University of Zagreb

Faculty of Mechanical Engineering and Naval Architecture

Vladimir Smojver

A MODEL OF INNOVATION EVOLUTION IN THE DEVELOPMENT OF TECHNICAL SYSTEMS

DOCTORAL DISSERTATION

Zagreb, 2022

University of Zagreb

Faculty of Mechanical Engineering and Naval Architecture

Vladimir Smojver

A MODEL OF INNOVATION EVOLUTION IN THE DEVELOPMENT OF TECHNICAL SYSTEMS

DOCTORAL DISSERTATION

Supervisor:
Prof. Mario Štorga, PhD

Zagreb, 2022

Sveučilište u Zagrebu

Fakultet strojarstva i brodogradnje

Vladimir Smojver

MODEL EVOLUCIJE INOVACIJA U RAZVOJU TEHNIČKIH SUSTAVA

DOKTORSKI RAD

Mentor:
Prof. dr. sc. Mario Štorga

Zagreb, 2022.

BIBLIOGRAPHY DATA

UDC	62:347.77
Keywords	Technology management; Technology life cycle analysis; Link prediction; Patent analysis; Patent citation analysis; Patent co-citation analysis
Scientific area	Technical Sciences
Scientific field	Mechanical Engineering
Institution	University of Zagreb, Faculty of Mechanical Engineering and Naval Architecture; CVH d.d.
Supervisor	Prof. Mario Štorga
Number of pages	166
Number of figures	30
Number of tables	37
Number of references	196
Date of oral examination	01.06.2022
Committee members	Prof. Neven Pavković *(University of Zagreb, Croatia)* Prof. Marina Dabić *(University of Zagreb, Croatia)* Prof. Mo Monsouri *(Stevens Institute of Technology, USA)*
Archive	University of Zagreb, Faculty of Mechanical Engineering and Naval Architecture

ACKNOWLEDGEMENTS

I would like to thank my supervisor Prof. Mario Štorga for his contribution as well as a never-ending amount of support, encouragement and constructive criticism. Moreover, I would like to thank the examiners, Prof. Neven Pavković, Prof. Marina Dabić and Prof. Mo Mansouri for their helpful comments in the process of reviewing this thesis.

I would like to thank my loving parents for their continuous help, support and encouragement. Finally, I would like to thank my friends for their patience and support and a special thank you to Morana.

TABLE OF CONTENTS

ABSTRACT

Developing new technologies is one of the most important goals of contemporary scientific and industrial research. Understanding how a technology domain evolves and its current state is invaluable in an ecosystem seeing the speed of technology evolution increasing at rapid pace. An overview of existing literature showed that while there is a significant volume of research focusing on using patents to study technology change, most of this research, in a technology evolution context, focuses on studying technology trajectories and convergence, with limited research combining insights from research based on other resources (i.e. paper citations) and applied to patent networks. Moreover, the review of literature shows that the majority of patent analysis methods focus on exploring technology development trajectories by examining the direct citations of patents. While this approach provides insight into the generational flow of knowledge, it provides little insight into how existing patents might combine and co-contribute to a future patent in the form of co-citations. Finally, a review of literature showed that the vast majority of patent-based methods for life cycle analysis as well as prediction base themselves on models derivate of the basic S – Curve model, providing little understanding of the underlying dynamics of patent attributes and their correlation to the life cycle phases. Quantitative methods not based on the S-Curve model mostly do not use patent data as a primary data source and give limited insight into the future knowledge flows.

This thesis aims to present a novel way of exploring the life cycle stages of a technology domain by conducting a dynamic growth analysis of a patent citation network, with patents being used as proxies for technological invention and patent citations representing the flow of knowledge. Additionally, new insights into the dynamics of the flow of knowledge within a technology domain are made by applying several link prediction algorithms to patent co-citation networks with the goal of identifying the link prediction algorithms most successful in describing the underlying intuition of co-citation network growth. Moreover, the dynamics of co-citation creation are explored by determining which part of a technology domains life cycle influences the link prediction algorithms precision the most and when the predicted links occur.

Two technology domains are explored; the car headlights technology domain representing a mature technology and the neuromorphic hardware technology domain representing an emerging technology. The choice of two technologies different in nature is deliberate; this way, the evolution of two different types of technologies can be explored and compared, helping to identify the particularities of each technologies evolution. The presented methodology for

exploring the evolution of a technology domain consists of creating a dataset containing patents representing the studied technology domain and conducting a pair of technology life cycle (TLC) analyses. The first life cycle analysis is performed using an established method based on the cumulative number of patent applications over time, while the second is performed using a novel method based on analysing the dynamic growth of a patent citation network. An algorithm is introduced to convert the patent citation network into a patent co-citation network. Several link prediction algorithms are applied to the created co-citation networks to explore the underlying intuition governing co-citation network growth.

The study results show that a correlation exists between the stages of a technology domains life cycle and changes in the dynamics of patent citation network growth. The transition of the mature technology domains life cycle stage from growth to maturation correlates with a noticeable change in patent citation growth dynamics. Additionally, examining the emerging technology domain, it is found that a correlation exists between the time when an exponential increase in the number of inventions starts and a change in the dynamic of patent citation network growth. The Preferential Attachment link prediction algorithm is shown to be the most successful in predicting missing links in a mature technology. The results indicate that the patent co-citation occurring at the end of the growth TLC stage and the start of the maturation TLC stage contribute the most to the algorithm's precision. Moreover, it is demonstrated that most of the predicted missing links occur in a time frame closely following the application of the link-prediction algorithm. The results of studying the emerging technology domain show that link prediction algorithms have a significantly lower success in predicting missing links. The Adamic/Adar, Resource Allocation Index and Jaccard Coefficient show a moderate to low success in predicting missing links while the Preferential Attachment shows no precision.

This thesis provides a contribution in both a theoretical and practical context. In a theoretical context, the theoretical background of previous studies is expanded with new insights in patent citation networks growth as well as the dynamics of patent co-citation network growth. In a practical context, it is demonstrated that companies involved in planning for the short term should consider the knowledge contained in the patents relevant to their respective fields, reinforcing the notion that proper knowledge management is an invaluable tool to companies aspiring to innovate or produce innovative products.

KEYWORDS :

Technology management; Technology life cycle analysis; Link prediction; Patent analysis; Patent citation analysis; Patent co-citation analysis

PROŠIRENI SAŽETAK

Razvoj novih tehnologija je jedan od glavnih ciljeva današnje znanosti i industrijskog razvoja. Kako bi stekle dominantan i povoljan položaj na tržištu, tvrtke koje posluju u kompetitivnom globalnom okruženju pokušavaju unaprijediti svoje razvojne procese, razviti nove proizvode ili ponuditi nove usluge na tržištu temeljeno na unaprijeđenim ili novim tehnologijama. Sukladno tome, suvremena istraživanja trendova nadolazećih tehnologija i njihovog razvoja temeljenog na tehničkim inovacijama predstavljaju sve značajniji dio istraživačkih i praktičnih napora u akademiji i industriji. Određivanje smjera razvoja tehnologije se može koristiti u industriji za potporu strateškom i dugoročnom planiranju razvoja proizvoda, procesa i usluga. Svrha određivanje smjera razvoja tehnologije je strukturirano razumijevanje i opisivanje odnosa između tehničkih inovacija, njihove implementacije u fizičke sustave i usluge, te razvoja tržišta kroz vrijeme. Iako je većina postojećih pristupa koji se za to koriste u praksi kvalitativna, istraživači pokušavaju razviti kvantitativne metode za podršku određivanju smjera razvoja budućih tehnologija. Trenutno ne postoji metoda koja omogućuje uspješno kombiniranje kvalitativnih i kvantitativnih pristupa na prikladan način. Organizacije koje posluju u visoko kompetitivnim okruženjima imaju potrebu za pravovremenim saznanjima o nadolazećim tehnologijama kako bi mogle pravovremeno planirati unapređenje proizvodnih i poslovnih procesa, te uvođenje novih proizvoda ili usluga na tržište. Istraživanja o uvjetima i načinim nastanka novih tehnologija, te proučavanje dinamike njihovog razvoja, značajna su u teoretskom kao i u praktičnom smislu. Svrha predviđanja smjerova razvoja tehnologije je minimizirati ili ukloniti iznenađenja saznanjima o svim mogućim ishodima tehničkog razvoja. Spoznajom da su se tradicionalni modeli predviđanja razvoja tehnologije, primjerice Moorov zakon ili Kryderov zakon, pokazali nepreciznim, javlja se potreba za novim modelima koji bi omogućili unapređenje uvida u smjerove razvoja tehnologije. Kod proučavanja razvoja tehnologije ključan je pojam „evolucija“ tehnologije koji podrazumijeva unapređenje performansi tehnologije kroz vrijeme. Literatura opisuje dva modela evolucije tehnologije: kontinuirani i diskontinuirani. Istraživači koji zagovaraju model kontinuirane i inkrementalne evolucije tehnologije tvrde da se taj proces odvija putem rekombinacije i sinteze osnovnih elemenata postojeće tehnologije, te tvrde kako je unapređenje performansi tehnologije u tim aktivnostima rezultat promjene u shvaćanju, vrijednostima, kulturi, organizacijskoj strukturi, resursima i ključnim kompetencijama ljudi koji rade u razvoju kao i društva u cjelini. Za te istraživače je inovacija društveni proces koji se zasniva na akumulaciji malih unapređenja, a ne na značajnom doprinosu genijalnih pojedinaca. Istraživači koji zagovaraju model diskontinuirane evolucije tehnologije, tvrde da se tehnologija unapređuje kroz razdoblja inkrementalnog unapređenja koja su isprekidana s diskontinuiranim pomacima. Oni tvrde kako proizvodi i usluge koji se temelje na potpuno novim tehničkim inovacijama kreiraju značajan napredak te postaju dominantna tehnologija što za posljedicu ima diskontinuirani pomak.

