An Explanatory Framework for the Formation of Patent Citation Relationships: An Exponential Random Graph Model Perspective (Postprint)

Authors: Yang Guancan, Chen Liang, Zhang Jing, Li Gang

Date: 2023-07-26T00:00:00+00:00

Abstract

[Purpose/Significance] In recent years, a substantial body of research has emerged focusing on the structural characteristics of patent citation networks. These findings collectively reflect that the formation of patent citation relationships is influenced by relational features beyond attribute characteristics, while existing statistical inference methods based on regression approaches struggle to incorporate these factors into the analytical framework. Therefore, there is an urgent need to explore novel methodologies. [Method/Process] From the perspective of relationship formation, this study represents the formation of patent citation relationships as three generalized relationship formation processes: self-organization influence process, intrinsic attribute influence process, and network covariate influence process, establishes a mapping relationship between relationship formation processes and network configurations, and ultimately forms a comprehensive explanatory framework for understanding the formation of complex patent citation relationships. [Results/Conclusion] This study proposes a comprehensive explanatory framework for understanding the formation of complex patent citation relationships. This framework serves as the theoretical foundation for further constructing network statistical models in the future. Additionally, the explanatory framework contains rich network configuration terms, indicating broad application prospects for exponential random graph models in bibliometrics and scientific network analysis.

Full Text

An Explanatory Framework for Patent Citation Formation: An Exponential Random Graph Model Perspective

Yang Guancan[1], Chen Liang[2], Zhang Jing[2], Li Gang[3] [1]School of Information Resource Management, Renmin University of China, Beijing 100872

[2]Institute of Scientific and Technical Information of China, Beijing 100038
[3]School of Information Management, Wuhan University, Wuhan 430072

Abstract: [Purpose/Significance] Recent research on patent citation network structural characteristics has produced numerous findings, all of which reflect to some extent that patent citation formation is influenced by relational features beyond attribute characteristics. However, existing statistical inference methods based on regression approaches struggle to incorporate these factors into their analytical frameworks, necessitating the exploration of new methodologies. [Method/Process] From a tie formation perspective, patent citation formation can be represented as three broad categories of relational formation processes: self-organizing influence processes, attribute-based influence processes, and network covariate influence processes. This paper establishes a mapping relationship between these formation processes and network configurations, ultimately developing a comprehensive explanatory framework for understanding complex patent citation formation problems. [Result/Conclusion] The paper proposes a complete explanatory framework for understanding complex patent citation formation problems, which serves as the theoretical foundation for future construction of network statistical models. Additionally, the framework contains rich network configuration terms, indicating broad application prospects for exponential random graph models in bibliometrics and scientific network analysis.

Keywords: patent citation formation; explanatory framework; statistical network analysis

Patent citations play a crucial role in technology evaluation because they can trace technological development trajectories, measure technology diffusion and spillovers between countries and regions, assess invention quality and value, and analyze innovation actors' technological strategic behaviors [**?**]. In recent years, scholars have introduced network analysis methods into patent citation analysis, generating a substantial body of research on patent citation networks. These studies have greatly enriched the perspectives of patent citation analysis and overcome the limitations of traditional approaches that relied solely on patent citation frequency. However, the lack of a comprehensive statistical network inference method has significantly restricted researchers' ability to obtain effective analytical results from social network analysis [**?**].

The formation of patent citation networks is a complex problem. Influencing factors may include the network's own evolution, patent attributes, and even external network factors. Traditional statistical inference methods (such as regression) are based on attribute data and assume independence—a prerequisite that is inappropriate for network data where the core object of analysis is relational data. This makes traditional regression analysis incapable of testing hypotheses about relational dependencies.

Consequently, establishing a statistical inference framework that can accommodate the complexity of network research has become a critical issue in current patent citation network studies. Solving this problem requires moving beyond traditional methods based on attribute data. Exponential Random Graph (ERG) models represent a research method focused on tie formation [?]. ERG models are based on relational data, assume dependence, and select local network structures as network statistics to observe the overall structural characteristics of complex networks, thereby obtaining a holistic understanding of network complexity, connectivity, and randomness [?]. Currently, ERG models are in a phase of rapid development.

