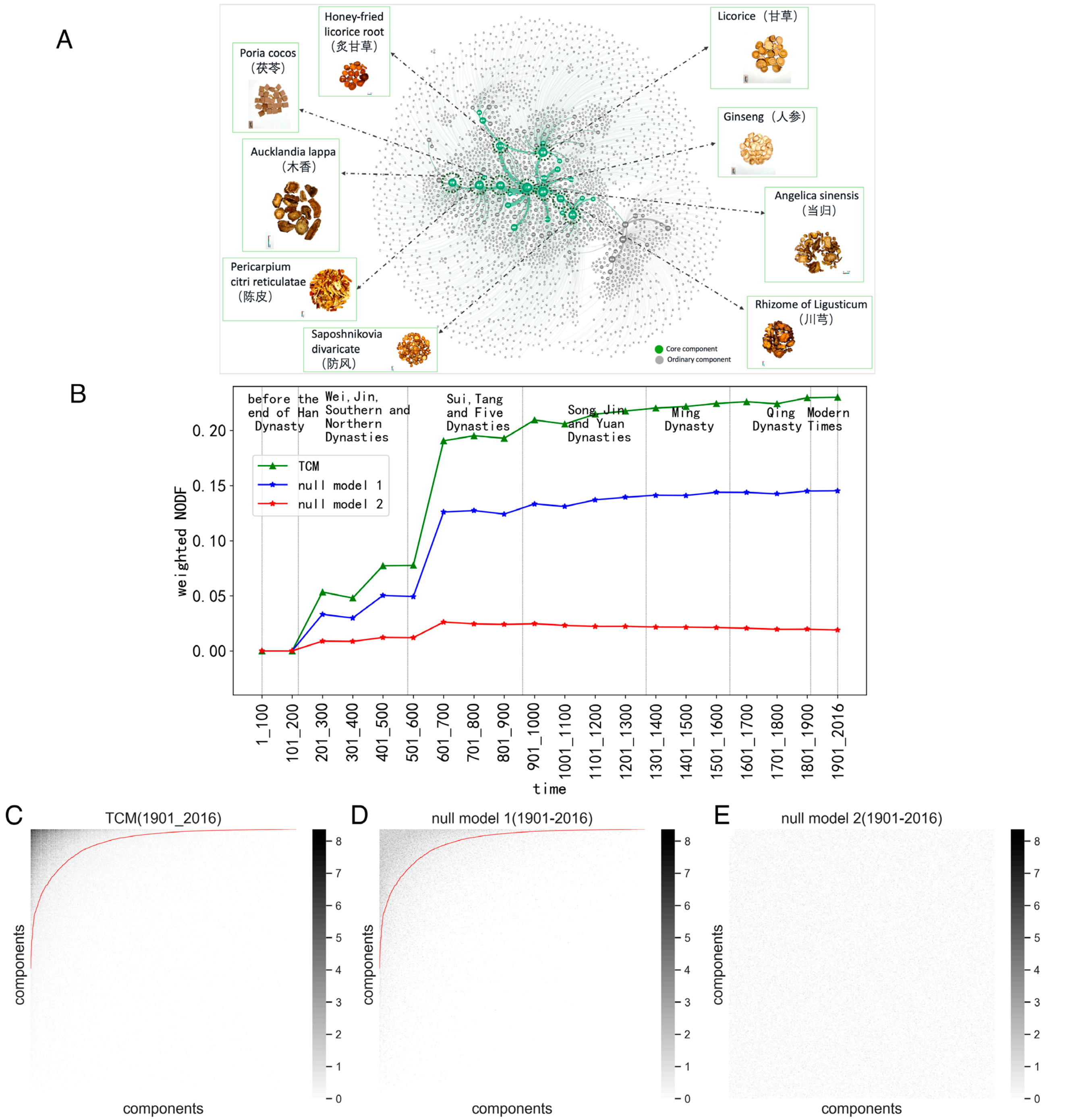


**Fig. 3.** Structure of the weighted TCM component networks. (*A*) Illustrates the weighted TCM component network in the form of a Maximum Spanning Tree, with TCM components as nodes and their co-occurrence in all the 59,063 TCM formulas as edges. The core components are denoted with green circles. The pictures show the top 9 core components. (*B*) Illustrates the weighted-NODFs of the TCM component networks in different periods, comparing with those of the corresponding null models. Null model 1 is generated by reshuffling the weights and edges preserving the degree sequence, and null model 2 is generated by reshuffling the weights and only preserving the numbers of nodes and edges. (*C*) Shows the logarithmic weighted adjacency matrix of the TCM component network during the period of 1901–2016. (*D*) Shows the logarithmic weighted adjacency matrix of the null model 1. (*E*) Shows the logarithmic weighted adjacency matrix of the null model 2. The red lines in (*C*) and (*D*) represent the borders of the perfected nestedness.

tuple with the largest frequency in this path. This example reflects the fact that identifying inheritance relationships among formulas is challenging, although it is an interesting topic (49).

## Discussion

The heavy-tailed distributions of the components' strength, core–periphery structures, and building-out from simple combinations to more complex ones, which were found in the evolution of TCM, are in accordance with the features found in modern technological systems. However, there are significant differences between the evolution of TCM and modern technological systems. For modern technological systems, changes in core components are frequently observed, leading to new technological trajectories (39–41). However, in the evolution of TCM, only slight fluctuations were observed in the core components. This may be due to two reasons. First, these core components can be used to cope with basic or widespread symptoms (*SI Appendix*, Table S4). Second, inheritance

