

# 基于创新网络社区主题发现和 S 曲线的技术预测

李 蕾<sup>1</sup>, 宋俭宁<sup>2</sup>, 宋天华<sup>1\*</sup>

(1. 哈尔滨工业大学 图书馆, 哈尔滨 150090; 2. 东北农业大学, 哈尔滨 150038)

**摘要:** [目的 / 意义] 从中观视角进行技术预测, 对于企业、行业和国家至关重要, 建立一套适合方法很有必要。[方法 / 过程] 本文综合应用专利引文网络社区甄别和文本挖掘方法, 建立了兼顾网络连接性和孤立节点的最大分割主题分类方法, 对技术创新社区进行侦测, 并应用 S-曲线模型对分离出来的创新社区进行发展趋势预测。[结果 / 结论] 对杂交水稻育种创新领域的实证研究表明, 研究提出的方法既丰富了技术预测理论和方法, 也从中观视角给出了相关创新机构和国家层面的策略建议。

**关键词:** 技术预测; 创新网络; 网络社区发现; 文本挖掘; 主题分类; 专利引文; 杂交水稻

**中图分类号:** G303

**文献标识码:** A

**文章编号:** 1002-1248 (2021) 04-0045-13

**引用本文:** 李蕾, 宋俭宁, 宋天华. 基于创新网络社区主题发现和 S 曲线的技术预测[J]. 农业图书情报学报, 2021, 33 (4): 45-57.

## Technology Forecasting Based on Topic Identification of Online Innovation Communities and S-Curve

LI Lei<sup>1</sup>, SONG JianNing<sup>2</sup>, SONG TianHua<sup>1\*</sup>

(1. Library Harbin Institute of Technology, Harbin 150090; 2. Northeast Agricultural University, Harbin 150038)

**Abstract:** [Purpose/Significance] It is very important for enterprises, industries and countries to carry out technology prediction from a meso-perspective, so it is necessary to establish a set of suitable methods. [Method/Process] Applying text mining and detection of patent citation network, this study built subject clustering based on biggest fragmenting to detect innovation communities and with s-curve to forecast technology development of innovation communities. [Results/Conclusions] The empirical research on hybrid rice breeding innovation field shows that the

收稿日期: 2020-09-25

基金项目: 国家自然科学基金项目“基于引文网络图数据挖掘的热点技术领域预测研究”(71003020)

作者简介: 李蕾 (ORCID: 0000-0002-5503-3358), 馆员, 研究方向为数字文献资源和竞争情报。宋俭宁 (ORCID: 0000-0002-7206-536), 本科生, 研究方向为农学

\*通信作者: 宋天华 (ORCID: 0000-0003-1549-3613), 博士, 研究馆员, 研究方向为图书情报、技术创新和知识产权。E-mail: tianhua\_s@163.com

proposed method not only enriches the theory and method of technology prediction, but also gives the relevant innovation institutions and national level strategic suggestions from the meso-perspective.

**Keywords:** technological forecasting; innovation network; network community discovering; text mining; subject clustering; patent citation; hybrid rice

## 1 引言

技术预测对于企业、行业和国家调整技术方向, 瞄准技术竞争制高点, 以及制定技术战略和技术规划具有重要意义。国内外对技术预测的研究, 应用的方法主要有德尔菲法、专家咨询法、头脑风暴法、需求分析法、关键技术选择法、趋势外推法、文本挖掘法、层次分析法、专利分析法、STEEP (社会、技术、经济、环境、政治) 分析法、SWOT (优势、劣势、机会、挑战) 分析法、决策树法、实地考察法、技术路线图等<sup>[1]</sup>。日本在第八次产业技术预测中, 出现了引文预测法, 它利用科技论文的引文数据, 通过统计方法处理后进行预测<sup>[2]</sup>。同样, 对技术预测的分类也有很多种, 如 MARIANO 把技术预测方法分为探索性方法和规范性方法<sup>[3]</sup>, 探索性方法是根据目前的情况推断未来

发展, 而规范性方法是依据未来的光景反推目前的做法。PORTER 等在 1991 年把技术预测方法分为直接法、相关法和结构法<sup>[4]</sup>, 但是流传比较广, 比较被认可的是还是 PORTER 等<sup>[5]</sup>根据第一届塞维利亚技术未来分析研讨会<sup>[6]</sup>提出“九类-51 种方法体系”基础上, 提出的“13 大类-48 种未来技术分析方法”, 如表 1 所示。

综观各类学者的观点, 多是根据各自需要执其一种方法或多种方法展开技术预测研究。自从 2000 年之后, 多方法融合以及大数据已经成为技术预测研究的主流。

专利是描述与刻画技术创新的一个成熟和客观的指标。专利的专利引文是对其技术的发展和运用, 引文越多, 则表明该专利的影响力越大, 技术扩散的越好<sup>[7]</sup>。最早提出专利引证概念的是在 1949 年<sup>[8,9]</sup>, 但是直到 2000 年之后, 随着专利电子数据库的陆续建立,

表 1 PORTER 等 13 大类—48 种技术未来分析方法

Table 1 Future analysis methods of 13 categories including PORTER - 48 kinds of technology

编号	方法族	代表方法
1	情报与监测	技术监测、技术挖掘
2	创造性方法	TRIZ、未来研讨、愿景
3	描述性方法	文献计量学、影响清单、未来状态指数、多视角评估
4	统计分析方法	风险分析、相关系数
5	路线图	回溯法、技术/产品路线图、科学路线图
6	情景分析	情景管理、定量情景分析
7	价值/决策/经济分析	成本-效益分析、层次分析法、数据包络分析、多目标决策分析
8	矩阵方法	类推、形态分析、交叉影响分析
9	趋势分析	生长曲线、领先指标、包络曲线、长波模型
10	专家意见	调查、德尔斐法、专题小组、参与性方法
11	模型与模拟	创新系统描述、复杂自适应系统模型、混沌模型、技术扩散和技术替代分析、投入-产出模型、基于代理的模型
12	逻辑/因果关系分析	需求分析、机构分析、利益相关者分析、社会影响评估、缓冲战略、可持续性分析、行动分析 (政策评估)、相关树、“未来轮”
13	组合	情景模拟 (博弈)、趋势影响分析

专利引文的研究才逐渐发展起来。DAIM 等用专利计量方法设计了预测技术创新的方法<sup>[10]</sup>。LAI 等用专利共引方法建立了一个新的专利分类系统, 便于进行研发规划、技术定位、专利战略的决策和技术预测<sup>[11]</sup>。YOUTIE 等利用专利与专利引文探索纳米技术的技术性质, 以界定其是否为通用目的技术 General Purpose Technology (GPT), 以了解技术的长期远景<sup>[12]</sup>。一些学者研究了如何通过专利地图为技术创新寻求机会以制订新的技术创新战略<sup>[13,14]</sup>。另外还有运用专利分析方法预测技术发展路径<sup>[15]</sup>。李婧等对 PEM 燃料电池专利引用情况进行实证分析, 从核心技术、技术趋势、竞争对手、专利引证率方面阐述了专利引文在技术竞争力研究中的重要作用<sup>[16]</sup>。康宇航等初步研究了技术跟踪的结构和过程, 设计了基于专利引文的技术跟踪工具, 为技术的发展战略提供决策支持<sup>[17]</sup>。