Metodologija

Metodologija istraživanja temelji se na općoj metodologiji istraživanja u znanosti o konstruiranju te se sastoji od 4 temeljna koraka: preliminarno istraživanje (razjašnjenje problema, definiranje ciljeva istraživanja i hipoteza), deskriptivno istraživanje I (definiranje teoretske podloge i pregled modela), preskriptivno istraživanje (izrada teoretskog okvira, kreiranje empirijske studije) i deskriptivno istraživanje II (empirijske studije te diskusija rezultata). Preliminarno istraživanje obuhvaća pregled postojeće stručne i znanstvene literature u području istraživanja. Na temelju pregleda područja, uspostavljen je inicijalni opis postojeće situacije, kao i opis željenih rezultata, s ciljem definiranja osnovnih pretpostavki istraživanja. Uvidom u postojeću literaturu napravljen je pregled teoretskih osnova korištenja patenata kao posrednika za tehnološke izume, teoretskih pristupa opisivanja evolucije tehnologije te teoretskih pristupa predviđanja razvoja tehnologije. Nadalje, uvidom u postojeću literaturu napravljen je pregled modela za opisivanje i predikciju razvoja tehnologije pri čemu je naglasak stavljen na modele koji kao svoj temelj imaju patentne prijave. Ishod ovog koraka je definiranje rupa u području te formuliranje istraživačkih pitanja, čime je usmjeren daljnji tijek istraživanja. Na temelju rezultata pregleda literature definirana je teoretska podloga istraživanja, koja sintetizira ranije prezentirane teoretske pristupe u jedan zajednički okvir unutar kojega će se vršiti istraživanje. Pojašnjen je odabir pojedinih teoretskih pristupa kao i njihov odnos s istraživačkim pitanjima i hipotezom. Empirijsko istraživanje sastoji se od dvije studije, pri čemu prva istražuje tehnologiju u zreloj fazi svog životnog ciklusa, a druga istražuje tehnologiju na početku svog životnog ciklusa. Obje studije su strukturirane na isti način, te se sastoje od empirijskog i eksperimentalnog dijela. Empirijski dio obje studije fokusiran je na analizu životnog ciklusa tehnologije, koristeći postojeću metodu temeljenu na kumulativnom broju patentnih prijava te koristeći novu metodu temeljenu na dinamičkoj analizi rasta mreže. Eksperimentalni dio obje studije fokusiran je na primjenu algoritma za predviđanje stvaranje veza na mreži kocitata patenata te istraživanje dinamike stvaranja kocitata. Verifikacija modela napravljena je kroz diskusiju rezultata, pri čemu se daje odgovor na ranije definirana istraživačka pitanja te hipoteze istraživanja.

Teoretska osnova

Preliminarnim pregledom literature fokus istraživanja usmjeren je na metode za modeliranje i predviđanje evolucije tehnologije temeljene na patentima. Izložena je teoretska podloga za korištenje patenata kao posrednika za tehnološke izume, pozivajući se na postojeću praksu koja patente promatra kao pouzdane formalizirane zapise izuma. Naglasak je stavljen na metode za kreiranje skupova patenata koji precizno opisuju tehnološku domenu. Nadalje, naglasak je stavljen na razumijevanje svojstava patenata te identifikacije meta podatka unutar patenta koji će se koristiti kao temelj za ovo istraživanje.

Predstavljen je koncept toka znanja unutar tehnološke domene te njegova povezanost s evolucijskim etapama tehnološke domene. Istraživanje se temelji na teoriji koja dijeli životni ciklus tehnologije na četiri etape pri čemu se kumulativna vrijednost određenih karakteristika tehnologije može opisati S-krivuljom.

Konačno, predstavljena je mreža kocitata patenata kao temelj za predviđanje toka znanja unutar tehnološke domene. Pretpostavka je da predviđanjem nastanka novih veza unutar mreže kocitata možemo predvidjeti utjecaj starijih izuma na nove.

Spajanjem teorija iz ova tri područja stvorena je teoretska osnova istraživanja, sintetizirana kroz tri pretpostavke. Ova teoretska osnova podupire analizu i interpretaciju rezultata te pomaže u donošenju širih generalizacija.

Empirijske studije

Izvršene su dvije studije s ciljem odgovaranja na istraživačka pitanja i verifikacije hipoteze. Obje studije slijede istu strukturu te se sastoje od sljedećih koraka. Prvi korak studije sastoji se od kreiranja skupa patenata koji opisuju tehnološku domenu, pri čemu je korištena prilagođena inačica postojeće i potvrđene metodologije. Samo empirijsko istraživanje sastoji se od dvije pod-studije. Prva pod-studija je empirijske naravi te se sastoji od provođena analize životnog ciklusa tehnološke domene. Napravljene su dvije analize životnog ciklusa tehnologije gdje je prva analiza napravljena koristeći utvrđenu metodu temeljenu na kumulativnom broju patentnih prijava dok je druga analiza napravljena metodom predstavljenom u ovom istraživanju, a koja se temelji na dinamičkoj analizi rasta mreže citata patenata. Druga pod-studija je eksperimentalne naravi, te se sastoji od primjene algoritama za predviđanje nastajanja veza unutar mreže na mrežu koja predstavlja kocitate patenata unutar tehnološke domene. Nadalje, proučava se dinamika stvaranja kocitata unutar tehnološke domene s naglaskom na otkrivanje koji patenti najviše utječu na stvaranje novih kocitata te kada se predviđeni kocitati stvaraju.

Koristeći navedenu strukturu studije, analizirane su dvije tehnološke domene. Prva tehnološka domena sastoji se od patenata koji opisuju svjetla na automobilima. Ova tehnološka domena predstavlja primjer tehnologije u zadnjoj fazi svog životnog ciklusa. Druga tehnološka domena sastoji se od patenata koji opisuju neuromorfno sklopovlje. Ova tehnološka domena predstavlja tehnologiju na početku svog životnog ciklusa.

Primjenom metodologije na tehnologije u različitim fazama njihovog životnog ciklusa ispituje se može li se metodologija generalizirati na sve tehnološke domene, ili samo tehnološke domene koje dijele određene karakteristike.

Rezultati empirijski studija

Rezultati prve studije, koja proučava tehnološku domenu u zadnjoj fazi svog životnog ciklusa, pokazuju da se dinamički rast mreže citata sastoji od dvije diskretne faze. U prvoj fazi krivulja koja opisuje rast prati generalno pozitivni trend nakon čega kreće faza negativnog trenda, pri čemu pozitivan trend krivulje označava period gdje se stvara više novih patenata od citata dok negativan trend označava period gdje se stvara više citata od novih patenata. Točka prelaska iz pozitivnog u negativni trend približno odgovara početku sazrijevanja životnog ciklusa tehnologije koja je određena analizom temeljenoj na kumulativnom borju patentnih prijava. Ovim je rezultatima pokazana korelacije između dinamike faza životnog ciklusa tehnologije i dinamike rasta mreže citata.

Rezultati empirijske pod-studije pokazuju da prva polovina životnog ciklusa tehnologije, koja se sastoji od faza uvođenja i rasta, prati trend gdje se stvara više novih patenata nego citata. U kontekstu evolucije tehnologije, ovo pokazuje da ove faze životnog ciklusa tehnologije generiraju pretežito originalne i inovativne izume. Druga polovica životnog ciklusa tehnologije, koja se sastoji od faze zrelosti i opadanja, stvara manje inovativne izume koje se većinom temelje na prijašnjim izumima.

Druga pod-studija prve studije pokazala je da mreža kocitata patenata raste sljedeći intuiciju preferencijalne vezanosti, odnosno da vjerojatnost da neki patent bude kocitiran raste s brojem kocitata koje ima. Nadalje, pokazano je da novi izumi većinom citiraju mlađe patente, odnosno da većine toka znanja proizlazi iz mlađih patenata. Konačno, pokazano je da većina predviđenih kocitata nastane u bližoj budućnosti.

Rezultati druge studije, koja proučava tehnološku domenu na početku svoga životnog ciklusa, pokazuju da graf dinamičkog rasta mreže citata kreće s fazom negativnog trenda nakon čega kreće faza pozitivnog trenda. Točka prelaska iz negativnog u pozitivni trend približno odgovara trenutku kada broj prijavljenih patenata unutar tehnološke domene počinje eksponencijalno rasti, što je vidljivo iz analize životnog ciklusa na temelju kumulativnog broja prijavljenih patenata.

Druga pod-studija druge studije nije pokazala je da mreža kocitata patenata koji opisuju tehnološku domenu u prvoj fazi svog životnog ciklusa ne prati nikakvu intuiciju rasta. Od četiri razmatrana algoritma za predviđanje veza unutar mreže, niti jedan nije uspješno opisao intuiciju rasta mreža. Štoviše, algoritam preferencijalne vezanosti pokazao je najmanju preciznost,

Vrednovanje istraživanja

Vrednovanje metode provedeno je raspravom kojom se adresiraju istraživačka pitanja i ciljevi istraživanja te hipoteza. Pri raspravi se također koriste saznanja iz dostupne literature.

Rezultati obje studije vezani za prvo istraživačko pitanje ukazuju da postoji korelacija između dinamike stvaranja citata patenata i faze životnog ciklusa promatrane tehnološke domene. Ova korelacija se primarno manifestira kao promjena u omjeru broja novo stvorenih patenata i novo stvorenih citata. Ova promjena je jasno uočljiva na vizualizaciji rezultata analize dinamičkog rasta mreže citata patenata. U slučaju zrele tehnologije, vrijeme tranzicije iz faze rasta u fazu zrelosti podudara se s promjenom u rezultatu dinamičke analize rasta gdje pozitivne trend rasta prelazi u negativni trend. Kod tehnologije na početku svog životnog ciklusa, nagli porast kumulativnog broja prijavljenih patenata podudara se s promjenom u rezultatu dinamičke analize rasta gdje trend rasta prelazi iz negativnog u pozitivni.

Rezultati vezani za drugo, treće i četvrto istraživačko pitanje doprinose razumijevanju dinamike stvaranje kocitata patenata unutar tehnološke domene. Rezultati prve empirijske studije pokazuju da dinamika rasta mreže kocitata patenata, kreirane od patenata koji predstavljaju zrelu tehnologiju, prati sličnu dinamiku kao i kocitati istraživačkih radova, točnije dinamiku preferencijalne vezanosti. Ovi rezultati ukazuju da tok znanja slijedi sličnu dinamiku širenja u oba nositelja znanja.