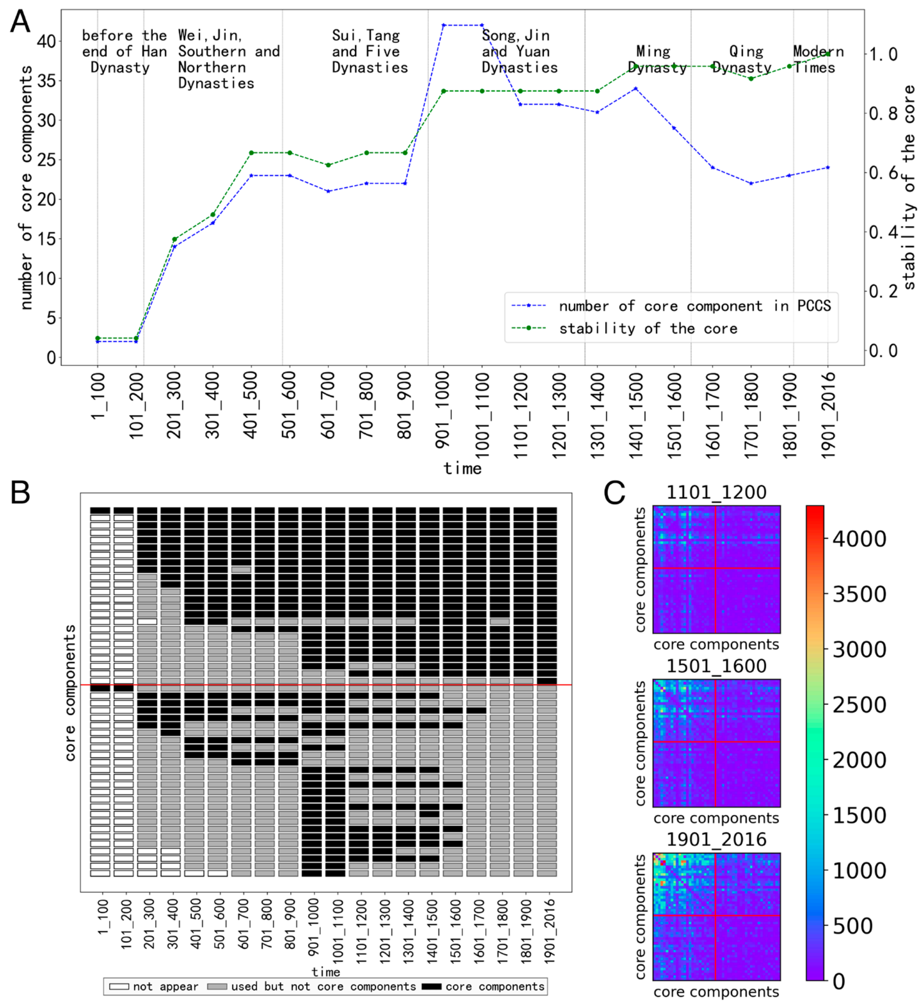


**Fig. 4.** The dynamics of the TCM core components. (*A*) Shows the number of core components in each period and the stability of the core. The *Left* y-axis denotes the number of core components and the *Right* y-axis denotes the stability of the core. The FCCS represents the core components set identified from all the 59,063 TCM formulas, and a PCCS refers to the core components identified from the accumulative TCM formulas by a period. (*B*) Shows the dynamics of the core components. A component in white color denotes it has not yet appeared in TCM formulas, in black color denotes it is identified as a core component in the PCCS, and in gray color denotes it is a noncore component but used in some formulas. The components above the red line are those that appear in the FCCS, while the components below the red line are those appear in the PCCS but not in the FCCS. (*C*) Shows the co-occurrence of pairs of the core components in the three periods (*SI Appendix*, Fig. S17 for all the periods). The *Top left* part segmented by the horizontal and vertical red line represents the co-occurrence of core components in FCCS.

played an important role in TCM development, and classical combinations worked as the basis for the invention of new formulas, which have been widely accepted (50).

The following two implications can be drawn from our study. First, in the further development of TCM, it is suggested to consider the core components we identified in this study as they have been believed and tested to be useful for thousands of years. Second, technological systems developed mainly based on experience might encounter few new technological trajectories, involving modern scientific explorations and engineering, which is becoming prevalent in TCM development (51), is important for maintaining the vitality of such systems.