CHANG 等利用专利引文网络的分析方法探索了商业分类方法的扩散规律。他们认为在众多技术预测指标中, 专利和专利引用是重要的指标<sup>[7]</sup>。WARTBURG 利用多级专利引文分析, 测量创新的过程, 指出单级专利引文分析不足以揭示技术发展具体路径, 应该依据文献耦合、共引、直接引用和间接引文等方法进行技术路径分析<sup>[18]</sup>。HUANG 等利用专利及专利引文研究了 2003 年纳米技术的最新发展, 从国家、机构和技术领域的角度展示了纳米技术的最新进展, 用引文网络揭示了技术扩散<sup>[19]</sup>。MINA 等通过专利引文网络和纵向路径方法研究了药品技术创新知识的涌现、成长和转变<sup>[20]</sup>。

王宏起运用专利共现网络和 Katz 链路预测方法, 对未来技术融合方向进行预测<sup>[21]</sup>。史璇等综合了文本挖掘、网络分析、文献计量和专利分析方法, 提出一套预测燃料电池技术机会的预测方法, 预测技术领域创新机会、识别可能应用的行业和商业领域的发展趋势<sup>[22]</sup>。覃兴等结合专利网络及专利地图方法, 提出专利地图分析技术预测分析框架, 对电力电缆的技术发展路线做出预测<sup>[23]</sup>。吴颖文等基于专利共现网络和回归分析, 对未来可能涌现的共性技术进行了预测<sup>[24]</sup>。CIGDEM 通过语义网络预测技术创新的发展阶段, 揭

示创新的动态演化<sup>[25]</sup>。

对专利网络开展预测研究比较多的是创新路径的预测研究, 如 VERSPAGEN 研究了燃料电池与数据通信标准技术演化, 绘制出完整的技术路径与专利引文网络图谱, 认为它们的技术弹道具有选择性的和累积性的特质, 在不同聚集水平上是非线性发展, 并揭示了网络中的重要技术(节点)和技术演化路径<sup>[26,27]</sup>。结合传统技术预测方法和语义挖掘技术, 研究与评价技术的创新路径<sup>[28]</sup>、技术主路径的演化与预测<sup>[29]</sup>、技术研究方向的识别<sup>[30]</sup>、技术涌现的早期识别<sup>[31]</sup>和对技术创新路径的揭示<sup>[32]</sup>。还有用技术演化路径预测纳米制药技术的发展<sup>[33]</sup>, 预测太阳能电池技术路径的发展<sup>[34]</sup>, 激光加工技术的发展<sup>[35]</sup>, 3D 打印技术的发展和演变等<sup>[36]</sup>。杨中楷等探索了不同技术领域间的专利知识流动的宏观网络特征, 并揭示技术的演进路径<sup>[37,38]</sup>。蔡力伟等一些学者运用专利文献和复杂网络方法, 对技术机会开展了链接预测<sup>[39-41]</sup>。

应用专利网络对技术发展趋势预测一直是研究热点。HEON 等利用社会网络分析方法预测了制药设备的核心技术<sup>[42]</sup>。李佳佳等比较了局部离群因子算法和社会网络分析方法, 在中国风能领域做了技术预测, 结论基本是一致的<sup>[43]</sup>。YOU 等基于专利引文网络, 对相干光发生器的技术发展趋势进行了预测<sup>[44]</sup>。周源等综合了文献计量、主题模型、社会网络分析等方法, 研究技术的演进趋势<sup>[45]</sup>。廖岭等运用文本挖掘和主路径分析方法, 开展了技术发展趋势的实证研究<sup>[46]</sup>。高建等综合 TRIZ 理论与逻辑曲线方法, 预测了破胎器新技术可能出现的时机和发展趋势<sup>[47]</sup>。一些学者, 如文雄辉<sup>[48]</sup>和钱娇等<sup>[49]</sup>基于 SAO 链文本结构模式揭示专利的文本语义, 对比专利间技术相似度, 构建专利网络, 分析技术创新的演化路径, 并结合 TRIZ 理论预测技术的未来发展趋势。张振刚等<sup>[50]</sup>基于聚类方法建立知识网络, 用以预测未来热点技术和技术的发展趋势。WEN 结合专利网络地图和 SAO 语义分析方法, 对产业机器人的技术演化和发展趋势进行了预测<sup>[51]</sup>。尹忠博等<sup>[52]</sup>全面总结了基于数据的技术预测现状和发展趋势。

上述研究表明, 利用专利或专利引文的研究方法,

能揭示技术的涌现、成长、转变、扩散和演化，也可以跟踪技术发展和技术预测。

虽然国内外对于技术未来和技术预测的研究有很多，基于专利引文研究技术演化脉络以及技术未来发展趋势的研究也不少。但是，这些研究或者是针对特定技术领域的微观分析，如路径识别、链路预测和机会识别等；或者是对特定技术的宏观研究，如技术演进图、未来技术机会和空白技术等，但很少有中观视角对技术预测的研究。本文则是从中观技术创新社区视角展开技术预测研究，综合运用专利引文网络、网络社区发现、文本主题分类和 S-曲线等方法，提出从中观视角研究技术未来发展的一套方法。

## 2 研究方法 with 数据

本文综合运用专利引文理论、网络拓扑和动态演化理论、网络社区侦测方法、语义主题分类和 S-logistic 曲线预测模型，对专利与专利之间通过引用关系形成的、以专利技术为节点、它们之间的引用关系为边的技术创新网络的技术创新进行数据挖掘和发展预测，以把握技术的发展进程和方向。

本文实证数据为杂交水稻育种的专利及其引文数据，数据来源于德温特专利数据库。数据获取方法是，确定“hybrid rice”为检索词，选取的检索项是主题途径，它包括专利题名与专利文摘，时间跨度是 55 年，既从 1963—2018 年。最后共得到 922 件杂交水稻专利，专利引文 565 件。从整体发展曲线（图 1）和专利主题分析，杂交水稻技术的发展经历了 3 个阶段。第一个阶段，1963—2001 年前后，在这个阶段，杂交水稻从萌芽、起步、迅速发展，逐渐发展成为相对完整的水稻培育技术领域，专利文献主题比较多样。第二阶段，2002—2014 年前后，这个阶段杂交水稻可能开始向一些特殊品性方向迅猛发展，专利主题中的如超高产量、高品质、高抗病性等频次增加较快。第三阶段，2014 年到现在，杂交水稻进入技术相对成熟阶段，一些具备特殊品性的杂交水稻育种技术开始接近成熟。