Na temelju vrednovanja istraživanja naglašeni su sljedeći doprinosi, kako u teoretskom tako i u praktičnom kontekstu. Prvi teoretski doprinos sastoji se od uvođenja novog načina za istraživanje faze životnog ciklusa tehnološke domene. Točnije, predstavljen je novi način

određivanja faze životnog ciklusa tehnologije na temelju analize dinamičkog rasta mreže citata patenata. Drugi teoretski doprinos sastoji se od pružanja uvida u to kako se elementi znanja šire unutar tehnološke domene. Razumijevanjem obrazaca difuzije znanja, dobiva se uvid u ishodišnu intuiciju toka elemenata znanja. Posljedično, ova se spoznaja može iskoristiti za predviđanje difuzije znanja u budućnosti. Nadalje, dodatno istraživanje dinamike toka elemenata znanja mogu pružiti uvod u to koje tehnologije iz životnog ciklusa tehnološke domene najviše utječu na buduće izume, kao i vrijeme pojave tih budući izuma.

U praktičnom kontekstu, ovi su rezultati primjenjivi na mikro i marko razini. Na mikro razini, metode iz ovog istraživanja mogu biti korištene u ideacijskoj fazi razvoj proizvoda. Primjene algoritma za predviđanja veza na mrežu kocitata patenata može korisniku pružiti uvid u dotad neistražene kombinacije patenata. Ovaj pristup se nadovezuje na postojeća istraživanja koja promatraju proces invencije kao kombinatorički proces. Na makro razini, metode iz istraživanja se mogu koristit na razvoju strategije razvoja tehnološkog portfolija.

Ključne riječi:

upravljanje tehnologijom; analiza životnog ciklusa tehnologije; predviđanje veza; analiza patenata; analiza citata patenata; analiza kocitata patenata

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND SYMBOLS

Abbreviations

ARC – Area of Relevance and Contribution

AUC – Area Under Curve

CPC – Cooperative Patent Classification

DRM – Design Research Methodology

ECLA – European Classification System

EPO – European Patent Organization

GTM – Generative Topographic Mapping

HOH – Hierarchy of Hypotheses

IPC- International Patent Classification

MC – Manual Code

NLP – Natural Language Processing

R & D – Research and Development

RFID - Radio Frequency Identification

RQ – Research Question

SAO – Subject Action Object

TLC – Technology Life Cycle

USPTO – United States Patent Organization

VC – Venture Capital

VLSI - Very Large Scale Integration

WIPO – World Intellectual Property Organization

Symbols

E – set of edges

e – edge

E^T – training subset of edges

E^P – testing subset of edges

G – undirected simple graph

i - step of study

$m(i)$ – number of edges

$m(p)$ – total number of edges

ML – set of missing links

$n(i)$ – number of nodes

$n(p)$ – total number of nodes

$n(U)$ – number of links contained in set U

NL – set of non-existent links

n – number of independent comparisons

$n^{'}$ – number of instances of missing link having a higher score then a randomly picked non-existent link

$n^{''}$ – number of instances of missing link having the same score then a randomly picked non-existent link

N_s – number of predicted links remaining after applying cut-off rate

N_{rs} – number of correctly predicted links

p – total number of steps

P – precision

$Sim(u,v)$ – similarity measure

T_n - training dataset

U – set containing all possible links within a network

u,v – node pair

V_n – verification dataset

V - Set of nodes

$|V|$ - number of nodes in set V

$\Gamma(x)$ – set of neighbours of node x

$|\Gamma(x)|$ - number of neighbours of node x

δ – network's rate of growth

$\delta(i)$ – actual rate of growth

$\delta_e(i)$ – measure taking into account the total size of a network at the end of a study, averaged over all step

$\delta_q(i)$ – relative network growth indicator per step

1. INTRODUCTION

This chapter aims to outline the importance of the topic explored in this thesis and provide the context of the presented research. The reader is introduced to the gap in research the presented work will fill as well as the research questions and problems the research addresses. Finally, the contribution of the research to the broader field is presented, as is an overview of the methods employed, the study's limitations and the structure of the thesis.

There is a consensus, in both an industrial and academic context, that knowledge is one of the key drivers of economic growth [1][2][3], with innovation being an integral part of the significant knowledge transformation in modern business [4]. This makes studying and explaining the spread of knowledge an attractive topic to scholars [5]. One of the facilitators of knowledge as a driver of growth is the development of new technologies. In order to attain a more dominant market position, companies operating in a competitive global environment seek to improve their development processes, develop new products, or offer new services based on improved or completely novel technologies. Consequently, research trends focused on studying emerging technologies and their development based on technical innovations represent an increasingly significant percentage of research and practical effort both in an academic and industrial context [6][7]. Furthermore, the management of knowledge is rising in popularity in an industrial context as organisations realize the importance of knowledge in creating new technologies. The survival of large organisations often depends on their ability to navigate the technological currents in their environments [8][9]. A result of this is the existence of a significant volume of research focused on the study of knowledge flow in different contexts, such as studying knowledge flow in the global economy [10][5][11], within organizations [12][13] and technology domains [14][15]. The latter emphasises studying knowledge in its manifestation as the evolution of technology, stating that the evolution of technology is typically guided by problem-solving activities integrating the knowledge from areas of technology and leveraging the cumulative nature of knowledge [15]. In an industrial context, an understanding of the evolution of technology has been shown to be invaluable to stakeholders at different levels, tasked with making decisions related to the implementation or development of technology. These stakeholders can be both on a project [16] or a strategic [17] level. In an academic context, insight into the nature of technology evolution can provide additional value for the fields of knowledge and innovation management [18], design theory [19], and risk management [20].

The purpose of understanding the evolution of technology and predicting the potential directions of technology development is to minimize or eliminate surprises by having an insight into the possible outcomes of the development of a technology. Since there is a growing awareness that traditional models of forecasting the development of technologies, such as Moore's Law [21] or Kryder's law, have been proved inaccurate in a contemporary context [22], there is a need for new models which would provide improved insight into the potential directions of technology development [22].

A key concept when studying the development of technology is "technology evolution", the process of continuous improvement of technology over time [6]. The timeframe in which a technology evolves is referred to as its life cycle [23][24]. This technology life cycle consists of multiple stages, each having its own set of characteristics. There are two main approaches to describing the principles governing the evolution of technology that can be found in the literature. The first one states that technology evolves incrementally, viewing the evolution of technology as a series of incremental improvements that accrue over time and result in significant technological advances [22][25][26]. An incremental new product is considered a refinement, adaptation, and enhancement of existing products [19]. The second theory of technology evolution proposes that a technology evolves in a continuous cycle of stagnation and radical improvements. Radical innovation facilitates the improvement of technologies, followed by periods of relative stagnation. This cycle then repeats during the technology's life cycle [27][28][29]. An additional important concept in technology evolution, separate from the first two, covers the appearance of disruptive innovations, innovations that create a new market and value network, eventually disrupting an existing market, resulting in the displacement of established market leaders [9].

The motivation for conducting this research is based on the continuous need to better understand the way technologies evolve by exploring the rules governing the evolution of technology, thereby increasing the ability to predict the potential development of technology and consequently reducing uncertainty in technology management. Moreover, any results of this research should be based on formalized records of technology inventions, minimising the need for expert knowledge when examining a technology. The formalized records of invention used should be accessible, structured and objective, enabling the results of this research to be easily repeatable. For this reason patents are considered the primary source of data in this research [3]. While several methods exist for exploring the life cycle stage of a technology, there is a limited number of approaches that focus on using the citations between patents as the basis for analysis. This is considered a significant gap as patents are considered among the most promising formalized written records of knowledge and have a history of being used as proxies for technical invention [2][30][31]. Additional research into patent citation may therefore provide new insight into how a technology domain evolves. Moreover, increasing the understanding of the correlation between patent citations and a technologies life cycle might provide researchers with a more profound insight into the underlying patterns of patent citations and consequently technology creation.

Specifically, the literature review shows a noticeable lack of research studying patent co-citations and the dynamics of their creation. This means that exploring patent co-citation could yield results that would provide a meaningful contribution to the broader field. Identifying an underlying intuition governing the creation of patent co-citations might further increase our understanding of how technologies evolve and potentially enable us to predict the future evolution of technology. Consequently, the dynamics of patent co-citation creation is one of the critical phenomena studied in this research.

Ultimately, any additional insight into the rules governing the evolution of a technology serves to partially mitigate risk to decision-makers at both a project and strategic level. This is evident when previous technologies which failed catastrophically are observed, primarily due to a lack of foresight of development trends. Examples include the devastation of Swiss watch manufacturers by the transition from mechanical to electric and quarts movements in watches [32], which caused the decline of the share the of Swiss watch export market (by volume) from 40 % in 1974 to just 10% in 1984. Because Swiss watch makers gathered in a watch cartel, they were insulated from the effect of inter-firm competition. However, they neglected external events and the introduction of a new technology by non-Swiss competitors. This crisis cost around two-thirds of the Swiss industry's employees their jobs, falling from 90 000 to 28 000. Another example is Blackberry's failure to predict the adoption of touchscreens and the inability of Nokia to imagine a use of mobile phones beyond simple communication devices [33]. This inability of Nokia to keep up with innovations in the mobile phone space, namely the introduction of Apple's iPhone and Google's Android, caused its global share of the smartphone market to drop from 39% to 29% in the span of two years (2008-2010), at the same time reducing operating profit by more than 50%, making the corporation as a whole unprofitable and ultimately resulting it the divestment of its entire mobile phone business unit to Microsoft in 2013 [34]. By then the number of employees fell from 125 829 in 2008 to 59 333 in 2013 [34].

Timely knowledge of technology evolution patterns might help prevent potentially highly impactful strategic mistakes when choosing technologies to be developed. This does not only apply to stakeholders operating at a strategic level, as stakeholders operating on a smaller, project-based scale might also benefit from this knowledge.