## Materials and Methods

**Data.** The TCM formula dataset used in this study is constructed by the Institute of Information on Traditional Chinese Medicine and includes 59,063 TCM formulas and 2,917 components. These TCM formulas are collected from 899 classical ancient books, dating mostly from AD 1 to 2016, with two books before AD 1 which included only four formulas and three components. It should be noted that some formulas developed and used in ancient history might not have been documented in existing ancient books; thus, they were lost and not included in the dataset.

**Constructing TCM Component Networks.** We construct TCM component networks for different periods using the accumulated TCM formulas documented at the end of the different periods, with components in formulas as nodes and their co-occurrence in formulas as edges. The weights of the edges were determined by the co-occurrence frequencies of the components in different formulas, and the edges have no direction; thus, TCM component networks are undirected weighted networks (as illustrated in *SI Appendix*, Figs. S1 and S2).

**Exploring Distributions of Component Strength.** We explored whether the distributions of component strength in TCM component networks are heavy-tailed ones with both power-law and lognormal functions (17, 18, 37) (*SI Appendix*, Text). We also used ordinary least squares to estimate the *preferential attachment* (attractiveness) of the components (*SI Appendix*, Text for more details).

**Exploring the Core–Periphery Structures.** We explored whether a TCM component network has a core–periphery structure and how it evolves with the nestedness of the network which is measured with an indicator called NODF (*SI Appendix*, Text for detailed definition of NODF). A high NODF value implies a core–periphery structure (38, 52). To explore whether a TCM component network has a core–periphery structure, we compare its NODF with those of randomly generated null models. We generate a kind of null model (denoted as null model 1) by reshuffling the weights and edges preserving the degree sequence and another kind of null model (denoted as null model 2) by reshuffling the weights