This paper's research objective is: under the guidance of ERG models, to review key research progress on factors influencing patent citation network formation in existing literature, establish a mapping relationship between this progress and local network structures in ERG models, and develop an explanatory framework to guide models of patent citation formation, laying the foundation for subsequent network statistical inference.

This paper is organized as follows: Section 2 provides a literature review, primarily summarizing factors influencing patent citation formation discussed in current literature. Section 3 proposes a statistical network explanatory framework for patent citation formation, consisting of three components: network self-organization effects, patent attribute effects, and network dyadic covariate effects. Finally, we summarize the explanatory framework, list potential application scenarios, and discuss future research directions.

2 Factors Influencing Patent Citation Formation

2.1 The Problem of Patent Citation Formation

The problem of patent citation formation is not new; it is essentially the same as patent citation network analysis—both explore the structural characteristics of patent citation networks. However, they differ in perspective. Patent citation network analysis argues that macro-level network structure determines individual tie formation, employing structural measures such as whole-network and ego-centric network analysis to observe network structural impacts from a descriptive statistical perspective. In contrast, patent citation formation adopts a micro-level perspective to understand overall network structural characteristics, following social selection process theory's basic assumption that social network structural features are determined by local network substructures formed through the accumulation of individual ties. Specifically: network formation results from local emergence; network tie formation is influenced not only by internal network self-organization and attributes but also by external factors; and multiple social processes occur simultaneously [?]. Based on these assumptions, scholars propose using statistical inference methods to achieve deeper understanding of social network structural characteristics, with ERG models being a mainstream statistical network method [?]. Due to these perspective differences, research on patent citation formation focuses more on statistical inference of network structure. Since patent citation network analysis developed earlier, its theories and methods have become the foundation for current network perspective research on patent citation issues. Therefore, this paper does not strictly distinguish between these concepts.

2.2 Exploration of Influencing Factors and Mechanisms

The influencing factors of patent citation formation constitute a complex problem. Early-stage patent citation analysis primarily focused on inventor attributes' impact on citation formation [?], especially for highly cited patents. Patent citation frequency studies [?, ?, ?] marked the starting point of citation research, establishing through empirical studies that patent attributes—citation frequency—affect citation formation. Researchers also found that temporal factors significantly influence citation formation, particularly in calculating citation frequencies, such as "time truncation," "citation inflation," and citation lag issues [?]. Observational studies suggest geographical factors also affect citation formation, as do inventor and applicant attributes like self-citations vs. others-citations [?], applicant type (e.g., university or research institution) [?, ?], H-index [?], and non-patent literature citations [?].

As research progressed, particularly with the development of patent citation network analysis methods, scholars began focusing more on how network structural characteristics themselves influence citation formation. Researchers have observed multiple dimensions from the network perspective: direct citation networks [?, ?, ?], coupling networks [?, ?], co-citation networks [?], and indirect citation networks [?, ?]. They have validated a series of micro-level measurement indicators applicable to patent citation networks, such as centrality, density, and clustering coefficients, and confirmed that patent citation networks exhibit specific structural features like small-world characteristics [?], power-law distributions [?], rich-club phenomena [?], shrinking diameters [?], and network brokerage and closure structures [?].

As the complexity of patent citation networks became widely recognized, scholars discovered that citation formation is heavily influenced by external factors, such as the combined effects of social structure and intellectual structure [?]. Some researchers have studied the mutual influences between patent citation networks and coupling, co-citation, and indirect citation networks [?, ?], finding that knowledge flows reflected in citation networks have certain associations with geographical distance [?, ?] and proposing that leveraging this multi-relational synergistic network mechanism based on citation relationships can improve patent citation analysis precision, compensate for missing links [?], and map more complete technology frontiers [?, ?].

2.3 The Failure of Traditional Statistical Inference Methods

Scholars have long attempted to apply statistical inference methods to patent citation and network analysis, hoping to guide research through statistical inference. Early statistical inference methods for patent citations were based on attribute data and independence assumptions, primarily using regression analysis such as Logistic regression [?], Logit regression [?, ?], and zero-inflated Poisson models [?]. Theoretical assumptions also imposed numerous limitations on patent citation research. In 1981, L.C. Smith's fifth basic assumption about citations was: "All patents are assumed to be equal" [?]. The core patent metric in citation analysis—citation frequency—ignores other relevant factors like citation relationships, attributes, and external factors, essentially treating citation frequency as independent votes by each inventor on patent quality. As B. Yoon criticized in a classic 2004 article: "Patent analysis only shows direct relationships between patent citation pairs, ignoring the overall relationships among all patents" [?]. The problem with these methods is that regression analysis assumes independence, requiring relative independence among multiple independent variables without high multicollinearity. Otherwise, multicollinearity problems render significance tests meaningless and model prediction functions ineffective. However, the basic unit of patent citation network analysis is the relationship between patents, making independence assumptions inappropriate for relational data.