Finally, stakeholders not directly involved in the development of technologies have a constant need for accurate evaluations of technology states. Primarily, these are stakeholders tasked with making high-level decisions, such as policy makers or venture capitalists. Even a minuscule reduction of uncertainty might enable these stakeholders to make better decisions. Understanding the current life cycle stage of a technology, as well as potential future development, might increase the ability of these stakeholders to introduce timely legislation governing the implementation of these technologies. Examples can be made of drone technology [35][36] and blockchain technology [37], both of which are highly disruptive technologies that policy makers did not predict, making their implementation into everyday life controversial as government policy attempts to keep up with innovation [35][38]. Stakeholders involved in venture capital (VC) funds are chasing optimal risk/reward opportunities. VC funds involved in the technology sector could have significant benefits from identifying technologies in stages suitable for investors and could also prevent potential losses by identifying technologies in the final stages of their life cycle [23].

1.1. Research Focus, Aim and Hypothesis

The evolution of technology is a complex phenomenon consisting of and being influenced by a multitude of external and internal factors [39]. While some of these factors can be qualitatively and quantitatively described, specific external influences shaping the development of technologies, such as government regulations or unpredictable creating a need for new technologies, cannot be predicted. Consequently, models attempting to describe the evolution of technology usually only focus on a small number of phenomena related to a technology's evolution. Examples include examining the rate of diffusion speed [40], external influence [41], development stage of technology [23], patterns of diffusion [42] as well as various technology life cycle matrices (a more thorough overview of these studies is presented in Chapters 2 and 3). An understanding of multiple phenomena should ideally enable the creation of a holistic and multifaceted model of technology change, incorporating contributions from different types of research, both quantitative and qualitative, in order to create a "bigger picture" overview of a technology's evolution.

The research presented in this thesis focuses on exploring how knowledge flows within a technology domain correlated with the life cycle stages of the examined technology using formalized records of invention as proxies for technologies. Patents are the formalized records of innovation chosen to be used these proxies, with patent citations being the primary metadata contained in patents being used as the basis for research. It is hypothesized that understanding the underlying phenomena related to the dynamics of patent co-citation might expand our understanding of the flow of knowledge within a technology domain, consequently enabling the exploration of potential future knowledge flow within a technology domain. This would further contribute to understanding the mechanisms governing the evolution of a technology domain.

This research aims to create a methodology for analyzing a technology's life cycle in a novel way, using formalized records of inventions. Moreover, based on the results of this analysis, an additional aim is to gain insight into previously undiscovered patterns governing the evolution of technology, exploring the intuition governing both mature and disruptive technologies and comparing the findings. Finally, it is explored whether the discovered intuition can be used to predict the future evolution of a technology. It is worth noting that a guiding principle in this research was the creation of a methodology that was as open and accessible as possible in all of its steps. This means that preference is given to tools and resources that are free to use, making the resulting methodology accessible to most researchers.

Research aim: *This research aims to develop a model of the evolution of technical inventions in a contemporary socio-technical context. An improved model to quantify the dynamics of evolution of technical innovation and technology implementation is to be used to explore potential future directions of the development of technology to reduce the uncertainty of decision-making in development projects.*

Hypothesis: *The proposed research will verify the hypothesis that, based on the existing records of technical inventions, it is possible to model the dynamics of a technology domains development and gain insights into the potential future directions of technology development.*

1.2. Research Methodology

Research in the field of knowledge and technology management involves the formulation of models and theories about phenomena in the environment, as well as the creation and validation of knowledge, methods and tools based on these models and theories with the aim of improving the process of modelling technology evolution and predicting the outcome of the development of technical systems and technologies. The current trend of research combines qualitative and quantitative methods and synergizes these two approaches in a combined model that would meet the goal of research. In general, when defining the research methodology in this area, it is necessary to consider the fact that the nature of research of socio-technical phenomena is heuristic. The research methodology to be used during the making of this thesis follows an approach presented in the Design Research Methodology (DRM), a general research methodology devised for design science [43]. The research methodology can be characterised by four stages, namely the Research Clarification stage, Descriptive Study I, Prescriptive Study I and Descriptive study II. An overview of these stages, as well as the corresponding thesis chapters, is presented as follows:

1) **Research Clarification** (Chapter 1)

The first stage of the research methodology focuses on providing an overview of the research focus and goals, clarifying the research problems and stating the aim of the research as well as the research hypothesis. The clarification of research problems includes forming the line of argumentations from the state of the art to the research goal.

2) **Descriptive Study I** (Chapter 2 and Chapter 3)

Preliminary research and literature review. The beginning of research requires a review of existing scientific and professional literature in the research area. Based on the literature review, an initial description of the current situation will be established, as well as a description of the desired results, with the aim of defining the basic assumptions of the research. For a detailed description of the current situation and guidance for further research, empirical research will be conducted in the form of an analysis of the existing development of technical innovations. As a result of this step, the research objectives are defined (under research goals). Also, the main research problems, questions, and hypotheses are identified. Moreover, the relevant disciplines and areas that need to be included in a literature review and existing approaches are also defined.

3) **Prescriptive study** (Chapter 4 and Chapter 5)

Based on the review and understanding of existing knowledge related to the problem, this stage aims to propose new methods and models to predict the dynamics of technology development. The development of models and methods involves the inclusion and synthesis of the theoretical principles of socio-technical systems, the development of technical systems, management of intellectual capital and the available empirical data. During this step, the main descriptive elements of intellectual property (patents) that can be used to model the development of technological innovation will be identified. The impact of the development of technical innovations on technology development will also be explored. Moreover, opportunities for improvement of existing models to predict the development of technologies and their applicability to the use of data obtained from patents will also be examined. Finally, rules and influential patterns discovered in the research so far will be identified, and their applicability in developing new models will be analysed. After that, the identification of impacts between individual indicators and the development of qualitative and mathematical a priori models will follow.

4) **Descriptive study II** (Chapter 6, 7, 8 and 9)

Initially, it is necessary to conduct a detailed patent analysis and an analysis of the history of the development of technical innovations in a particular technology domain. Furthermore, attention should turn to other potential sources of information such as scientific journals and white papers and their applicability to predict the development of technology should be assessed. Finally, we should not neglect data collection methods grouped in the category of "expert knowledge", that gather data from experts in a particular area.

Methods for data analysis will be applied to the data collected. As the research presented in this thesis focuses on exploring macro trends, it is assumed that the collected data samples will be extensive. Consequently, the applied data analysis methods should be suitable for the application on large sets of data. The data will also be modelled as a graph that will then be visualized and analysed. An analysis of changes in the value of data over time to define trends during technical innovations and technology development will also be conducted. The results of the research in theoretical and practical terms will be confirmed in the final step, and so will the accomplishment of the targeted theoretical and practical scientific contributions. By evaluating the research, achieved results will be compared with the research goals, and the advantages and disadvantages of the methods applied will be pointed out. The outcome of this phase will include proposals for improvements and implementation guidance for forecasting real technology development processes. Based on the conclusions and findings of the final research phase, guidelines for future research may be pointed out. Given that the topic of research is modelling the evolution of technical innovations, verifications of the model is only possible by comparing predicted outcomes with the actual outcomes, which is hardly feasible as a part of this thesis due to the time constraints. One solution to this problem is the application of the method of backtracking, which will be used in this research.

1.3. Expected Contribution

This research aims to contribute in both a theoretical and industrial/managerial context. The contribution in the theoretical context consists of expanding the theoretical fields of study related to the field of knowledge and technology management. This is accomplished by gaining new insight into how a technology domain evolves, first by deepening the understanding of how a technology's life cycle can be determined based on formalized records of invention, namely patents, and then by exploring the intuition governing the flow of knowledge within a technology domain. Finally, the aim is to demonstrate that having a deeper understanding of this intuition enables researchers to explore the existing knowledge flow within a technology domain and predict the future flow of knowledge.

The expected contribution in a managerial context consists of reducing uncertainty for different decision-makers and stakeholders. Companies involved in the creation of innovative products are the primary potential benefactors of this research, having a need for up-to-date information of technologies life cycles as well as the potential development of a technology domain, reducing uncertainty in decision making on both strategic and project levels. Determining the direction of the development of a technology domain can support strategic and long-term planning of the development of products, processes, and services as well as broader corporate strategy dealing with expanding a company's knowledge portfolio, whether by internal research and development or by acquisitions of external intellectual property.

The expected contribution in the theoretical context consists of expanding the current body of knowledge studying technology change, more specifically the field of studying technology change based on patent data. This is primarily accomplished by introducing a new method for exploring the life cycle of a technology domain based on patent citations, correlating the dynamics of patent citation network growth with technology domain life cycle stages. Moreover, a contribution is made in exploring the underlying intuition of knowledge flow within a technology domain by exploring the creation of patent co-citations within a technology domain.

The expected contribution of the proposed research, as part of this thesis, is manifested through:

- The development of a model for quantifying the dynamics of evolution of technical invention and the implementation of technology.

- The development of a tool, or tools, for simulating the potential future directions of the development of technology, that will be used for decision-making in development projects.

1.4. Thesis structure

This thesis is structured into nine chapters, and its outline is shown in Figure 1.