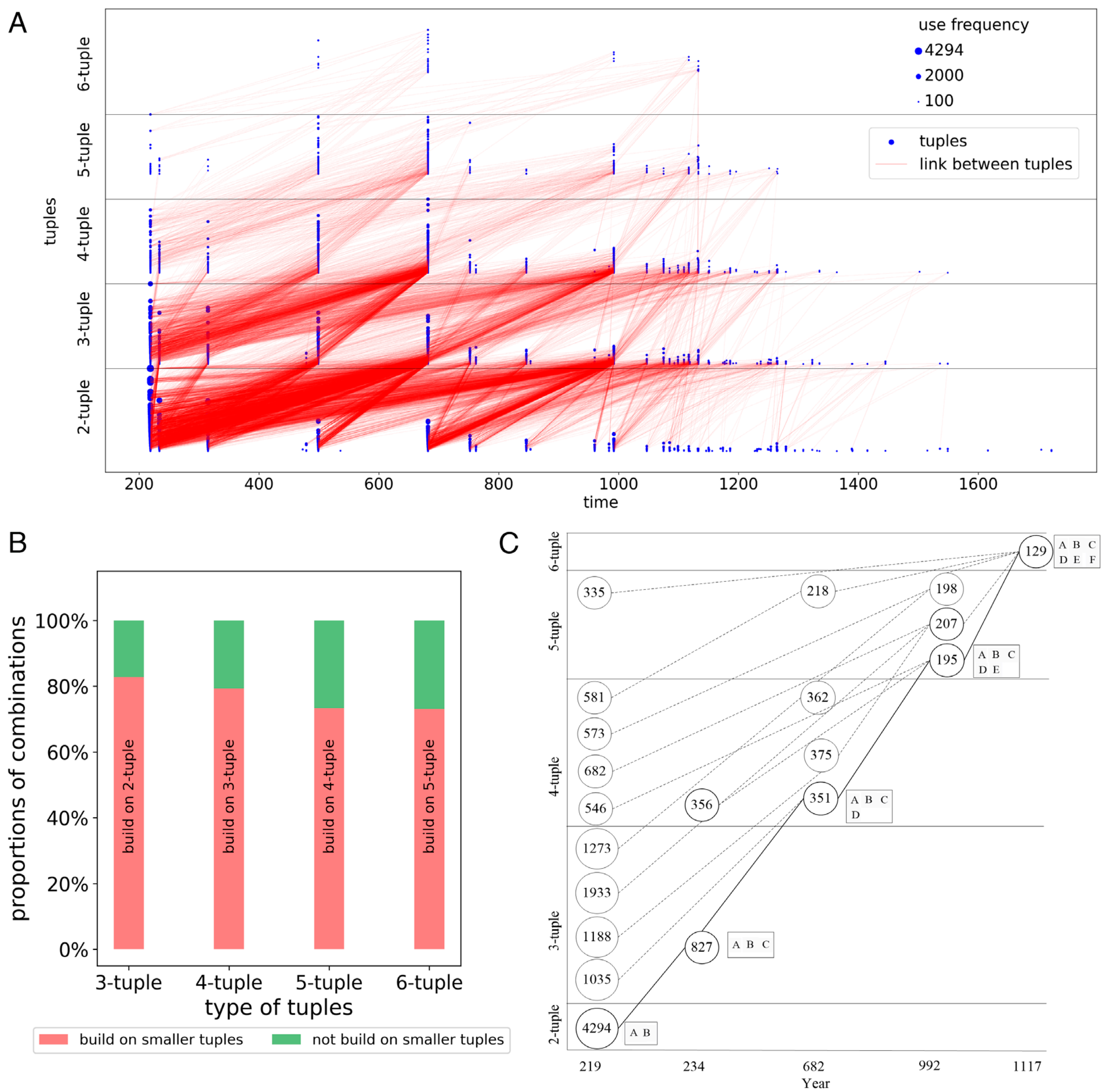


**Fig. 5.** The building-out of tuples. (*A*) Illustration of linkages between the large tuples and one-size smaller ones. The x-axis represents the appearing time of different tuples and y-axis represents sizes of tuples. The blue dots represent tuples, and the larger the size of the dots, the more frequently used of the tuples. A red line linking a large tuple to a small one represents that the large tuple is constituted of the corresponding small one. (*B*) Shows the proportion of large tuples built-out from one-size smaller ones. The x-axis represents different sizes of tuples and y-axis represents the proportion that the large tuples are based on the one-size smaller ones. (*C*) Illustrates a concrete example of the building-out of a 6-tuple from small tuples. A circle represents a tuple with the number denoting its frequency in different formulas, and the horizontal solid lines segment different sizes of tuples. The rectangles around the circle includes the specific components [A (Ginseng), B (Poria cocos), C (Prepared rehmannia root), D (Rhizome of Ligusticum chuanxiong), E (Atractylodes), and F (Paeonia lactiflora)] in different size of tuples. The x-axis represents the time that corresponding tuples appeared, and the y-axis represents different size of tuples. The solid lines link circles represent the path which can trace the 6-tuple back to a 2-tuple, while the dotted lines represent the paths that can be traced back to other small tuples but not to the 2-tuple.

and only preserving the numbers of nodes and edges following Bascompte et al. (53). If the NODF of the TCM component network is significantly higher than that of null model 2, the TCM component network is thought to have high nestedness, which implies that it has a core–periphery structure. If the NODF of the null model 1 is also significantly higher than that of null model 2, it means the core–periphery structure detected (NODF) could be explained by null models taking the degree distribution into account.