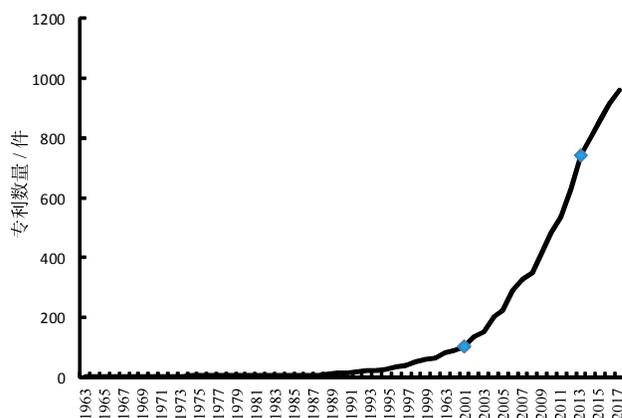


图 1 杂交水稻技术发展

Fig. 1 Growth of hybrid rice

### 2.1 最大分割主题分类法

网络社区 (Network Community) 主要是指网络中的子网络，子网络中的节点相互连接比较紧密，而子网络中的节点与外面的节点连接比较少或者没有连接<sup>[53]</sup>。网络社区的侦测主要是基于网络中节点的连接性进行聚类分区的方法。具体地说，就是要抽取网络中的重要模块。网络社区的发现对于预测网络成长过程中涌现的重要的、关键的性质具有重大意义。一些社区发现的算法已经发展起来，如主要应用于大规模物理网络的层次的 GN 算法 (Girvan and Newman's Divisive Algorithm)<sup>[54]</sup>、K-近邻算法 (Kernighan-Lin)<sup>[55]</sup>、用于社区图划分的谱方法 (Spectral Methods)<sup>[56,57]</sup>、贪婪算法 (Newman's Greedy Optimization of Modularity)<sup>[58]</sup> 和应用于多水平网络划分的马尔可夫聚类算法 (Markov Clustering)<sup>[59,60]</sup>等。

本文采用的数据为文本数据，数据有约 1 500 个样本量，对数据的语义理解比较容易处理。网络中还存在着一些孤立点也可能对技术的发展具有作用，而且这些创新的孤立，有可能是数据系统规范要求造成的，如中国专利不要求必须标注引文。因此，忽略了孤立点，也可能造成数据不完整。但上述提到的那些算法都不适合这些孤立点的社区划分。为避免了上述情况发生，有必要采用新的方法，本文主要是基于独立创新的技术主题，比较它们与网络中有连接的技术创新社区的主题，按照匹配原则归属它们的社区划分。

本文设计的社区发现方法命名为最大分割主题分类法, 分为网络社区发现与主题分类两个步骤。

(1) 网络社区发现。对数据集生成的技术创新网络直接进行分割, 规则是最大化节点的链接, 其定义为: 如果在子网络中, 节点之间存在连接, 而这些节点与子网络外的节点不存在连接, 则该子网络为网络中的一个网络社区  $N_i$ 。

(2) 文本挖掘与主题分类。首先对网络社区 ( $N_1, N_2, \dots, N_i$ ) 进行文本挖掘与语义分析, 利用词频分析和 Wordnet 专业词典提取反映每个社区创新技术主题的高频核心关键词  $K_n$ , 建立每个网络创新社区的主题关键词集  $\{K_{nj}\}$ 。然后, 抽取并建立反映每个孤立点主题的高频核心关键词集  $\{K_{iso}\}$ , 比较每个孤立点的关键词集  $\{K_{iso}\}$  与每个创新社区的关键词集  $\{K_{nj}\}$  的匹配度  $M_i$ , 如果  $M_i \geq 60\%$ , 则孤立点  $i$  归入网络创新社区  $N_i$ , 如果  $M_i = 0$ , 则孤立点  $i$  不属于任何网络社区。这样就保证了再网络社区数据尽可能完备的情况下, 对技术创新的发展趋势进行预测。

## 2.2 成长曲线预测模型

S-logistic 曲线模型也叫 S-曲线模型, 是生物学家 VERHULST 在 1845 年研究人口增长时建立和发展起来的<sup>[6]</sup>。S-曲线模型广泛应用在经济、政治、人口统计、化学、植物增长动态、昆虫生态和森林增长等领域。S-曲线模型如下:

$$Y_t = \frac{k}{1 + e^{-a(t-b)}} \quad (1)$$

其中,  $Y_t$  和  $t$  分别代表 S-曲线发展变量和时间变量,  $a$  和  $b$  是模型参数, 决定着曲线的形状与位置, 而  $k$  是一个代表当  $t$  达到极限时  $Y_t$  极限固定值。每个 S-曲线都有 3 个阶段: 萌芽期、成长期和成熟期, 如图 2 所示。

I、II 和 III 分别代表萌芽期, 成长期和成熟期。也有的 S-曲线模型分为 4 个阶段, 在成熟期之后, 还有衰退期。本文的研究重点是探讨技术的前 3 个成长阶段, 而对技术衰退期不做深入探讨。

技术生命周期理论已经被广泛的接受和采用, 它

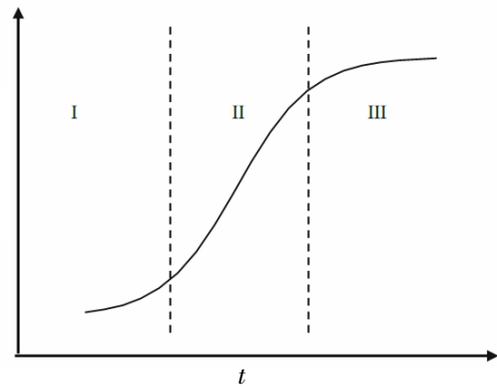


图 2 S-曲线模型

Fig. 2 S-curve model

反映了技术成长最主要的特征, 比较符合技术的成长演化形态。一般地情况下, 技术周期分为 4 个阶段, 初始阶段、发展阶段、成熟阶段和成熟后阶段。而体现技术生命周期的技术成长曲线与 S 曲线比较相似, 因此用 S-曲线模型预测技术的发展趋势是可行的, 也是适合的。当然存在着技术突变和其它变化, 可能会产生一个升级的 S-曲线或其他成长曲线模型, 那么, 可以先拟合出最优符合成长曲线, 然后运用该曲线模型进行预测。本文对于技术发展趋势的研究正是基于技术周期理论的前提条件, 对代表技术创新的专利历史数据进行分析, 最后选用了最符合杂交水稻社区成长的 S-曲线模型外推出技术的发展趋势。整个预测过程如图 3 所示。

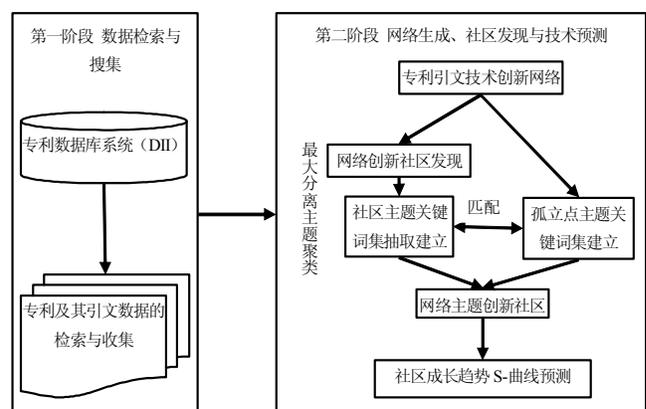


图 3 技术创新发展趋势预测过程

Fig.3 Forecasting process of developing trend for technology innovation

### 3 实证研究

#### 3.1 专利引文技术创新网络的社区侦测

接下来，利用杂交水稻专利及其引文数据，通过引用关系生成杂交水稻专利引文技术创新网络。然后利用本文设计的社区分割分类方法对网络进行社区发现与划分，划分出杂交水稻的技术创新社区。