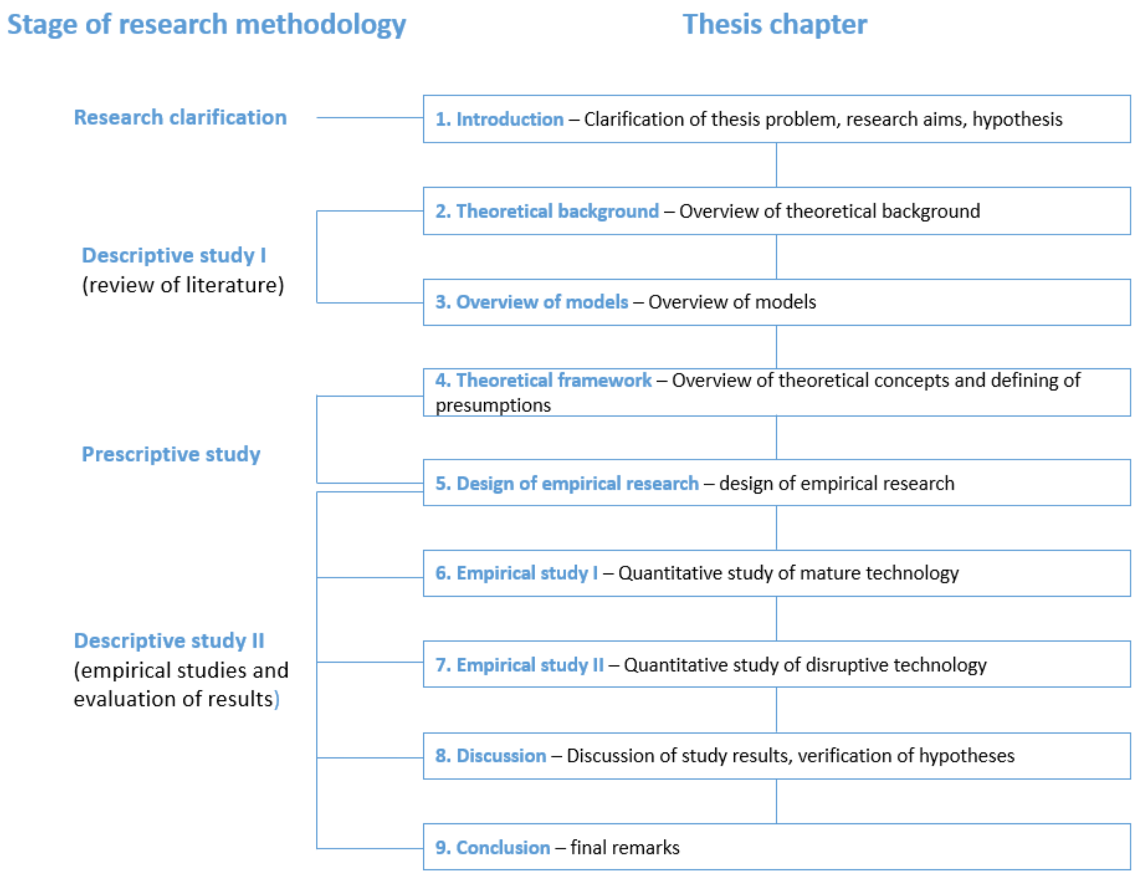

Figure 1 Overview of thesis chapters

Chapter 1 introduces the thesis problem statement, presents the research aims, expected contribution and the central thesis hypothesis. An overview of the research methodology used to conduct the study is also presented.

Chapter 2 provides a review of literature, identifying the most important theoretical concepts related to studying the evolution of technology, as well as the fundamentals of patents theory. Key findings and concepts related to the field are presented, and a structured presentation of the most crucial research and existing approaches is provided. The review of the state-of-the-art literature is segmented into three main sections, each aimed at reviewing the literature from the main areas studied in this thesis, namely patent theory, technology evolution and technology forecasting. Gaps in the literature are identified, and the presented research is contextualized along with those gaps.

Chapter 3 continues the review of literature, focusing on models and methodologies predominantly using patents to explore the evolution of technical inventions within a technology domain, provide insight into its current state and attempt to predict its future development. Based on the review of literature presented in Chapters 2 and 3, the research gaps are identified, and research questions are formulated.

Chapter 4 provides a theoretical framework upon which this thesis is built. Theoretical concepts used in the research are defined, and prepositions hypothesized to be true are stated, contextualizing the research. The choice of each theory is elaborated, and its strengths and weaknesses compared to alternative approaches are presented. Finally, the relationships between the various elements and concepts within the model are described and visualized.

Chapter 5 presents the design of the empirical research used to achieve the research aims. The process of collecting and sorting data is presented, as well as the related data engineering process used to prepare the data for further analysis. The approaches and tools presented in chapter 3 are expanded and presented in the context of a unified research methodology. The tools considered for forecasting are presented, and the methodology for choosing the tool to be used is described, as are the methods used to verify the results.

Chapter 6 presents the setup and results of the first empirical analysis conducted based on the methodology outlined in chapter 4. The setup of the empirical study is shown first, first providing a short overview of the technology domain being studied followed by an overview of the collected patent dataset representing the studied technology domain highlighting key characteristics. The results of the empirical analysis are presented and contextualized within previous research. Insights from the results are presented and commented on.

Chapter 7 presents the setup and results of the second empirical study, this one focusing on the technology domain of an emerging and disruptive technology.

Chapter 8 discusses and relates the findings of the two empirical studies to the literature review and theoretical framework, contextualizing these results into the broader research. The results of the two empirical studies are presented and contextualized within the research. Each of the research questions is addressed and discussed based on the presented results, and the extent to which the research questions have been clarified is used to confirm the guiding hypothesis of this research.

Chapter 9 provides a conclusion, summarizing what was done in this thesis, what the discoveries were and the implications to the broader field of study. The answers to the research questions are given and the significance and implications of research findings are presented. The limitations of the study are discussed and questions and guidelines for future work are stated.

2. THEORETICAL BACKGROUND

Chapter two provides the theoretical background of this research by presenting a review of the literature related to the fields of study. The fundamental concepts related to the exploration and prediction of the evolution of technology and the fundamentals of patent theory are reported. This is the first part of the review of literature, focusing solely on the theoretical background. A continuation of the review of literature is made in Chapter 3, focusing on presenting an overview of the methods and models used in the related research as well as defining the research gaps and research questions.

This research is interdisciplinary in nature, combining theories from different fields of study, the synthesis of which allows for the creation of a theoretical framework on which the research will be based on. Figure 2 shows an Areas of Relevance and Contribution (ARC) diagram as described by Blessing and Chakrabarti [43]. This representation clarifies the foundation on which the presented research is based and areas of contribution of this research. Consequently, Figure 2 provides a visualization of the fields relevant to the topic of the research, their relative importance to the research as well as the areas of the researcher's contribution.

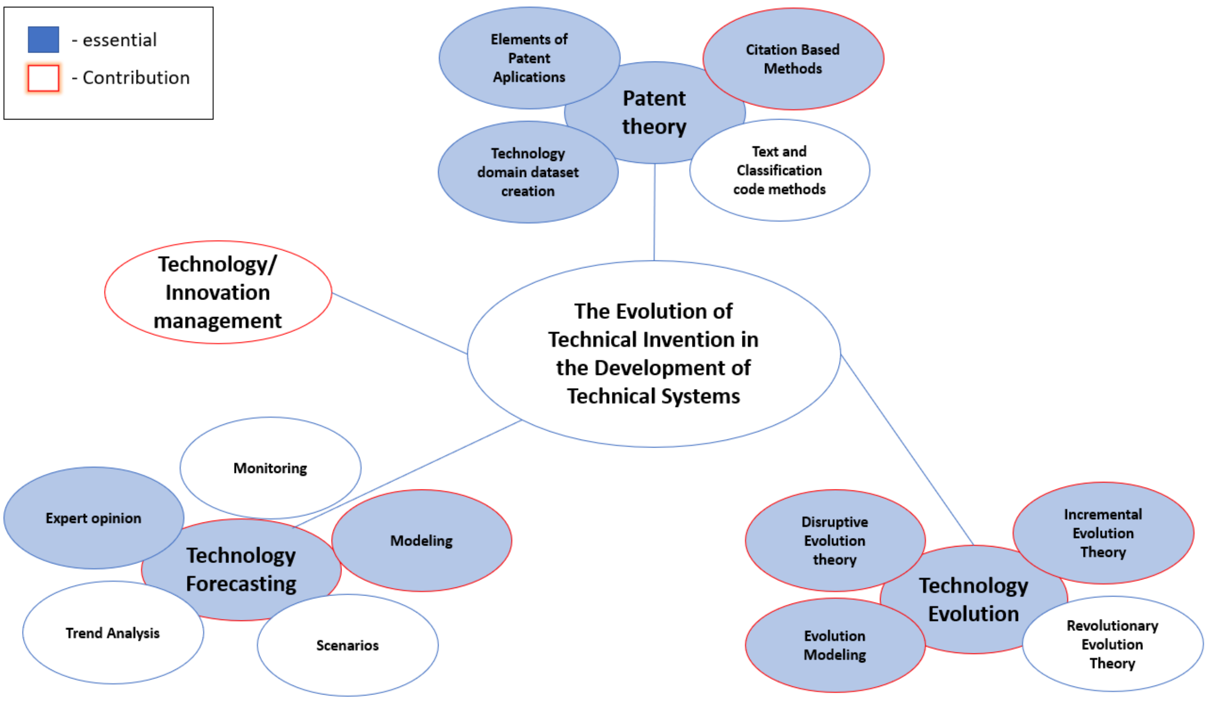

Figure 2 ARC diagram

As this research uses patents as the primary source of data and a proxy for inventions [2], a thorough overview of patent theory is made, focusing on the elements of the patent application that can be used to study the evolution of technology and the architecture of patent applications themselves.

An overview is also made of theories related to exploring the evolution of technology. The theoretical background of technology evolution is explored, presenting the most prominent theories which attempt to explain how technologies evolve. These theories are compared and their respective distinctions noted.

Finally, a review is made on the state-of-the-art literature concerning forecasting the development of technology. The most popular methods for forecasting technology development are presented and an overview of the relevant research is made. Finally, a comparison of all of the presented methods is made and the gap in the literature is identified.

Before proceeding with the literature review, it is essential to define the key concepts from the research fields influencing this study. The reason for this a noticeable discrepancy in the nomenclature and definitions of terms through multiple fields of study and even different researchers within the same field of study. Consequently, it is crucial to define the scope and definitions of important terms as they will be used in this thesis. Table 1 shows an overview of the key concepts used in this research as well as their definitions.