**Identifying Core Components.** We adopt the definition of cores proposed by Borgatti and Everett (38), which considers not only the edges of core components but also the cohesion of core components, and a core refers to a set of components that not only link to many peripheral components but are also themselves linked densely. An illustration of an ideal core–periphery structure following this definition is shown in *SI Appendix*, Fig. S3. This study adopts a classical continuous model to identify the core components. Compared with the dichotomous model, the continuous model considers the weights of the edges and measures the coreness value of each node rather than simply assigning the node as a core or periphery (38). As robustness checks, we use the rich club method (42), continuous extension model method (43), and random walk method (44) to identify cores. Details of these methods can be found in *SI Appendix*, Text.

**Exploring the Building-Out of Complex Tuples/Combinations.** To explore, whether some early relatively simple combinations serve as building blocks for more complex ones that would in turn serve as building blocks for even more complex ones in the development of TCM, we adopt the term tuples (54), which denote frequently used combinations of components. Tuples have different sizes in terms of the number of components they comprise: a 2-tuple comprises two components, and a 3-tuple comprises three components.

We extract tuples from all the formulas, and a larger tuple is thought built out from the smaller tuple(s) if the formula including the larger one appeared later than formula(s) including the corresponding smaller ones. We find that the frequency of tuples decreased sharply with an increase in their size (*SI Appendix*, Fig. S4), thus, we limit our study from 2-tuples to 6-tuples. We explore the building-out of large tuples by counting the proportion that large tuples can be traced back to the one-size smaller tuples considering time lag as well as regression analysis on the linkages among tuples with different sizes (*SI Appendix*, Text).

**Data, Materials, and Software Availability.** Data used for the illustrations, regressions, tables, and corresponding codes, which can support all the results in this paper. Data have been deposited in Github (https://github.com/wdmwmw/evolution_TCM) (55).

**ACKNOWLEDGMENTS.** We thank the anonymous reviewers for providing valuable comments and suggestions which are very helpful for improving our study. We acknowledge the support of the NSFC for this research. National Natural Science Foundation of China 71125002 (T.J.M.). National Natural Science Foundation of China 82174534 (Y.Z.).

Author affiliations: [a]Department of Management Science and Engineering, School of Business, East China University of Science and Technology, Shanghai 200237, China; [b]Department of Information, Technology and Innovation, Antai College of Economics & Management, Shanghai Jiao Tong University, Shanghai 200030, China; [c]International Institute for Applied Systems Analysis, Laxenburg A-2361, Austria; [d]Institute of Information on Traditional Chinese Medicine, China Academy of Chinese Medical Sciences, Beijing 100700, China; and [e]China Institute for History of Medicine and Medical Literature, China Academy of Chinese Medical Sciences, Beijing 100700, China