##### 3.1.1 专利引文技术创新网络的社区分割

首先通过专利、专利引文以及它们之间的引用关系，生成包含 922 杂交水稻专利和 565 件专利引文的杂交水稻技术创新网络。然后用前面设计的最大分割主题分类法对该网络进行社区探测。约简去比较小的创新网络社区之后，得到 3 个创新社区的网络图，如图 4 所示。最早形成的也是最大的社区是第一个创新社区（图中左边红色圆点形状节点的子网络），该社区由 654 件专利互相连接而成，通过文本挖掘与人工判别，其主要研究方向是具有“高产量高抵抗力”特点的杂交水稻技术。图中右上（绿色方块形状节点的子网络）的是第二个创新社区，该社区有 300 件连接的专利，同样的文本挖掘与识别方法，判定该创新社区

的特点是培育具有“高品质高抵抗力”种子的杂交水稻技术。最下方的（蓝色三角形节点的子网络）是第三个创新社区，该社区的主要创新特点是，估测与区别与杂交水稻有很大关系的能育性基因的技术。第三个创新社区的规模和发展速度上都比较低，数据量满足不了预测的规模要求，而且从 2005 年之后，数据出现了中断，无法进行预测。因此，我们主要对第一和第二技术创新社区进行实证研究。

##### 3.1.2 专利引文技术创新网络社区的主题分类

对孤立专利的社区划分要通过前面设计的最大分割主题分类法的第二步——主题分类分析来完成。具体是通过对专利数据的文本分析，主要包括专利名称与专利摘要的文本主题分析与词频分析，运用成熟的词频分析软件“英语词频分析器”和 Wordnet 语义词典，经过分析、清洗、分类等提取出代表各技术创新社区的高频主题关键词，形成社区主题词集。第一创新社区主要的研究方向是培育具有“高产量高抗性”的杂交水稻种子技术，其汉英对照的主题词集合如表 2 所示。

第二创新社区主要的研究方向是培育具有“高品质高抗性”的杂交水稻种子技术。它汉英对照的主题词集合如表 3 所示。

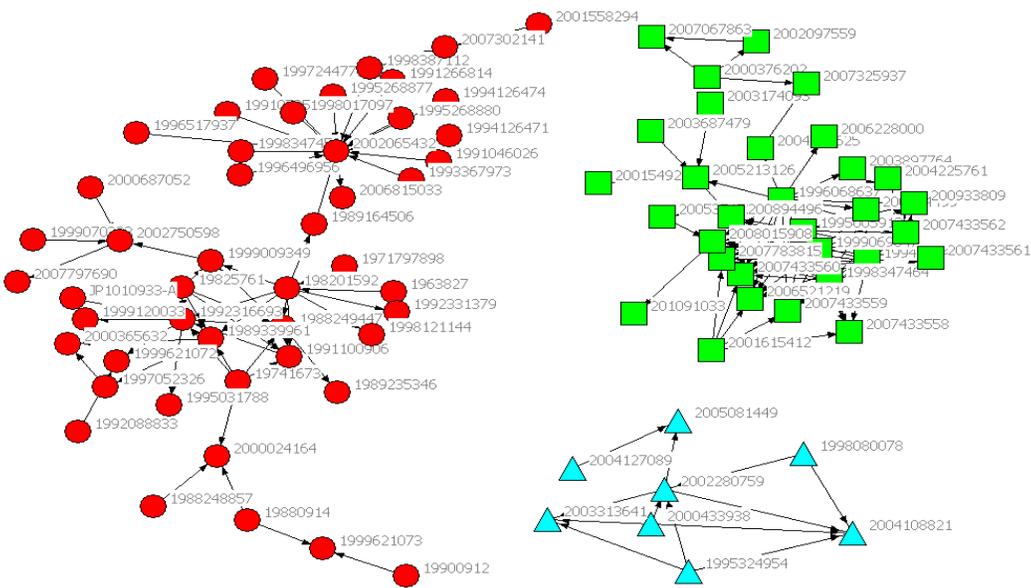


图 4 杂交水稻主要的技术创新社区

Fig. 4 The main innovation communities of hybrid rice

表 2 杂交水稻第一创新社区主题词集

Table 2 Subject words set of the first innovation community for hybrid rice

主题分类	主题词
杂交水稻育种	Rice、Hybrid*、Cross、Mating、Seeds、Breed*
育种技术与方法	Linear、System、Line
系、种、亚种、细胞、分子、基因等	Cell、Cytoplasm
授粉、选择、储存、恢复、存活等	Indica、Japonica、Sativa、Oryza Gene、Isogenic、Genomes、Molecular Pollen、Choose、Select、Restore、Recovery、Surviving、Retained
作物抗性	Resistance、Resisting、Bacterial、Disease、Killed
抗疾病、抗病菌、抗虫害、抗低温, 多年生等	Killing、Pathology、Winter、Warm、Perennial
作物品质	High、More、Increase、Yield
高产量	

表 3 杂交水稻第二创新社区主题词集

Table 3 Subject words set of the second innovation community for hybrid rice

主题分类	主题词
杂交水稻育种	Rice、Seed (S) Hybrid、Crossing、Backcross Plant (S) Produce (D)、Producing
育种方法	Morphological、Physiological Modified Transgene、Transforming Encoding
转基因、组织细胞培养等	Tissue、Cells
近交系、不育系、授粉、后代等	Line、Inbred、Pollen、Sterility、Progeny
作物抗性	Resistance、Resistant
抗疾病、抗害虫等	Disease、Insect、Herbicide
作物品质	Metabolism、Atty Acid、Carbohydrate
脂肪酸、碳水化合物、新陈代谢等	

孤立创新的主题分析与匹配, 首先对每个孤立专利 (主要是题名与文摘) 进行文本语义分析和词频分析, 提取出每件专利的主题词集, 这些主题词汇也是由代表每件专利主题的高频关键词组成。之后, 把每件孤立专利的主题词集与每个创新社区的主题词集进行语义分析与匹配, 根据每件专利主题与每个创新社区主题的相似度确定每件孤立专利的所属创新社区。如果 {主题词集} 孤立 x 与 {主题词集} 第 n 社区的匹配度达到 60% 以上, 则孤立点 x 归入第 n 创新社区, 否则, 不归入任何创新社区。主题分析与匹配的结果是, 235 件孤立专利归属到第一创新社区, 49 件孤立专利