With the development of network analysis methods, some approaches emerged to test correlations between relational variables and between attribute and relational variables, such as the Quadratic Assignment Procedure (QAP), which studies similarity between one network and multiple networks, thereby addressing correlation evaluation between relational data [?]. This method can also incorporate attribute data through certain network transformations. Early network statistical analysis models like the P1 model adopted similar approaches. However, these methods cannot be applied to more complex scenarios, such as mixed measurements based on weighted and binary attribute features. Therefore, the lack of a comprehensive statistical analysis framework covering multiple influencing factors has limited the development of patent citation network analysis methods.

3 Establishing an Explanatory Framework for Patent Citation Formation

3.1 Exponential Random Graph (ERG) Models

ERG models are research methods focused on tie formation, originating from the social network statistical analysis model (Bernoulli graph distribution) proposed by P. Erdos and A. Renyi in 1959. Subsequently, W. Holland and L. Samuel proposed the dyadic relationship model, a statistical model expressing the relationship between tie occurrence probability and actors' "expansiveness" and "attractiveness" in exponential form. In 1986, O. Frank and S. David introduced Markov dependence, and in 1996, S. Wasserman expanded this model into the ERGM/p* model that can include any statistical configuration in a graph. In 1999, J. Anderson proposed parameter estimation methods for these models, representing important progress [?]. Currently, ERG models are in a phase of rapid development.

ERG models are expandable models that can be adjusted according to research content. Their most general form is:

$$Pr(Y = y) = \frac{\exp(\sum_A \eta_A g_A(y))}{\kappa}$$

where the summation includes all configurations A, _A is the parameter corresponding to configuration A, g_A(y) = {y{ij} A} y_{ij} is the network statistic for the corresponding configuration, and is the normalization constant ensuring the formula is a proper probability distribution [?].

ERG modeling involves observing ordered local network structures, and through specific parameter estimation procedures, the parameter values corresponding to local network structures can be calculated, enabling statistical inference of complex network structures. For domain scholars, the key is to extract a set of theoretically meaningful concepts, hypotheses, algorithms, and indicators from existing literature and establish mapping relationships with local network structures in ERG models.

3.2 Basic Characteristics of Patent Citation Relationships

Patent citation networks are formed by citation relationships between patents. These networks differ from other networks such as collaboration networks in three fundamental characteristics: they are typically unweighted, directed, and cross-sectional. The unweighted characteristic means that patent citation relationships are binary ties linking citing and cited patents—generally only citation or non-citation relationships exist, without considering multiple citations of one patent by another. The directed characteristic arises from the temporal sequence of patent citation behavior: a patent can only be cited by patents published after its own publication date and can only cite patents published before its publication date. Therefore, patent citation relationships are unidirectional and temporally backward. The cross-sectional characteristic means citation networks represent a cross-sectional view of all data up to a certain point in time.

These unweighted, directed, and cross-sectional features give patent citation networks distinct structural characteristics compared to inventor collaboration networks or patent text similarity networks, specifically affecting the composition of dyads, triads, and higher-order structures, which in turn influences corresponding configurations, network statistics, and the explanatory framework. See Figure 1 [Figure 1: see original paper].

3.3 Explanatory Framework for Patent Citation Formation

Based on the basic characteristics of patent citation relationships, combined with existing literature and fundamental relational formation theory, we combine key effects that may influence patent citation formation to develop an explanatory framework for statistical network inference (see Figure 2 [Figure 2: see original paper]). This framework consists of three components: network self-organization influence processes, attribute influence processes, and network dyadic covariate influence processes. These correspond to specific network configurations and effects, where configurations are possible local subgraphs representing regularities in social network structures, and ERG models achieve statistical inference of entire networks by transforming these patterned local subgraphs into network statistics. Network effects represent the causal relationships between network configurations and tie formation that researchers hope to test through ERG models. The framework's core is to bridge existing patent citation network research theory with ERG models, providing a basis for subsequent network statistical inference.