1. P. Anderson, M. L. Tushman, Technological discontinuities and dominant designs: A cyclical model of technological change. *Administrative Sci. Q.* **35**, 604–633 (1990).
2. L. Fleming, Recombinant uncertainty in technological search. *Management Sci.* **47**, 117–132 (2001).
3. M. Ganco, NK model as a representation of innovative search. *Res. Policy* **46**, 1783–1800 (2017).
4. W. B. Arthur, W. Polak, The evolution of technology within a simple computer model. *Complexity* **11**, 23–31 (2006).
5. W. B. Arthur, *The Nature of Technology* (Free Press, New York, 2009).
6. A. A. Yassine, Managing the development of complex product systems: An integrative literature review. *IEEE Trans. Eng. Manage.* **68**, 1619–1636 (2021).
7. S. Valverde, R. V. Sole, M. A. Bedau, N. Packard, Topology and evolution of technology innovation networks. *Phys. Rev. E* **76**, 056118 (2007).
8. S. Yayavaram, G. Ahuja, Decomposability in knowledge structures and its impact on the usefulness of inventions and knowledge-base malleability. *Adm. Sci. Q.* **53**, 333–362 (2008).
9. S. Yayavaram, W. R. Chen, Changes in firm knowledge couplings and firm innovation performance: The moderating role of technological complexity. *Strategic Manage. J.* **36**, 377–396 (2015).
10. M. A. Schilling, E. Green, Recombinant search and breakthrough idea generation: An analysis of high impact papers in the social sciences. *Res. Policy* **40**, 1321–1331 (2011).
11. B. Uzzi, S. Mukherjee, M. Stringer, B. Jones, Atypical combinations and scientific impact. *Science* **342**, 468–472 (2013).
12. C. R. Myers, Software systems as complex networks: Structure, function, and evaluability of software collaboration graphs. *Phys. Rev. E* **68** (2003).
13. A. MacCormack, J. Rusnak, C. Y. Baldwin, Exploring the structure of complex software designs: An empirical study of open source and proprietary code. *Management Sci.* **52**, 1015–1030 (2006).
14. C. Baldwin, A. MacCormack, J. Rusnak, Hidden structure: Using network methods to map system architecture. *Res. Policy* **43**, 1381–1397 (2014).
15. E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, A. L. Barabási, Hierarchical organization of modularity in metabolic networks. *Science* **297**, 1551–1555 (2002).
16. A. L. Barabási, R. Albert, Emergence of scaling in random networks. *Science* **286**, 509–512 (1999).
17. A. D. Broido, A. Clauset, Scale-free networks are rare. *Nat. Commun.* **10**, 1017 (2019).
18. P. Holme, Rare and everywhere: Perspectives on scale-free networks. *Nat. Commun.* **10**, 1016 (2019).
19. B. R. K. Merton, The matthew effect in science, II cumulative advantage and the symbolism of intellectual property. *ISIS* **79**, 606–623 (1988).
20. R. Apa, I. De Noni, L. Orsi, S. R. Sedita, Knowledge space oddity: How to increase the intensity and relevance of the technological progress of European regions. *Res. Policy* **47**, 1700–1712 (2018).
21. K. B. Clark, The interaction of design hierarchies and market concepts in technological evolution. *Res. Policy* **14**, 235–251 (1985).
22. R. M. Henderson, K. B. Clark, Architectural Innovation: The Reconfiguration of Existing Product Technologies and the Failure of Established Firms. *Administrative Sci. Q.* **35**, 9–30 (1990).
23. H. Gatignon, M. L. Tushman, W. Smith, P. Anderson, A structural approach to assessing innovation: Construct development of innovation locus, type, and characteristics. *Management Sci.* **48**, 1103–1122 (2002).
24. T. S. Kuhn, The Structure of Scientific Revolutions (University of Chicago Press, Chicago, 2012).
25. C. Wang, S. Rodan, M. Fruin, X. Xu, Knowledge networks, collaboration networks, and exploratory innovation. *Acad. Manage. J.* **57**, 484–514 (2014).
26. L. Fu, Y. Wang, Y. Zhao, L. Huang, The concept of traditional chinese medicine: History, theory and empirical research. *Chinese J. Med. History* **52**, 195–205 (2022).
27. X. Pan, R. Valerdi, R. Kang, Systems thinking: A comparison between Chinese and western approaches. *Procedia Computer Sci.* **16**, 1027–1035 (2013).
28. J. Zhu, The History of Traditional Chinese Medicine (Xueyuan Press, Beijing, 2009).
29. S. Yan, The History of Traditional Chinese Medical Science (Science Press, Beijing, 2021).