归属到第二创新社区, 还有 28 件孤立专利没有归属到任何创新社区。下面将是根据杂交水稻创新社区的时间序列发展来预测创新的发展趋势。

### 3.1.3 技术创新发展趋势预测

首先, 根据杂交水稻创新社区, 对创新社区的专利数据进行最优模型拟合, 然后按最优模型外推, 进而预测技术创新社区的未来发展趋势。由于第三个创新社区样本量比较小, 而且在 2005 年之后, 数据出现中断, 因此不能产生预测结果。所以本文只对杂交水稻第一和第二个创新社区发展趋势进行分析与预测。

采用的数据是每个杂交水稻技术创新社区的年代

累积专利数，确定拟合的最优模型是 S- 曲线模型，因此采用该模型进行预测。S 曲线外推预测首先要拟合创新社区的成长模型，确定社区成长曲线的饱和点 (Saturation)、反曲时间点 (Midpoint) 和成长时间 (Growthtime) 3 项参数。之后，把累积专利数据输入，按照确定的预测模型参数，推测出创新社区的发展趋势。预测模型的 3 项参数的意义如下。

(1) 饱和点 (Saturation): 指在杂交水稻技术创新主题社区中产生的最大专利数，也是技术发展到成熟阶段专利累积数量的极值点。

(2) 反曲时间点 (Midpoint): 是 S- 曲线二次微分值为 0 的点，曲线形状在反曲时间点由上凸转变为下凸，专利累积增长速率由正数转为负数，由增加转为减少。

(3) 成长时间 (Growthtime): 指专利累积成长从饱和值的 10% 数量开始，至饱和值的 90% 数量为止所需的时间。

首先，对每个创新社区的专利年代累积数据进行曲线拟合。杂交水稻第一技术创新社区最优的拟合曲线符合 S- 曲线分布。杂交水稻第一创新社区的累积成长曲线 (图中方块连线) 与最优拟合曲线 (图中红线) 如图 5 所示。

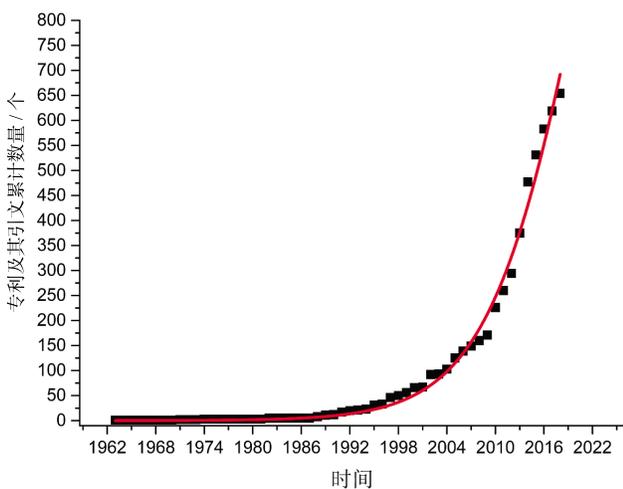


图 5 杂交水稻第一创新社区成长曲线

Fig. 5 Growth curve of the first innovation community for hybrid rice

第一创新社区成长的最优拟合曲线模型为 Slogistic

分布，模型具体形式如方程 (2):

$$Y_G = \frac{1863}{1 + e^{-0.168(t - 2021.2)}} \quad (2)$$

由拟合模型得出累积专利数的饱和值为 1 863，根据这个计算出来的最优饱和值，输入第一创新社区年代累积专利数据，调节 S- 曲线至收敛到最大置信区间，最后得出预测曲线 (图 6)。

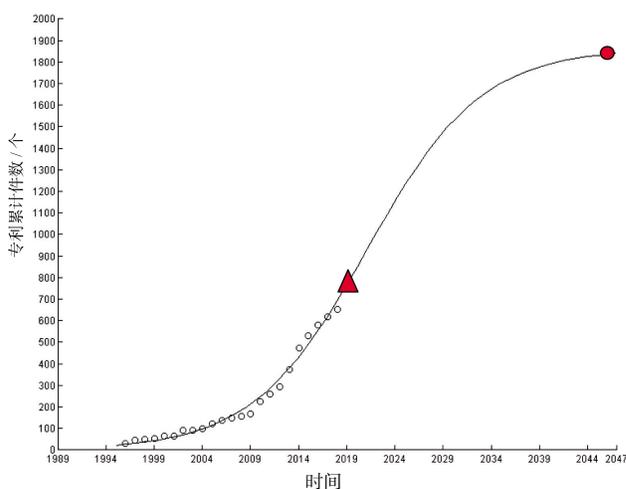


图 6 杂交水稻育种技术第一创新社区发展趋势

Fig. 6 Developing trend of the first innovation community for hybrid rice

预测的 S- 曲线如图 6 所示，由预测曲线可知，杂交水稻第一主题创新社区成长的反曲时间点 (图中三角形状点) 大约为 2021.2 年，估算的 10%~90% 的成长时间 1973—2043 年共 70 年，而接近 90% 成长饱和点即接近技术成熟 (图中平行四边形点) 的时间约为 2047—2050 年左右，饱和点专利累积件数就是前面拟合模型给出的饱和点数值，为 1 863 件。

之后对杂交水稻第二技术创新社区发展进行预测。首先依据杂交水稻第二创新社区的累积增长数据拟合出最优曲线，其数据线与拟合曲线如图 7 所示。

图中黑色方块形状连线为数据线，红色实线为拟合曲线。杂交水稻第二技术创新社区很好地符合 S 曲线分布，具体拟合曲线模型如方程 (3):