Table 1 Key concepts and their definitions

Concept	Definition
Technology	The definition of technology used in this research is the one given by Burgelman [44]: “Technology refers to the theoretical […] knowledge, skills, and artefacts that can be used to develop products and services as well as their production and delivery systems”. This definition is supplemented by Bunge, stating that the technology is applied scientific knowledge for the attainment of practical goals [45]. The distinction of this definition is the breadth of concepts it considers a technology. Moreover, it emphasizes the need of technology to be a facilitator of value creation and restricts it to explicit manifestations.
Invention	The creation of a product or introduction of a process for the first time [46]
Innovation	An improvement of significant contribution to an existing product, process or service. Innovation includes new technologies, transactions, and its stakeholders such as collaborators and customers [4]. As a rule, innovation flows from invention.
Technical Invention	An invention in a technical field, a new solution to a technical problem [47]
Technology Domain	Like technology, this term has multiple vaguely different definitions in literature. Moreover, different researchers use different terms to describe roughly the same concept (technology space [48], technology field [49], technology domain [50]). In this research, it is settled on using the term "technology domain" and this uses is consistent with its use in previously published research. The definition of technology domain as outlined by Boyack et al. [51] is used, viewing a domain as a sphere of knowledge. In the context of technology, a technology domain is considered a sphere of knowledge pertinent to that technology.
Technology Evolution	An organic analogy that explains the appearance and selection of novel artefacts in the technology space [25].
Technology Forecasting	Exploring future changes in technology, such as its functional capacity, timing or significance, among other attributes [20]
Technology Life Cycle	A description of a technologies journey from initiation to its eventual decline consisting of stages, with each stage having some identifiable and unique attributes. Usually consisting of two dimensions, the competitive impact and integration in products or processes. [24]

2.1. Technology Evolution

Technology evolution is an umbrella term covering a breadth of theories on how technology changes and improves over time. Taking inspiration from the process of biological evolution, it seeks to describe the development of technology using analogies of the same mechanisms which govern the evolution of living things in nature. There is a significant volume of research covering the evolutionary nature of technology change and attempting to model this change [52][53][54][22] [9]. From this literature, an established archetype of a technology evolution can be observed, describing generalized patterns and rules governing the evolution of a technology.

The early emergence of a new technology is usually characterized by high technological turbulence as competing technologies vie for acceptance within a field. Also, this time sees a significant level of technical and market uncertainty [55]. Companies engage in trial-and-error, attempting to resolve this uncertainty, and eventually, a technology emerges as the new industry standard. This emergence is often influenced by legislative, economic, social and political forces [26] [56][57][58]. The initial period of uncertainty is followed by a period of incremental technology evolution. This period of incremental change is accompanied by a reduction of meaningful innovation and focus is placed on cost reduction and minor improvement to components or subsystems [59]. Nevertheless, this incremental innovation is punctured by technological discontinuities and radical technological changes facilitating radical industrial change [8][55][27][60].

Consequently, it can be argued that a technologies evolution consists of both incremental and discontinuous changes. This cycle of incremental and discontinuous technology evolution has been studied extensively and in multiple technology cases, both as discrete cases and as parts of a technology evolution cycle.

There is a consensus among researchers studying the evolution of technology that technology evolution generally follows a Darwinian model of natural selection [61][8][9]. However, while the concept of technology evolution does take inspiration from biological evolution, certain researchers have noted that this Organic - Mechanical analogy is merely an approximation and has its limits [44]. Certain traits found in biological systems are not found in technology, primarily self-replication and survival of the fittest. While technologies compete for market dominance, it is the owners of the technologies, not the technologies themselves, who do the fighting. Moreover, several more key differences are noted by Albert [44] :

- Technologies have no "individuals", meaning they can only be observed as the results of a process. Consequently, technologies being developed in different institutions may cause evolutionary branches that are different in their development.
- State changes in technologies may not be obvious by clearly observable events or transition phases.
- In biological organisms (i.e. mammals), the time of death is intrinsic; it is not related to the life cycle of a descendent. However, there is a direct observable correlation between the decline of one technology and the emergence of a new one to supplant it.

Nevertheless, the analogy of technology evolution is still a valuable illustration of the development of a technology [25].

Theories of technology evolution can generally be segmented into three groups, each describing a different underlying pattern of evolution. The first two theories are those of incremental evolution and radical evolution. Most researchers agree that the evolution cycle of a technology consists of alternating incremental and radical phases [22]. However, a third theory of technology evolution arises, that of disruptive evolution describing a special case of evolution that is highly unpredictable and causes a significant impact within a technology field and related industries. It should be noted that radical and disruptive theories are often labelled together as simply "radical" [26]. However, in this research these two theories, and their respective distinctions, are presented separately while acknowledging that they too share some characteristics, as it is assumed that a reduction of the two theories to a single theory would not provide an adequate representation of technology evolution theories.

In the following subchapters, the three dominant theories of evolution will be presented in more detail and their most significant characteristics will be highlighted.

2.1.1. Incremental evolution

The incremental technology evolution theory views innovation as a gradual accumulation of slight variations over time that yields a novel innovation. Change is slow and inevitable, and there is no room for radical leaps [25] as technology architecture remains stable and companies place their focus on innovations related to the optimisation of the manufacturing process, cost reduction, component improvement and customer segmentations [59][61][27][62]. This type of evolution is still the most dominant in most industries representing a low-risk/low-reward approach to technology management and is marked by organizational, social and political stability [59]. A primary reason for this approach is the high risk associated with developing radical products [19]. Iyer et al. [26] show that approximately 85% of new products fail, reinforcing the notion that incremental innovation represents lesser risk.

Consequently, organizations focused on incremental innovations develop technology at a linear and steady pace, investing in incremental improvements of strategic technologies [4]. These improvements often mark a shift from product to process innovation, focusing on cost reduction and minor component and subsystem innovation [27], and the period of incremental evolution usually lasts until a radical change in invention or the introduction of a disruptive technology. The period of incremental change usually occurs in the later stages of a technologies life cycle, where a mature technology reaches a physical or economic application limit [63].

The level of innovation decrease in the incremental period of the evolution of the technology is observable in the nature of inventions being created. While the initial stage of a technologies life cycle sees invention with a high level of novelty, incremental development sees minor new developments to existing technologies, as most new inventions are modifications of existing products [16] or recombinations of existing technologies [64] [65][31][25].

However, this type of approach at developing technologies has several drawbacks. Strong adherence to accepted beliefs and established technologies make companies less likely to create a radically new and innovative product [16]. Moreover, it makes companies susceptible to unpredictable leaps in technology evolution. There is a number of examples of companies facing problems as a result of disregarding technology leaps and focusing on incremental development of existing technologies [59].

2.1.2. Radical (discontinuous, revolutionary evolution)

Revolutionary innovations are innovations that are not the results of an incremental (linear) development [4]. They are a radical departure from the norm of incremental improvements and are therefore termed discontinuous. These technological discontinuities are technologies that have a definitive cost or quality advantage and strike at existing firms' foundations [27]. However, these discontinuities do not necessarily depend upon the emergence of a completely new technology, as a sufficient recombination of existing ideas, technologies, or knowledge artefacts may also result in technological discontinuities [65]. Hoisl et al. [28] state that technological discontinuities can be determined in two ways. One is externally determined, i.e. by a crisis or regulatory changes introduced by governments and standard-setting bodies. However, the more common ones are endogenous, i.e. determined by economic forces and consequently influenced by companies from a relevant industry.

Tushman et al. [66] state that technology evolves incrementally until this incremental change is punctuated by a significant advance. These advances cause a discontinuity in progress and are often reflected in the emergence of new product classes (automobiles), product substitution (transistors replacing vacuum tubes) or fundamental product improvement. However, a categorical distinction between incremental and radical innovation is often hard to determine, as the demarcation lines can be context-specific and unclear [67][68]. Nevertheless, there are enough differences between incremental and radical innovations to discern the two [4]. A period of radical innovations usually occurs at the start of a technologies life cycle [28], characterised by a significant number of original inventions [66], unlike incremental innovations, which mostly consist of recombination of existing technologies. However, radical innovations also occur in the other stages of a technologies life cycle, signifying a noteworthy shift in the technology field.

Technological discontinuities caused by radical innovation can be classified as competence destroying or competence enhancing [66][29]. Competence destroying discontinuities, so-called because they destroy the competence of existing firms within the industry, require new skills and abilities as the mastery of the new technology alters the set of necessary competencies within a product class. This type of discontinuities either create a new product class or create new substitutes for existing products. On the other hand, competence enhancing discontinuities offer significant price or performance improvements of a product class, building on existing know-how.

Regarding the causes of radical innovations, Nemet studies the extent to which demand-pull policy measures, government policies that induce investees, stimulate discontinuous technology change [69]. However, he finds no evidence that demand-side policies encouraged discontinuous technical change. In fact, a negative relationship is suggested. In his description of a model that describes the way discontinuities occur, Funk [70] combines three arguments related to discontinuous innovation: The first two are how incremental improvements in the components of a product impact the performance and design of the product and how these incremental improvements can lead to discontinuities in product design through their impact on the design trade-off. The third argument is how these incremental components can combine to create entirely new products. The factors that cause discontinuities in a mature industry are also studied by Tripsas [55], claiming that discontinuities occur when products that are based on fundamentally different principles invade an industry, displacing products based on the prior technology.

2.1.3. Disruptive Evolution

The idea of disruptive technologies was popularized by Christensen in his work The Innovator's Dilemma [9]. There, the author provides examples of disruptive technologies and defines them as “technologies that provide different values from mainstream technologies and are initially inferior to mainstream technologies along the dimensions of performance that are most important to mainstream customers”. In other words, it can be argued that disruptive innovations cater to undefined markets. Much like radical leaps in a technology’s evolution, disruptions are also products of technological discontinuities. However, while radical technologies simply signify a significant departure from a linear development, disruptive technologies by definition cause a disruption in the current technology space and transformation of the mainstream market. Therefore, for a discontinuous innovation to be disruptive, successful implementation is vital [71][72].

An issue when studying the theory of disruptive research is the scattered and conflicting nature of the relevant literature. Yu et al. [73] seek to rectify this by comparing the different research related to disruptive innovation and clarifying the basic concepts of disruptive innovation theory. Expanding on the definition provided by Christensen [9], they state disruptive innovation as “a powerful means of broadening and developing new markets and providing new functionality, which in turn, man disrupt existing market linkages.”

Early in the development stages of a disruptive product, disruptive technologies serve only specific niches which have a need for its non-standard attributes [73]. As a disruptive technology further develops, it experiences further widespread adoption. Regardless, the performance of the disruptive technology still lacks behind the incumbent technology. The market disruption occurs when the new, disruptive product displaces the mainstream product in the mainstream market, despite its inferior performance on attributes valued by customers of the existing product. In order for this to occur, two conditions must be met: Performance overshoot on the principal mainstream attributes of the existing product and asymmetric incentives between existing healthy business and potentially disruptive business [74][71].