Figure 2 shows that the explanatory framework comprises three parts: network self-organization, attribute factors, and network dyadic covariates. Network self-organization influence processes refer to situations where citation relationships can spontaneously facilitate the formation of other citation relationships, gradually forming an ordered network structure through accumulation. This is an endogenous network effect arising purely from internal relational dynamics. Attribute influence processes emphasize that individual attribute characteristics affect network formation—patents differ in geography, technology fields, scope, quality, and influence, and these differences together with network self-organization characteristics influence citation formation. Additionally, external factors can affect patent citation formation through network dyadic covariates [?], such as semantic relationships between patents, geographical distance, and interpersonal connections. While dyadic covariates have structures similar to network data, their focus is on covariate effects on the network itself. Therefore, a complete explanatory framework for patent citation formation consists of these three components [?].

It should be noted that the framework in Figure 2 assumes directed, unweighted, cross-sectional data, requiring adjustments for specific complex patent citation formation problems. However, the proposed framework already covers major factors currently influencing patent citation formation, such as citation sparsity, preferential attachment, structural holes, transitivity, network closure, geographical (or institutional, disciplinary) homophily or heterophily, and relationships between knowledge networks (semantic networks) and citation networks. ERG models are still rapidly developing, with software packages (like Pnet and Statnet) already providing hundreds of configurations [?], and research on patent citation networks continues to progress, incorporating new theoretical algorithms. Indeed, ERG models' vitality lies in their extensibility.

3.4 Network Self-Organization Influence Processes

Network self-organization is a major research branch in network studies. Under ERG models, network self-organization effects related to patent citation formation mainly include five types: sparsity effects, expansiveness effects, popularity effects, connectivity effects, and transitivity effects. See Figure 3 [Figure 3: see original paper].

3.4.1 Sparsity Effects Sparsity effects refer to the influence of the number of citation relationships on patent citation formation, assuming other factors remain constant. Regarding patent citation quantities, previous literature identifies two characteristics: E. Yan noted that patent citation networks are extremely sparse [?], while A.B. Jaffe proposed the phenomenon of "citation inflation" over time [?]. Although both characteristics exist, sparsity is often more prominent. As time evolves, the growth scale of sparse features (growing as the square of patent numbers) far exceeds the growth in citation quantities [?]. Therefore, the effect of relationship quantity on patent citation formation is typically small or even negative. This effect is considered because it forms the basis of simple random graphs, representing the range of maximum uncertainty and serving as a foundation for comparative studies. In ERG models, the network configuration measuring this sparsity effect is the arc (see Figure 3).

3.4.2 Expansiveness Effects Expansiveness effects refer to the influence of out-degree distribution on patent citation formation, assuming other factors remain constant. The corresponding network configuration is a star structure with arcs emanating from a central node to two or more other nodes. In patent citation networks, this represents a patent's behavior of citing multiple other patents. This expansiveness effect can be measured by node out-degree. From a triadic perspective, the configuration resembles patent coupling structures [?], while from a hub-authority perspective, it can also explain hub characteristics [?]. In ERG models, the network configuration measuring this effect is called "Out-2-star" or "Out-K-stars."

3.4.3 Popularity Effects Popularity effects refer to the influence of in-degree distribution on patent citation formation, assuming other factors remain constant. The corresponding configuration is a star structure with arcs pointing to a central node from two or more other nodes. In patent citation networks, a few patents receive high-frequency citations while others are rarely cited—a phenomenon known as "rich-club," "Matthew effect," or "preferential attachment." From a triadic perspective, this configuration resembles co-citation structures [?] and authority characteristics in hub-authority studies [?]. In ERG models, the configuration measuring this effect is called "In-2-star" or "In-K-stars."

3.4.4 Connectivity Effects Connectivity effects refer to the influence of 2-path configurations on patent citation formation, assuming other factors remain constant. The 2-path configuration (whether simple or multiple) is a special local structure where a central node both receives arcs from and sends arcs to other nodes. In patent citation networks, this represents indirect connectivity between two patents via an intermediate patent, similar to indirect citation structures, and can represent potential "missing links" [?]. From a nodal perspective, the central patent plays a brokerage role [?]. Based on these studies, 2-path configurations can also measure structural holes in patent citation networks. In ERG models, this is called "2-path" or "Multiple 2-path."