$$Y_G = \frac{294}{1 + e^{-0.369(t - 2008.19)}} \quad (3)$$

从模型中得出饱和点为 294。根据饱和点数值

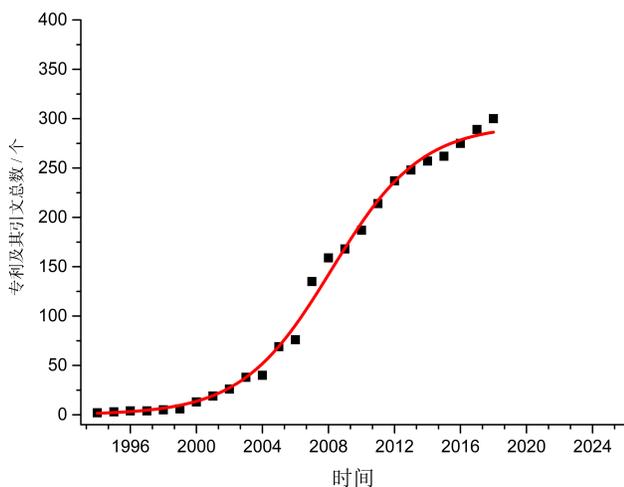


图 7 杂交水稻第二创新社区成长曲线

Fig. 7 Growth curve of the second innovation community for hybrid rice

294, 调节 S- 曲线模型曲线, 使置信区间最大, 非置信区间最小 (原方差最小), 从而得到最优化的预测拟合曲线, 结果如图 8 所示。

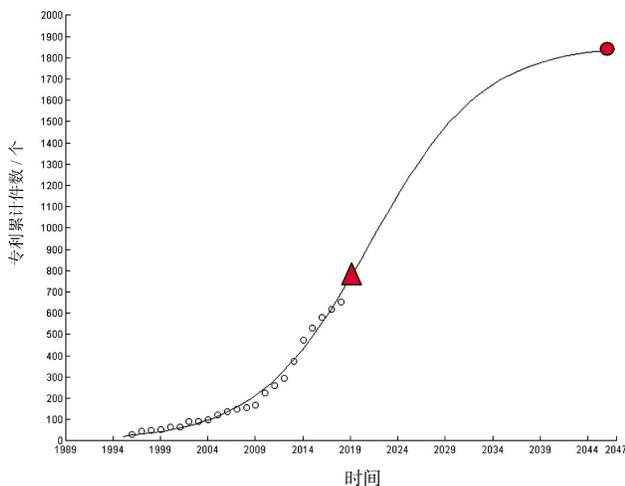


图 8 杂交水稻第二创新社区发展趋势

Fig. 8 Development trend of the second innovation community for hybrid rice

由图 8 的最优 S- 曲线可知, 杂交水稻第二主题创新社区成长的反曲时间点 (图中红色圆点) 约为 2008.2 年, 估算的 10%~90% 的成长时间 (1996—2016 年) 为 20 年, 而接近 90% 成长饱和点即接近技术成熟 (图中红色方块点) 的时间约为 2022—2024 年左右, 饱和点专利累积件数就是前面拟合模型给出的饱和点

数值, 为 294 件。

## 4 讨论与结论

### 4.1 对技术创新研究的启示

技术创新是比较难度量与刻画的。由于专利及其引文能很好地度量与刻画技术创新, 自从电子专利数据库的大量出现之后, 基于专利及其引文研究技术创新成为一个主要的技术创新研究流派。又由于专利制度在世界各国的普及发展, 专利化的技术创新就已经占据技术创新主要的地位。因此, 用专利及其引文刻画与度量技术创新是合适的。

本文提出了从专利引文技术创新网络社区中观视角, 预测技术发展趋势的方法。技术创新是一个系统, 只有包含专利引文和网络中孤立创新点的专利网路, 才是更完整的, 预测的结果才能贴近实际。本文从理论上完善了用专利文献数据预测技术创新的方法, 提出了从复杂系统角度预测中观技术发展趋势的专利文献数据预测方法, 使得其他学者或者任何人都可以在此理论方法的指导下, 对其他的技术创新领域进行发展趋势预测。

本文发展了一个社区侦测方法。主要是融合了网络连接、模块社区划分等传统社区侦测理论与文本主题挖掘理论, 建立了新的社区侦测方法。社区划分最合理的方法是从节点与网络模块的语义主题给予划分。但是, 这种方法同时带来了计算上的复杂, 在计算多数大型复杂网络的时候会使计算成本高到无法承担的地步。所以, 在低成本语义挖掘方法出现之前, 这种方法无法利用, 更无法普及。另一方面, 传统通过网络连接性和模块性划分社区的方法却遗漏了网络中的孤立点, 而这些孤立点有可能是下次链接发生的起点, 也可能由于数据结构规范引发的。因此, 本文考虑了社区侦测方法的两难限制, 融合了传统社区划分与主题语义分析方法, 在尽量不增加计算成本的前提下, 提高社区侦测的准确性, 并包含对孤立点的划分。这种社区侦测方法是融合两种思想的初步尝试, 为进一

步的研究探索指明了方向。

本文超越了整体宏观预测技术发展趋势的限制,从中观的技术创新社区角度,建立了预测技术创新中观社区发展趋势的方法。充分地利用复杂网络的拓扑特征——主要是社区拓扑结构特性进行技术发展预测,使得预测的视角更加细微、更加具体,但又不失系统性。对于今后利用专利法预测技术中观发展趋势,提供了一定的借鉴。

## 4.2 对实证技术领域的建议

目前,高抗性高产量的杂交水稻育种技术处于快速发展的转折阶段,在现有技术条件下,下一阶段的发展速度将逐渐减低,大约将在2047—2050年左右发展成熟;而高品质高抗性的杂交水稻育种技术目前处于成长阶段的后期,接近技术成熟的阶段,大约将在2022—2024年间技术发展成熟。

从本文预测的结果看,现有技术条件下,杂交水稻第一创新社区处于快速发展的中期,也是发展的转折阶段。在此阶段,该领域杂交水稻育种技术的研究者,应该从两个方面加强研究行为。第一,对现有育种技术的集成和改进完善。因为该技术创新社区处于技术发展中期阶段,因此还会有近20年左右的快速发展时期。在此阶段,该领域的研究者应该不断地加强研究,对现有技术不断地完善,不断地综合集成现有各种技术,使高产量高抗性的杂交水稻育种技术发展完善,尽可能抢占技术竞争的制高点。第二,瞄准未来杂交水稻育种高地,研发与储备下一代杂交水稻育种技术。由于技术周期规律的作用,如果没有突破性技术的产生,只对现有技术修补与集成的话,那么20年之后技术发展成熟的时候,也会遇到技术发展瓶颈。而按照技术生命周期理论,突破性技术的更新换代能改变技术生命周期曲线和发展周期,从而使技术创新进入更高层次的技术成长周期曲线上,使得技术在更高层次上发展。由于育种技术的周期比较长,以及未来技术突破过程中可能遇到各种技术难题,因此,应该尽早开始新一代高产量高抗性杂交水稻育种技术的研发。

从本文的预测结果看,杂交水稻第二创新社区已经处于技术发展的后期,接近技术发展顶峰。在该阶段,该领域的研究者没有太多选择,除了用很少的一部分精力完善现有技术,保持该技术创新领域的领先外,应该转变研究方向争取取得突破性技术创新。只有这样,才可能在未来的杂交水稻育种技术的竞争中占有一席之地。

### 参考文献:

- [1] 赵玉堂. 知识主题动态发展模式之建立——以供应链管理为例[D]. 台湾: 国立云林大学全球运筹管理研究所, 2007.
- [2] ZHAO Y T. The establishment of dynamic development model of knowledge subject - Taking supply chain management as an example[D]. Taiwan: Institute of global operations research management, National Yunlin university, 2007.
- [3] 陈峰. 日本第八次科学预见项目的竞争情报学解析[J]. 竞争情报, 2007, 春季刊: 2-5.
- [4] CHEN F. Competitive intelligence analysis of Japan's eighth science foresight project[J]. Competitive intelligence, 2007, Spring quarterly: 2-5.
- [5] Martino J P. Technological Forecasting for Decision Making[M]. New York: McGraw-Hill, Inc., 1993.
- [6] PORTER A L, ROPER A T, MASON T W, et al. Forecasting and management of technology[M]. New York: Wiley, 1991.
- [7] 李国秋, 龙怡. 预测市场应用于技术预见的优势分析——对13中常用技术预见方法的20个维度的实证研究[J]. 图书馆杂志, 2014(8): 11-28.
- [8] LI G Q, LONG Y. The Strengths of the prediction markets being applied to the technology foresight: 20 dimensions empirical studies for 13 commonly used methods of technology foresight[J]. Library journal, 2014(8): 11-28.
- [9] Technology futures analysis methods working group. Technology futures analysis: Toward integration of the field and new methods[J]. Technological forecasting & social change, 2004, 71(3): 287-303.
- [10] CHANG S B, LAI K K, CHANG S M. Exploring technology diffusion and classification of business methods: Using the patent citation network[J]. Technological forecasting & social change, 2009, 76