Examples of disruptive technologies include, but are not limited to, drone technology, autonomous robots, blockchain technology, neuromorphic hardware [75][38].

2.2. Overview of Patents

Patents are an instrument of protecting one's intellectual property (IP) [76][2]. Historically, patents have been the predominant form of intellectual property protection in the fields of mechanical, electrical and chemical engineering as well as the field of thermodynamics. Recently, biological patenting has also become increasingly important [76] with the rise of the field of biotechnology. The volume of knowledge contained in patents is immense and is constantly increasing, with there currently being over 6 million patents. This number increases by approximately 150 000 patents per year [3]. The study of technological change has traditionally been aggravated by a lack of relevant and accessible data. The use of patents seems to contribute to the solving of this problem and patent are widely considered as the best-structured records of inventive activity, covering a breadth of innovation fields [3].

Patents add a number of advantages as technological indicators. Primarily, patents are publicly available information, providing an abundance of detailed information, both structured and unstructured [77]. Additionally, each patent produces an exceptionally structured document containing accurate information about the invention itself, the technological area it belongs to, the inventors and the organization to which the property right is assigned [3].

For a patent to be granted, the examination process must conclude that it is both novel, innovative and useful. A patent is intended to represent only one invention consisting of several closely related and integrated technologies that, acting in synergy, accomplish a particular task [78]. However, it is essential to note that not only granted patents are used as technological indicators in research, as patent applications (patents not yet granted) also contain knowledge about inventions, even though these inventions might not meet the criteria necessary for a patent to be granted.

Patent applications consist of the following elements [76]:

- Abstract – provides a summary of the technical information disclosed in the patent.
- Specification – Provides the technical details of an invention as well as its novelty compared to previous inventions. It consists of three parts: The background, the description and the claims.
- Background – provides an overview of the prior art related to the patent. Prior art is a description of all relevant technologies preceding the patent as well as

other pertinent information, showing work that the patent cites (both previous patents and previous scientific papers). It is crucial in determining the novelty of the patent, i.e. how it differs from prior inventions.

- Description –provides a complete description of the patent and detail how it works. As defined by US patent law, a patent's description must be detailed enough "so that any person of ordinary skill in the pertinent art, science, or area could make and use the invention without extensive experimentation".
- Claim – provides a very narrow, precise statement of what the invention is. It is arguably the most crucial part of the patent because it defines the scope or boundaries of the patent's invention. This scope in a patentable invention is specified with a series of numbered claims illustrating the patent's novelty. In technical and precise terms, the claim must state the subject matter of the invention, as well as its purpose, operation, properties, and methods it employs [78]. A single patent may contain several claims.
- Drawings – most, but not all, patents contain drawings that illustrate all of the features declared in the claims section.
- Declaration – a declaration by the inventor that he has disclosed all relevant information important for the patent application and that he is the original inventor.

Another important piece of information contained in a patent is the classification codes used to classify the patent. Classification codes provide a hierarchical system of language independent symbols for the classification of patents and utility models according to the different areas of technology they pertain to. The classification codes themselves differ in different patent offices, although recently a new classification system, the Cooperative Patent Classification (CPC), has been introduced. In this harmonization process under the IP5 Common Hybrid Classification initiative, the United States Patent and Trademark Office (USPTO) and the European Patent Office (EPO) have agreed to work toward the development of a joint classification system based on the European Classification System (ECLA) and aligned with the WIPO classification standards and the IPC structure [79][80][81]. CPC also follow a hierarchical structure but has more subgroups than the IPC [82].

Classification codes describe the set of technological abilities at a given time [78]. As patents by definition contain technological novelty, the examiner assessing a new patent application must decide whether an existing code can be used to describe the patent or whether a new classification code must be introduced. The introduction of a new code requires a backward reclassification of all patents which may contain the coded technology, meaning even older inventions will still be classified under the newly introduced code if applicable. Therefore, at any given time, the existing set of technology codes essentially provide a description of the current set of technological capabilities [64].

Table 2 shows the nine CPC sections with their respective definitions.

Table 2 The nine CPC sections

Code	Name
A	Human necessities
B	Performing operations; transporting
C	Chemistry; metallurgy
D	Textiles; paper
E	Fixed constructions
F	Mechanical engineering; lighting; heating; weapons; blasting engines or pumps
G	Physics
H	Electricity
Y	Emerging Cross-Sectional Technologies

The complete classification code contains information about a patents Section, Class, Subclass, Group ad Sub-Group. Figure 3 shows an example of a classification code and its breakdown.

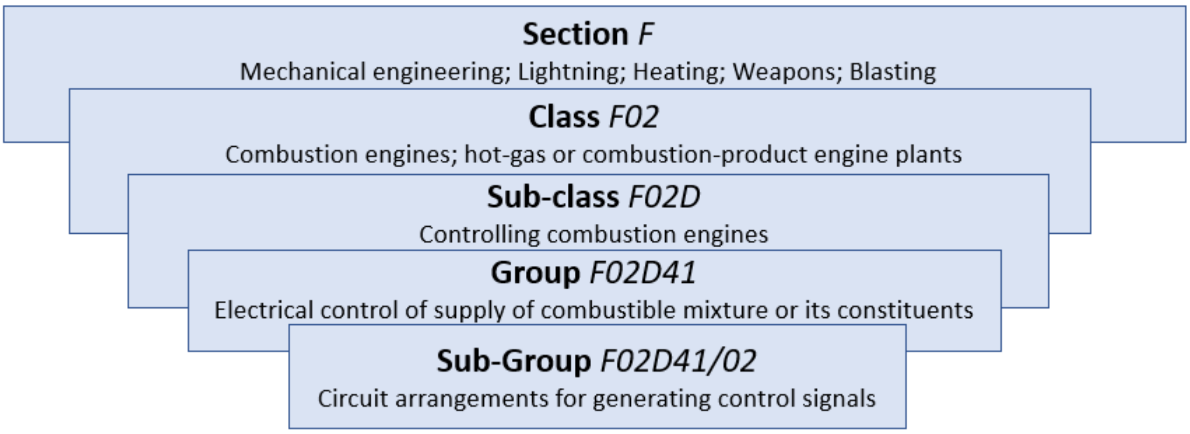

Figure 3 Complete patent classification symbol scheme [83]

In Figure 3, the label *F02D 41/02* consists of section *F*, the class *F02*, the subclass *F02D*, the main group *F02D41* and the subgroup *F02D 41/02*.

Alongside classification codes, an element of patent applications suitable for analysis are the citations related to the patent. Patents include citations of previous patents (backward citations), which show how knowledge embedded in previous inventions influence new inventions [84] and also include patents that cite the examined patent (forward citations). As no other measure exists representing the influence from one invention on another, patent citation is a promising measure of knowledge flow [85]. Patent citations allow for the study of knowledge spillovers and the creation of indicators of the technological impact of individual patents [3].

Patents as proxies for technical invention

In the context of exploring the evolution of a technology, patents have a history of being used as proxies for innovation and technologies in research [2][86][31][87]. The underlying assumption is that most research and development activities generate innovations which are in turn protected by innovators as their intellectual property in the form of a patent. Consequently, patents can be considered proxies of innovation [88]. According to the work of Yoon et al. [18], patents can also be considered as a container of knowledge elements, with Phelps et al. [89] calling them "discrete knowledge artefacts", alongside other knowledge artefacts i.e. papers and products. Pereira et al. talk about the contents described in patents being an important source of information and basis for knowledge, containing technical details and particularities of the author's inventive intentions [2]. The citation information contained in patents captures the directed relationships between patents and clarifies technological antecedents and descendants [3]. At its core, citation-based analysis of the patents presumes a cumulative view of the process of the development of technology, with each inventor benefiting from the work of previous inventors and subsequently influencing the work of future inventors [3]. These directional relationships can illustrate the flow of information (or knowledge) from patent to patent, demonstrating the spillover of technology through clearly separated entities. In the context of knowledge management, a patent citing another implies that the cited patent contains a piece of existing knowledge that the citing patent builds upon [86]. The types of citations being analysed differ among researchers, with the most common being direct citations, co-citations and bibliographic coupling [90]. Direct citations are defined as an antecedent–descendent link, while co-citations are defined as linking patents being cited by the same patent and bibliographic couplings are defined as linking patents that cite the same patent [90].

It has been demonstrated that patents are formalized records of technical inventions as well as containers of knowledge elements. Direct patent citations represent the generational flow of knowledge between patents, providing insight into how previous inventions influence future inventions, i.e. the knowledge relationship between antecedent and descendant patents. Therefore, a collection of patents from the same field can be said to form a representation of a technology domain, i.e. a collection of knowledge artefacts representing a knowledge domain. Consequently, it is demonstrated that patents can be used as proxies for technical inventions and increasing the understanding of the dynamics of their citations can provide insight into the dynamics of knowledge flow between technical inventions within a technology domain representing the evolution of the technology domain. Furthermore, based on the definition of technology used in this research, stating that technology is the application of scientific knowledge for practical purposes, patents can be used as proxies for new technologies as they by definition must be new and have a practical application.

Some significant limitations of using patents as a source of data exist. Primarily, not all inventions are patented, whether because not all inventions meet the criteria for patentability (the invention has to be novel and nontrivial; the invention has to have commercial applicability) or the inventor chooses not to patent his invention. However, despite these drawbacks, patents are considered the best formalized records of technological invention. Moreover, in the context of this research, the mentioned limitations are assumed to have a negligible impact on the results, primarily because the large volume of patents being analyzed is deemed sufficient to extrapolate trends and patterns. The research also emphasizes exploring patterns of citations and co-citations, consequently making non patented inventions utterly irrelevant in the research context.