3.4.5 Transitivity Effects Transitivity effects refer to the influence of transitive closure on patent citation formation, assuming other factors remain constant. Transitive closure is a well-studied structure in patent citation networks, traditionally measured through clustering coefficients or network clustering features. More comprehensively, transitive closure has two key aspects: (1) it adds an arc to a 2-path configuration, making "missing links" explicit and creating more robust internal relationships, useful for analyzing technology evolution paths [?]; (2) degree distributions in transitive closure are uneven, with some nodes having more in-degree, and this in-degree advantage in transitive closure configurations exceeds that in simple popularity configurations. Therefore, transitivity effects can also identify sources in knowledge flow processes [?]. In ERG models, this is called "transitive triad" or "multiple transitive closure."

3.5 Patent Attribute Influence Processes

Patent attributes also significantly influence citation formation. Research shows patents differ in geography, applicants, inventors, technology breadth and depth, examiner types, technology fields, and dependence on scientific knowledge, all of which affect citation formation to varying degrees.

3.5.1 Sender Effects (or Receiver Effects) In patent citation networks, certain attribute characteristics affect citation formation by influencing the degree to which other patents engage in citation behavior. For directed networks, these effects divide into two types: receiver effects, where a patent's attributes affect its likelihood of being cited more frequently by other patents. For example, U.S. patents may receive more citations due to their disclosure level, quality, and market attention [?]. The corresponding configuration is shown as "receiver" in Figure 4 [Figure 4: see original paper]. Sender effects refer to attributes that make a patent more likely to cite other patents. Research suggests the USPTO's "burden of proof" requirement for patent citations means applicants who fail to disclose prior art face disadvantages in litigation, leading U.S. patents to cite more references on average than other countries [?]. The corresponding configuration is shown as "sender" in Figure 4.

3.5.2 Homophily Effects (or Heterophily Effects) Beyond single patent attributes, attribute characteristics of both citing and cited patents affect citation formation. These are called homophily effects or heterophily effects. Homophily effects mean that when both patents in a citation dyad share the same attribute, it influences citation formation (see "homophily" configuration in Figure 4). Heterophily effects mean different attributes between the two patents affect citation formation (see "heterophily" configuration in Figure 4).

For example, studies show U.S. patents preferentially cite other U.S. patents [**?**], and similar patterns exist for European patents [**?**]. This suggests citation preferences based on common patent office characteristics (homophily) or different patent office characteristics (heterophily). Non-patent citations are also important, as patents may cite both patent and non-patent literature. This can be addressed by treating examiner citations as different node attributes (patent literature vs. non-patent literature) and using homophily or heterophily effects.

3.6 Dyadic Covariate Influence Processes

Research shows that beyond citation relationships, factors like collaboration among patent owners, inventor mobility, semantic similarity between patent texts, technological classification similarity, and geographical distance also affect citation formation. These external features can be incorporated as network dyadic covariates. For instance, patent examiner citations can be tested through covariate effects. Patent citation relationships can be divided into examiner citations and applicant citations, making this a multi-type relational problem (also called multi-relational or multilevel). By constructing network dyadic covariates, we can test examiner citation networks' influence on citation formation. Since examiner and applicant citations don't overlap, measuring dyadic covariates can explain, under constant other conditions, how increased examiner citations affect citation formation relative to applicant citations. Covariate effects refer to how the co-occurrence of dyadic covariate networks with patent citation networks influences citation formation, assuming other factors remain constant. The corresponding configuration is shown as "covariate effect" in Figure 4. For example, if Patent P9 cites Patent P10 in the citation network and they also have text similarity in the semantic similarity network, these two relationships constitute a covariate relationship. In network analysis, covariates can take many forms, including collaboration networks, inventor mobility, semantic similarity, technological similarity, and geographical distance.

4 Applications and Prospects

Theoretically, the citation formation explanatory framework aims to encompass a system that explains all patent citation features and substructures. Under this framework, ERG models can test whether various effects exist and their impact on citation formation. Ideally, the framework should include all substructures and patterns affecting patent citation formation (realized through network configurations). In practice, patent citation data is complex, with multi-relational, heterogeneous, weighted, and temporal factors making it unrealistic to build an all-encompassing framework. Nevertheless, the proposed framework offers comprehensive explanatory power for patent citation relationships as a special network system.