- (1): 107–117.
- [8] SEIDEL. Citation system for patent office[J]. Journal of patent office society, 1949, 31: 554.
- [9] HART. Re: Citation system for patent office[J]. Journal of patent office society, 1949, 31: 714.
- [10] DAIM T U, RUEDA G, MARTIN H, et al. Forecasting emerging technologies: Use of bibliometrics and patent analysis[J]. Technological forecasting & social change, 2006, 73: 981–1012.
- [11] LAI K K, WU S J. Using the patent co-citation approach to establish a new patent classification system[J]. Information processing and management, 2005, 41: 313–330.
- [12] YOUTIE J, IACOPETTA M, GRAHA S. Assessing the nature of nanotechnology: Can we uncover an emerging general purpose technology?[J]. J techno transfer, 2008, 33: 315–329.
- [13] 翟东升, 王明吉. 基于专利地图理论的专利分析系统研究[J]. 情报杂志, 2006, 3: 5–7.
- ZHAI D S, WANG M J. Research on patent analysis system based on patent map theory[J]. Journal of intelligence, 2006, 3: 5–7.
- [14] 武建龙, 陶微微, 王宏起. 基于专利地图的企业研发定位方法及实证研究[J]. 科学学研究, 2009, 27(2): 220–225.
- WU J L, TAO W W, WANG H Q. Research on R&D position methods of enterprises and related empirical study based on patent map[J]. Studies in science of science, 2009, 27(2): 220–225.
- [15] 黄鲁成, 蔡爽. 基于专利的技术轨道实证研究[J]. 科学学研究, 2009, 27(3): 363–367.
- HUANG L C, CAI S. Empirical study on the technological trajectory based on patent information[J]. Studies in science of science, 2009, 27(3): 363–367.
- [16] 李婧, 张旭, 赵蕴华, 等. 基于专利引文的技术竞争力研究[J]. 数字图书馆论坛, 2008, 11: 22–26.
- LI J, ZHANG X, ZHAO Y H, et al. Research on technology competitiveness based on patent citation[J]. Digital library forum, 2008, 11: 22–26.
- [17] 康宇航, 苏敬勤. 基于专利引文的技术跟踪系统—理论模型与工具开发[J]. 科学学与科学技术管理, 2008, 4: 24–27.
- KANG Y H, SU J Q. Technology tracking system based on patent quotation – Theory model and tool development[J]. Science of science and management of S.& T., 2008, 4: 24–27.
- [18] WARTBURG I, TEICHERT T, ROST K. Inventive progress measured by multi-stage patent citation analysis[J]. Research policy, 2005, 34: 1591–1607.
- [19] HUANG Z. International nanotechnology development in 2003: Country, institution, and technology field analysis based on USPTO patent database[J]. Journal of nanoparticle research, 2004, 6: 325–354.
- [20] MINA A, RAMLOGAN R, TAMPUBOLON G, et al. Mapping evolutionary trajectories: Applications to the growth and transformation of medical knowledge[J]. Research policy, 2007, 36(5): 789–806.
- [21] 王宏起, 夏凡, 王珊珊. 新兴产业技术融合方向预测: 方法及实证[J]. 科学学研究, 2020, 38(6): 1009–1017, 1075.
- WAGN H Q, XIA F, WANG S S. Forecast on technological fusion direction of emerging industries: Method and empirical research[J]. Studies in science of science, 2020, 38(6): 1009–1017, 1075.
- [22] 史璇. 基于科技文献的燃料电池汽车企业技术机会研究[D]. 太原: 山西大学, 2020.
- SHI X. Research on technology opportunities of fuel cell vehicle films based on scientific and technological text[D]. Taiyuan: Shanxi university, 2020.
- [23] 覃兴. 专利分析技术预测方法及其在电缆附件行业中的应用[D]. 广州: 华南理工大学, 2017.
- QIN X. Patent analysis technology prediction method and its application in cable accessories industry[D]. Guangzhou: South China university of technology, 2017.
- [24] 吴颖文, 纪杨建, 顾新建. 基于专利技术共现动态网络的产业共性技术预测研究[J]. 计算机集成制造系统, 2019, 10: 1–18.
- WU Y W, JI Y J, GU X J. Research on industrial generic technology prediction based on technology co-occurrence dynamic network of patents[J]. Computer integrated manufacturing systems, 2019, 10: 1–18.
- [25] CIGDEM B, SUAT A, YAVUZ E. Forecasting of innovation in the light of semantic networks[J]. Procedia computer science, 2019, 158 (158): 443–449.
- [26] VERSPAGEN B. Mapping technological trajectories as patent cita-