30. Y. Cui, B. Gao, L. Liu, J. Liu, Y. Zhu, AMFormulaS: An intelligent retrieval system for traditional Chinese medicine formulas. *BMC Med. Inform. Decision Making* **21**, 56 (2021).
31. C. A. Hidalgo, B. Klinger, A. L. Barabasi, R. Hausmann, The product space conditions the development of nations. *Science* **317**, 482–487 (2007).
32. J. Huenteler, J. Ossenbrink, T. S. Schmidt, V. H. Hoffmann, How a product's design hierarchy shapes the evolution of technological knowledge—Evidence from patent-citation networks in wind power. *Res. Policy* **45**, 1195–1217 (2016).
33. J. Guan, N. Liu, Exploitative and exploratory innovations in knowledge network and collaboration network: A patent analysis in the technological field of nano-energy. *Res. Policy* **45**, 97–112 (2016).
34. S. Lozano, L. Calzada-Infante, B. Adenso-Díaz, S. García, Complex network analysis of keywords co-occurrence in the recent efficiency analysis literature. *Scientometrics* **120**, 609–629 (2019).
35. J. Choi, S. Yi, K. C. Lee, Analysis of keyword networks in MIS research and implications for predicting knowledge evolution. *Information & Management* **48**, 371–381 (2011).
36. A. Barrat, M. Barthélemy, R. Pastor-Satorras, A. Vespignani, The architecture of complex weighted networks. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 3747–3752 (2004).
37. J. Alstott, E. Bullmore, D. Plenz, Powerlaw: A Python package for analysis of heavy-tailed distributions. *PLoS ONE* **9**, e85777 (2014).
38. S. P. Borgatti, M. G. Everett, Models of core/periphery structures. *Soc. Networks* **21**, 375–395 (1999).
39. M. L. Tushman, L. Rosenkopf, Organizational determinants of technological change: Toward a sociology of technological evolution. *Res. Organ. Behav.* **14**, 311–347 (1992).
40. J. P. Murmann, K. Frenken, Toward a systematic framework for research on dominant designs, technological innovations, and industrial change. *Res. Policy* **35**, 925–952 (2006).
41. G. Dosi, Technological paradigms and technological trajectories. *Res. Policy* **11**, 147–162 (1982).
42. T. Opsahl, V. Colizza, P. Panzarasa, J. J. Ramasco, Prominence and control: The weighted rich-club effect. *Phys. Rev. Lett.* **101**, 168702 (2008).
43. M. P. Rombach, M. A. Porter, J. H. Fowler, P. J. Mucha, Core-periphery structure in networks. *SIAM J. Appl. Math.* **74**, 167–190 (2014).
44. F. D. Rossa, F. Dercole, C. Piccardi, Profiling core-periphery network structure by random walkers. *Sci. Rep.* **3**, 1467 (2013).
45. Chinese Pharmacopoeia Commission, Pharmacopoeia of the People's Republic of China (China Medical Science and Technology Press, Beijing, 2020).
46. https://www.yaozh.com/.
47. J. Wang, Discussion on concept of "the monarch and his subjects assistant and envoy" and applied principle from documents. *Chin. J. Basic Med. Tradit. Chin. Medicine* **10**, 63–65 (2004).
48. Y. Liu, *The Mapping between Traditional Chinese Medicine and Signaling Molecules(* (Cambridge Scholars Publishing, Newcastle, 2022).
49. Y. Zhu, B. Gao, M. Cui, Investigation and formal representation of analogous formulae derivation relationship. *World Chinese Medicine* **12**, 2809–2813 (2017).
50. X. Chi *et al.*, Research progress on classical traditional Chinese medicine formula Lily Bulb and Rehmannia Decoction in the treatment of depression. *Biomed. Pharmacother.* **112**, 108616 (2019).
51. X. Gan, Z. Shu, X. Wang, D. Yan, J. Li, Network medicine framework reveals generic herbsymptom effectiveness of traditional Chinese medicine. *Sci. Adv.* **9**, eadh0215 (2023).
52. M. S. Mariani, Z. M. Ren, J. Bascompte, C. J. Tessone, Nestedness in complex networks: Observation, emergence, and implications. *Phys. Rep-Rev. Sec. Phys. Lett.* **813**, 1–90 (2019).
53. J. Bascompte, P. Jordano, C. J. Melian, J. M. Olesen, The nested assembly of plant-animal mutualistic networks. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 9383–9387 (2003).
54. D. Strumsky, J. Lobo, Identifying the sources of technological novelty in the process of invention. *Res. Policy* **44**, 1445–1461 (2015).
55. M. Wu, T. Ma, Y. Zhu, H. Ren, L. Fu, Data used in the study on evolution of Traditional Chinese Medicine as recombinant inventions. TCMData. https://github.com/wdmwmw/evolution_TCM. Deposited 23 October 2024.