As noted, establishing a comprehensive statistical inference model is important for patent citation networks. We present three examples of how ERG models can address limitations in existing research and provide guidance.

(1) Inference for Complex Network Effects. The highly cited patent problem has long been a research focus, but it results from multiple complex effects operating simultaneously, such as preferential attachment, receiver effects, and homophily effects. Determining the impact of complex network effects is a core academic concern. Preferential attachment suggests new citations form based on existing citation relationships, particularly in-degree distributions; receiver effects suggest patents receive more citations due to individual attributes like country of origin or emerging technology fields; homophily suggests citation behavior tends to occur between patent pairs with similar attributes, such as shared country of origin. A similar example is C. Zhang's use of ERG to test homophily, transitivity, and preferential attachment in scientific collaboration formation [**?**].

(2) Inference for Highly Nested Variables. As reviewed in the literature, patent citation networks have multiple dimensions: direct citation, coupling, co-citation, and indirect citation networks, all influencing citation formation. However, real networks often contain these relationships simultaneously, and no good methods exist for testing such highly nested relationships. ERG models can precisely test effects of highly nested network variables. Indeed, G. Robins et al. provided a set of methods to simultaneously test transitivity, connectivity, and degree distribution features (popularity or expansiveness effects) [**?**].

(3) Weight Setting for Integrating Multi-Relational and Heterogeneous Networks. Integrating multi-relational and heterogeneous networks for multi-perspective observation is a recent research hotspot. Weight setting becomes challenging during this integration. ERG models can obtain parameter estimates for various factors' impacts on overall network formation through model estimation, providing a solution for scientifically setting weights. Relevant research can refer to D. Lusher et al.'s Chapter 10, "Extensions of ERG Models: Multi-Relational and Bipartite Networks" [**?**].

Currently, ERG model research is still in its early stages. As research deepens, network configurations for more complex situations like multi-relational and dynamic networks are being incorporated into ERG models. The explanatory framework proposed here for patent citation formation is exploratory. In the future, we will attempt to introduce more existing theoretical results into ERG models to gain deeper insights into patent citation formation through statistical network inference.

References

[1] OECD. OECD Patent Statistics Manual[M]. Paris: OECD Publishing, 2009: 14-16. [2] Yang GC, Li G, Li CY. Using the comprehensive patent citation network (CPC) to evaluate patent value[J]. Scientometrics, 2015, 105(3): 1319-1346. [3] Van Raan AFJ. Patent citation analysis and its value in research evaluation: a review and a new approach to map technology-relevant research[J]. Journal of data and information science, 2017, 2(1): 545-38. [4] Shu-