- tion networks: A study on the history of fuel cell research[J]. Working papers, 2005, 10(1): 93–115.
- [27] FONTANA R, NUVOLARI A, VERSPAGEN B. Mapping technological trajectories as patent citation networks: An application to data communication standards[R]. Working paper, 2008, 2.
- [28] 郭俊芳. 基于语义挖掘的技术创新路径分析与评价方法研究[D]. 北京: 北京理工大学, 2016.
- GUO J F. A semantic mining-based method analyze and evaluate technology innovation path[D]. Beijing: Beijing institute of technology, 2016.
- [29] 汪雪锋, 张硕, 韩晓彤, 等. 技术预测研究现状与未来展望[J]. 农业图书情报, 2019, 31(6): 4–11.
- WANG X F, ZHANG S, HAN X T, et al. Technology forecasting: Status and trends[J]. Agricultural library and information, 2019, 31(6): 4–11.
- [30] 王京安, 牛建, 王坤, 等. 基于专利引文网络的技术范式转换预见方法体系构建——以液晶材料技术的发展为例[J]. 现代情报, 2017, 37(8): 135–141.
- WANG J A, NIU J, WANG K, et al. Construction of the prediction method system of the technological paradigm transformation based on patent citation network – Taking the example of liquid crystal materials[J]. Journal of modern information, 2017, 37(8): 135–141.
- [31] CHANGYONG L, OHJIN K, MYEONGJUNG K, et al. Early identification of emerging technologies: A machine learning approach using multiple patent indicators[J]. Technological forecasting & social change, 2018, 127(127): 291–303.
- [32] IAN M P L, ALEX F D P, GECIANE S P. Patent-based network analysis to understand technological innovation pathways and trends[J]. Technology in society, 2019, 59(59).
- [33] AHMAD B, BRUNO A, CATHERINE B. Discovering and assessing fields of expertise in nanomedicine: a patent co-citation network perspective[J]. Scientometrics, 2013, 94(3): 1111–1136.
- [34] 范维熙. 基于专利地图的太阳能电池技术发展趋势研究[D]. 南京: 南京工业大学, 2015.
- FAN W X. Research on technological trend of solar cell technology based on patent map[D]. Nanjing: Nanjing tech university, 2015.
- [35] 金丹. 基于专利信息分析的激光加工技术发展趋势预测 [D]. 武汉: 华中科技大学, 2015.
- JIN D. A thesis submitted in partial fulfillment of the requirements for the degree for the master of engineering[D]. Wuhan: Huazhong university of science and technology, 2015.
- [36] 许冠南, 谢梦娇, 潘美娟, 等. 3D打印产业技术的演变与预测研究——基于专利主路径分析[J]. 北京邮电大学学报(社会科学版), 2016, 18(4): 77–85.
- XU G N, XIE M J, PAN M J, etc. Technology forecasting of 3D printing industry – Based on main path analysis of patent[J]. Journal of Beijing university of posts and telecommunications (social sciences edition), 2016, 18(4): 77–85.
- [37] 杨中楷, 刘倩楠, 于霜. 基于专利引文网络的技术领域间关系探析[J]. 科研管理, 2011, 32(5): 149–155.
- YANG Z K, LIU J N, YU S. The relationship among technical fields based on patent citation network[J]. Science research management, 2011, 32(5): 149–155.
- [38] 杨中楷, 刘佳. 太阳能光伏技术演进路径的可视化分析[C]. 第六届中国科技政策与管理学术年会论文集, 2010.
- YANG Z K, LIU J. Visual analysis of evolution path of solar photovoltaic technology [C]. Proceedings of the 6th annual meeting of China science and technology policy and management, 2010.
- [39] 柴庆凤, 翟东升, 蔡力伟, 等. 基于专利网络链接模型的技术链接机会预测方法研究[J]. 情报理论与实践, 2020, 7: 1–18.
- CHAI Q F, ZHAI D S, CAI L W, et al. Study on the method of the technology link opportunity forecasting based on patent network link model[J]. Information studies: theory & application, 2020, 7: 1–18.
- [40] 蔡力伟. 基于链接预测的技术机会预测系统研究[D]. 北京: 北京工业大学, 2016.
- CAI L W. A research on technology opportunity prediction system based on link prediction[D]. Beijing: Beijing university of technology, 2016.
- [41] 翟东升, 刘鹤, 张杰, 等. 一种基于链路预测的技术机会挖掘方法[J]. 情报学报, 2016, 35(10): 1090–1100.
- ZHAI D S, LIU H, ZHAN J, et al. Approach to mining technology opportunity based on link prediction[J]. Journal of the China society for scientific and technical information, 2016, 35(10): 1090–1100.

- [42] CHUN J H, LEE C S, LEE S J. Central technology deriving for the patents of medical device using social network analysis[J]. *Management & information systems review*, 2016, 35(2): 221–254.
- [43] 李佳佳, 马铁驹. 利用局部离群因子算法探测核心技术发展趋势——以中国风能专利数据为例[J]. *情报杂志*, 2017, 36(3): 119–124, 195.
- LI J J, MA T J. Detecting core technology's development trend with LOF based on wind energy patent data of China[J]. *Journal of intelligence*, 2017, 36(3): 119–124, 195.
- [44] HANLIN Y, MENGJUN L, KEITH W H, et al. Development trend forecasting for coherent light generator technology based on patent citation network analysis[J]. *Scientometrics*, 2017, 111(1): 297–315.
- [45] 周源, 刘怀兰, 廖岭, 等. 基于主题模型的技术预见定量方法综述[J]. *科技管理研究*, 2017, 37(11): 185–196.
- ZHOU Y, LIU H L, LIAO L, et al. Literature review of quantitative technology foresight methods based on topic modeling[J]. *Science and technology management research*, 2017, 37(11): 185–196.
- [46] 廖岭. 基于文本挖掘和主路径分析的技术趋势预测方法及案例研究[D]. 武汉: 华中科技大学, 2017.
- LIAO L. Method and case study based on text mining and main path analysis technological tendency forecasting[D]. Wuhan: Huazhong university of science and technology, 2017.
- [47] 高建. 基于冲突网络和逻辑曲线模型的技术预测研究[D]. 天津: 河北工业大学, 2018.
- GAO J. Technology forecasting based on the network of contradiction and logistic curve models[D]. Tianjin: Hebei university of technology, 2018.
- [48] 文雄辉. 基于专利挖掘的技术预见方法及其应用[D]. 广州: 华南理工大学, 2019.
- WEN X H. The method technology foresight based on patent mining and its application[D]. Guangzhou: South China university of technology, 2019.
- [49] 钱娇. 基于 SAO 的专利挖掘方法及在工业机器人领域的应用[D]. 广州: 华南理工大学, 2018.
- QIAN J. The method of patent mining based on SAO and its application in industrial robot[D]. Guangzhou: South China university of technology, 2018.
- [50] 张振刚, 罗泰晔. 基于知识网络的技术预见研究[J]. *科学学研究*, 2019, 37(6): 961–967, 985.
- ZHANG Z G, LUO T Y. Research on technology foresight based on knowledge network[J]. *Studies in science of science*, 2019, 37(6): 961–967, 985.
- [51] XIONGHUI W. Technology foresight research of industrial robot based on patent analysis[J]. *Journal of data analysis and information processing*, 2019, 7(2): 74–90.
- [52] 尹忠博, 罗威, 罗准辰. 数据驱动的技术预测: 现状、技术与趋势[J]. *情报理论与实践*, 2018, 41(12): 35–40.
- YIN Z B, LUO W, LUO Z C. Data-driven technology forecasting: status, technology and trends [J]. *Information studies: Theory & application*, 2018, 41(12): 35–40.
- [53] DIANE J C, LAWRENCE B H. *Mining graph data*[J]. New Jersey: John Wiley & Sons, Inc., Hoboken, 2007.
- [54] NEWMAN M E J, GIRVAN M. Finding and evaluating community structure in networks[J]. *Phys. Rev. E*, 2004, 69(2): 026113.
- [55] KERNIGHAN B, LIN S. An efficient heuristic procedure for partitioning graphs[J]. *The bell system technical J*, 1970, 49.
- [56] SHI J, MALIK J. Normalized cuts and image segmentation[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2000, 22(8): 888–905.
- [57] LUXBURG U V. A tutorial on spectral clustering[J]. *Statistics and computing*, 2007, 17(4): 395–416.
- [58] NEWMAN M E J. Fast algorithm for detecting community structure in networks[J]. *Physical review E*, 2004, 69(6): 066133.
- [59] SATULURI V, PARTHASARATHY S. Scalable graph clustering using stochastic flows: Applications to community discovery [C]. In *KDD' 09*, New York, ACM USA, 2009: 737–746.
- [60] DONGEN S V. *Graph clustering by flow simulation*[D]. Utrecht: University of Utrecht, 2000.
- [61] GUO J-L. S-Curve networks and an approximate method for estimating degree distributions of complex networks[J]. *Chin. Phys. B*, 2010(120503): 1–7.