mate M, Palazzolo ET. Exponential random graph (p*) models as a method for social network analysis in communication research[J]. Communication methods and measures, 2010, 4(4): 341-371. [5] Robins G, Pattison P, Kalish Y. An introduction to exponential random graph (p*) models for social networks[J]. Social Networks, 2007, 29(2): 173-191. [6] Goodreau SM, Handcock MS, Butts CT. Statnet: software tools for the representation, visualization, analysis and simulation of network data[J]. Journal of statistical software, 2008, 24(1): 1-11. [7] Lusher D, Koskinen J, Robins G. Exponential random graph models for social networks[M]. New York: Cambridge University Press, 2012: 16-28. [8] Harris JK. An introduction to exponential random graph modeling[M]. London: SAGE Publications, 2013: 1-13. [9] Jaffe AB, de Rassenfosse G. Patent citation data in social science research: overview and best practices[J]. Journal of the Association for Information Science and Technology, 2017, 68(6): 1360-1374. [10] Narin F, Carpenter MP, Woolf P. Technological performance assessments based on patents and patent citations[J]. IEEE Transactions on engineering management, 1984, EM-31(4): 172-183. [11] Trajtenberg M. A penny for your quotes: patent citations and the value of innovations[J]. The RAND journal of economics, 1990, 21(1): 172-187. [12] Hall BH, Jaffe AB, Trajtenberg M. The NBER patent citation data file: lessons, insights and methodological tools[R]. Cambridge: National Bureau of Economic Research, 2001. [13] Alcácer J, Gittelman M, Sampat B. Applicant and examiner citations in U.S. patents: an overview and analysis[J]. Research policy, 2009, 38(2): 415-427. [14] Alcácer J, Gittelman M. Patent citations as a measure of knowledge flows: the influence of examiner citations[J]. Review of economics and statistics, 2006, 88(4): 774-779. [15] Jaffe AB, Trajtenberg M. Patents, citations, and innovations: a window on the knowledge economy[M]. New York: MIT Press, 2002: 4-19. [16] Criscuolo P, Narula R, Verspagen B. Role of home and host country innovation systems in R&D internationalisation: a patent citation analysis[J]. Economics of innovation and new technology, 2005, 14(5): 417-433. [17] Owen-Smith J, Powell WW. The expanding role of university patenting in the life sciences: Assessing the importance of experience and connectivity[J]. Research policy, 2003, 32(9): 1695-1711. [18] Guan JC, Gao X. Exploring the h-index at patent level[J]. Journal of the Association for Information Science and Technology, 2009, 60(1): 35-40. [19] Brusoni S, Criscuolo P, Geuna A. The knowledge bases of the world's largest pharmaceutical groups: What do patent citations to non-patent literature reveal?[J]. Economics of innovation and new technology, 2005, 14(5): 395-415. [20] Chang SB, Lai KK, Chang SM. Exploring technology diffusion and classification of business methods: using the patent citation network of e-business patents[J]. World patent information, 2009, 31(2): 115-131. [21] Gress B. Properties of the USPTO patent citation network: 1963-2002[J]. World patent information, 2010, 32(1): 3-21. [22] Sternitzke C, Bartkowski A, Schramm R. Visualizing patent statistics by means of social network analysis tools[J]. World patent information, 2008, 30(2): 115-131. [23] Huang MH, Chiang LY, Chen DZ. Constructing a patent citation map using bibliographic coupling: a study of Taiwan's high-tech companies[J]. Scientometrics, 2003, 58(3): 489-506. [24] Egghe L, Rousseau R.

Co-citation, bibliographic coupling and a characterization of lattice citation networks[J]. Scientometrics, 2002, 55(3): 349-361. [25] Boyack KW, Klavans R. Co-citation analysis, bibliographic coupling, and direct citation: which citation approach represents the research front most accurately?[J]. Journal of the Association for Information Science and Technology, 2010, 61(12): 2389-2404. [26] Hu X, Rousseau R, Chen J. On the definition of forward and backward citation generations[J]. Journal of informetrics, 2011, 5(1): 27-36. [27] Atallah G, Rodríguez G. Indirect patent citations[J]. Scientometrics, 2013, 67(3): 437-465. [28] Hung SW, Wang AP. Examining the small world phenomenon in the patent citation network: a case study of the radio frequency identification (RFID) network[J]. Scientometrics, 2009, 82(1): 121-134. [29] Brantlet PF, Fallah MH. Complex innovation networks, patent citations and power laws[C]//2007 Portland International Conference on Management of Engineering & Technology. Piscataway: IEEE, 2007: 540-549. [30] Smilkov D, Kocarev L. Rich-club and page-club coefficients for directed graphs[J]. Physica A: statistical mechanics and its applications, 2010, 389(11): 2290-2299. [31] Leskovec J, Kleinberg J, Faloutsos C. Graphs over time: densification laws, shrinking diameters and possible explanations[C]//Proceeding of the eleventh ACM SIGKDD international conference. New York: ACM, 2005: 177-187. [32] Wang JC, Chiang CH, Lin SW. Network structure of innovation: can brokerage or closure predict patent quality?[J]. Scientometrics, 2010, 84(3): 735-748. [33] White HD, Wellman B, Nazer N. Does citation reflect social structure? Longitudinal evidence from the "Globenet" interdisciplinary research group[J]. Journal of the Association for Information Science and Technology, 2004, 55(2): 111-126. [34] Shibata N, Kajikawa Y, Takeda Y. Comparative study on methods of detecting research fronts using different types of citation[J]. Journal of the Association for Information Science and Technology, 2009, 60(3): 571-580. [35] Yan E, Ding Y. Scholarly network similarities: How bibliographic coupling networks, citation networks, co-citation networks, topical networks, co-authorship networks, and co-word networks relate to each other[J]. Journal of the Association for Information Science and Technology, 2012, 63(7): 1313-1326. [36] Thompson P. Patent citations and the geography of knowledge spillovers: evidence from inventor- and examiner-added citations[J]. Review of economics and statistics, 2006, 88(2): 383-388. [37] Bottazzi L, Peri G. Innovation and spillovers in regions: evidence from European patent data[J]. European Economic review, 2003, 47(4): 687-710. [38] Chen DZ, Huang MH, Hsieh HC. Identifying missing relevant patent citation links by using bibliographic coupling in LED illuminating technology[J]. Journal of informetrics, 2011, 5(3): 400-412. [39] Morris SA, VanderVeer Martens B. Mapping research specialties[J]. Annual review of information science and technology, 2008, 42(1): 213-295. [40] Braam RR, Moed HF, Van Raan AFJ. Mapping of science by combined co-citation and word analysis. I. Structural aspects[J]. Journal of the Association for Information Science and Technology, 1991, 42(4): 233-251. [41] Criscuolo P, Verspagen B. Does it matter where patent citations come from? Inventor vs. examiner citations in European patents[J]. Research policy, 2008, 37(10): 1892-1908. [42] Bakker J. The

log-linear relation between patent citations and patent value[J]. Scientometrics, 2016, 110(2): 1-14. [43] Lee YG, Lee JD, Song YI. An in-depth empirical analysis of patent citation count using zero-inflated count data model: the case of KIST[J]. Scientometrics, 2007, 70(1): 27-39. [44] Nicolaisen J. Citation analysis[J]. Annual review of information science and technology, 2007, 41(1): 609-641. [45] Yoon B, Park Y. A text mining based patent network: analytical tool for high-technology trend[J]. The journal of high technology management research, 2004, 15(1): 37-50. [46] Lee WJ, Lee WK, Sohn SY. Patent network analysis and quadratic assignment procedure to identify the convergence of robot technologies[J]. PloS ONE, 2016, 11(10): e0165091. [47] Snijders TAB, Pattison PE, Robins GL. New specifications for exponential random graph models[J]. Sociological methodology, 2006, 36(1): 99-153. [48] Rosek MJY, Howard M, Cox PAHNEE. Understanding network formation in strategy research: exponential random graph models[J]. Journal of statistical software, 2008, 24(4): 1548-7660. [49] Robins G. Doing social network research[M]. London: SAGE, 2015: 81-85. [50] Morris M, Hunter DR, Handcock MS. Specification of exponential-family random graph models: terms and computational aspects[J]. Journal of statistical software, 2008, 24(4): 1548-7660. [51] Batagelj V. Efficient algorithms for citation network analysis[EB/OL]. [2017-12-06]. https://www.researchgate.net/publication/1956732_{{{Efficient}}*{{Algorithms}}}*{{{for}}*{{Citation}}}*{{{N
[52] Dernis H, Hoisi K, Webb C. Analysing European and international patent citations[EB/OL]. [2017-12-06]. http://www.oecd-library.org/science-and-technology/analysing-european-and-international-patent-citations{883002633010}. [53] Meyer M. What is special about patent citations? Differences between scientific and patent citations[J]. Scientometrics, 2000, 49(1): 93-123. [54] Cockburn IM, Kortum S, Stern S. Are all patent examiners equal? The impact of examiner characteristics[R]. Cambridge: National Bureau of Economic Research, 2002. [55] Zhang C, Bu Y, Ding Y. Understanding scientific collaboration: homophily, transitivity, and preferential attachment[J]. Journal of the Association for Information Science and Technology, 2018, 69(1): 72-86. [56] Robins G, Pattison P, Wang P. Closure, connectivity and degree distributions: exponential random graph (p*) models for directed social networks[J]. Social Networks, 2009, 31(2): 105-117.

Author Contributions

Yang Guancan: Responsible for conceptual framework construction, main content writing, and experimental results analysis. Chen Liang: Responsible for literature review, participated in experimental data collection, processing, and results analysis. Zhang Jing: Participated in experimental data collection and results analysis. Li Gang: Provided guidance and suggestions on paper concept and framework.

Note: Figure translations are in progress. See original paper for figures.