2.3. Technology forecasting

Technology forecasting is a set of methods and models that, focused on changes in technology, attempt to predict the future characteristics of a technology or technology domain and related procedures and techniques [20]. The term can, and often is, be used interchangeably with terms like "tech mining" [76], "technological forecasting" [91], "technical intelligence" [76], "impact assessment" [92], among others. While all are intended to enable better decision making, they differ in their targeted audience, problem conceptualization and methodology. However, in the scope of this thesis, the term *Technology forecasting* will be used to cover all attempts to predict the future development of technology.

Several approaches exist for classifying technology forecasting methods. Traditionally, technology forecasting methods were classified as either exploratory or normative [93]. Exploratory methods are outward bound methods for examining where present events and trends might take us in the future. On the other hand, normative methods are inward bound, beginning with an initial view of a possible future then working backwards to see if and how these futures might or might not develop from the present. While this distinction is useful, problems arise when methods cannot be clearly defined as either exploratory or normative. Thus, Porter and Rossini propose grouping forecasting methods into five families of methods [20]:

- Monitoring
- Expert opinion
- Trend extrapolation
- Modelling
- Scenarios

A further categorization of forecasting methods can be made depending on whether they are direct, correlative, or structural [20]. Direct methods are used to forecast the parameters that measure functional capacity or some other pertinent characteristic of a technology. Correlative methods relate a technology's development to the growth or change of one or several elements in its context considered to be analogous.

The most often forecasted technology attributes are [20] :

- Growth in functional capability, i.e. the ability of a company to manage technology efficiently
- Rate of replacement of an old technology by a newer one
- Market penetration
- Diffusion of technology
- Likelihood and timing of technological breakthroughs

It should be noted that technology forecasting is, in principle, not deterministic, i.e. it does not attempt to predict a single future. Instead, it attempts to predict a number of possible futures as well as their respected likelihoods [20].

In this thesis, the classification of forecasting methods proposed by Porter and Rossini is used as a basis for structuring this part of the review of literature.

2.3.1. Monitoring

While not a strictly forecasting method, monitoring is a crucial part of the forecasting process as it is the method most used to gather information about the state of the technology being studied. The sources for this information are numerous, from academic journals and technical databases to less reputable information sources available on the internet [94]. This abundance of data sources is also the main drawback of this method, as it becomes increasingly challenging to filter useful data as the total volume of available data increases. Therefore, finding reputable sources for technology monitoring is a constant priority. However, even reputable sources of technical data, such as patent databases, might prove challenging as navigating these databases requires a certain amount of skill.

2.3.2. Expert opinion

These methods are fundamentally based on opinions given by a panel of experts. Expert opinion methods are based on the assumption that experts have insights into a technology field not available to outsiders. These experts are often academics researching the technology field, stakeholders in charge of managing products based on the technology field or R&D engineers involved in developing products based on the target technology [95]. However, historically, expert opinion methods have a very poor track record when predicting future technology, their predictions being especially susceptible to radical innovations and disruptions [96]. A major drawback of expert opinions is the expert's ingrained biases when dealing with a single expert. When dealing with a panel of experts, problems of groupthink are often manifested, as well a tendency of dominant individuals forcing the rest of the group to adjust their opinion [97]. Regardless, if conducted properly, the expert opinion method can be a valuable method for technology forecasting [94].

The most representative expert opinion method is the Delphi method, a structured approach to extracting a forecast from a group of experts, with an emphasis on producing an informed consensus view of the most probable future. It is a form of a survey intended to provide anonymity to participants, iterative responses, controlled feedback, estimates of the likelihood/timing of technological developments and statistical response measure. A large number of derivative methods exist, all expanding the Delphi method to a wide range of applications [58].

Another popular expert opinion method is the Gartner Hype Curve [98] [99], used to characterise a typical progression of an emerging technology to its eventual position in a market or a domain. It consists of five stages and is designed to help stakeholders decide which companies are promising investments reinforcing the notion that companies should not invest a technology simply because it is “hyped” [99]. It should be noted that the Gartner Hype Cycle has been criticised for a lack of an underlying mathematical model [100].

2.3.3. Trend Analysis

Trend Analysis is a method based on analysing a large volume of historical time series data and using this data to extrapolate future trends [101][102]. The underlying assumption is that the future represents a logical extension of the past and can therefore be inferred by extrapolating the appropriate trends from the available data. However, this requires a deep understanding of all relevant factors that influenced the past trends, as well as any disruptions and non-repeatable events, the lack of which might make trend extrapolation not feasible.

2.3.4. Modelling

Modelling assumes the sufficient availability of information to create a model that will lead to a forecast at some future point, occasionally referred to as a causal model [20]. This is comparable to developing and solving a set of equations in order to represent some physical phenomenon. In their nature, models can be computer-based and judgement based. Computer-based models usually use quantitative parameters, consequently often omitting important qualitative subtleties necessary to create an accurate prediction. Judgement based models are more expert oriented, relying on the forecaster's ability to make predictions. In any case, the specific forecast produced by the model is less relevant than the underlying trends governing the model or insight into how different combinations of input parameters influence the model's output.

2.3.5. Scenarios

At their core, scenarios are stories told about the future, or more accurately, about possible alternative futures [103]. First used by the RAND Corporation during the Cold War, scenarios have since been used in order to explore the development paths of technologies as well as the method of their introduction into the world. Creating scenarios requires an understanding of the underlying science behind a technology as well as external contributing factors which might influence its future appropriation, ranging from economic and geopolitical influences to individual biases and needs [104]. One of the more popular scenario approaches is backcasting, where forecasters envision various possible future scenarios and then create paths that would lead to these scenarios. The distinction of this method is that, while other methods begin by analysing current trends and try to project the future, backtracking starts with a projection of the future and continues by going backwards. The purpose of the backcast is to identify signposts or tipping posts that might serve as leading indicators. In general, scenarios are an excellent way to communicate the results provided by other forecasting techniques. Moreover, the construction of a complete future scenario, or event leading to it, often reveal holes in a previous analysis. Therefore, they can be used to integrate quantitative data with qualitative information.

2.3.6. Overview and Comparison of Forecasting Approaches

Table 3 provides an overview of the presented approaches to forecasting the development of technology, stating their respective strengths and weaknesses.

Table 3 Overview and comparison of technology forecasting method groups

Group of methods	Summary	Strength	Weakness
Monitoring [94]	Gathering and organizing information about a technology	Large volume of information	Volume of information can be overwhelming
Expert opinion [95]	Experts give opinions about potential development of technology	Experts often have unique insights	Hard to identify experts; largely speculative
Trend Analysis [101]	Used to extend time series to the future	Data based quantitative method	Vulnerable to disruptions, requires a large amount of data
Modeling [20]	A simplified representation of reality	Models can capture the essence of a problem, filtering out noise	Sensitive to initial assumptions; bias toward quantifiable parameters
Scenarios [103]	A series of snapshots of paths leading from the present to the future		Highly speculative

Based on the insights from the review forecasting approaches, which are summarized in Table 3, it is clear that no methods exist which will provide forecasting of the development of all of the technology characteristics outlined by Roper et al. [20]. Moreover, while a forecasting method can be either qualitative or quantitative, it can be assumed, based on Table 3, that an effective method should consist of both a quantitative and qualitative approach to forecasting, taking advantage of the strengths of both approaches. The following chapter further explores these approaches, where models for forecasting technology development will be presented.

2.4. Summary of the State of the art

This chapter provides the state of the art of the three main fields of study explored in this research. The theoretical background of patents is presented, demonstrating their use as proxies for technological inventions and representations of a technology domain (Chapter 2.2). The theoretical background related to the study of technology evolution is presented, describing the most popular approaches to studying the evolution of a technology. An overview of the three types of technology evolution recognized by the literature is made, and the main characteristics of each type of evolution are presented as well as their connection to the life cycle stages of a technology domain. Finally, an overview of technology forecasting methods is presented from a theoretical perspective, providing a systematization of approaches as defined by Roper et al. [20], along with the strengths and weaknesses of each theoretical perspective. To conclude, this chapter presents an overview of the theoretical part of the literature review, providing a theoretical context for the research presented in this thesis. The literature overview is continued in the following chapter, Chapter 3, which provides an overview of methods and models used to describe and predict the evolution of technology using patent applications as the primary source of data.

3. OVERVIEW OF MODELS AND METHODS

Chapter three provides an overview of models and methods used to explore the evolution of technical inventions, provide insights into the current state of a technology domain and attempt to predict its future development.

In this chapter, an overview of patent-based models for exploring the evolution of inventions within a technology domain is made. These models represent the application of the concepts outlined in Chapter 2 and are focused on exploring the development of a technology domain using patents as proxies for technical invention. It should be noted that some of the presented models, considered in this research to be relevant to the state of the art, are not patent-based but are nevertheless included in this review.

Patent analysis is a collection of techniques, methods and tools for studying the knowledge contained in patents, and includes, but is not limited to, searching for relevant patents, extracting information for relevant patents and analyzing the extracted information. The importance of patent analysis has increased with the shift into a knowledge-based economy, where it is paramount to know whether a business depends on someone else's patent [76]. Recently, the strategic importance of patent analysis in a high technology context is increased due to the more complex nature of the innovation process, the shortening of the innovation cycle and the increase in market demand volatility [102].

In general, patent analysis methods can be segmented into two large groups of approaches: A micro approach, consisting of a detailed examination of a small number of patents, and a macro approach, studying a large number of patents [105]. On a micro level, patent analysis has primarily been used to study technological change on a single company basis, whether to develop technology plans [77], evaluate R&D investments [106] or monitor a company's competitiveness [107]. On a macro level, patent analysis is used to explore large scale trends, exploring the link between technology development and economic growth [108][106], estimate technological knowledge flows and compare the performance of innovations in an international context [102].

The research presented in this thesis focuses on studying trends on a macro scale, exploring how technical inventions evolve within a technology domain. Consequently, in creating an overview of models and methods, preference was given to patent analysis models and methods exploring technology on a macro scale. These patent analysis methods are usually systemically grouped into three main groups.

- Citation based models – use patent citations as a basis for analysis
- Keyword/text mining based – use extracted keywords from patent abstracts/claims as a basis for analysis
- Classification code based – use patent classification codes as a basis for analysis

Because a large number of methods combines keyword/text mining based methods with classification code based methods, the rest of this chapter sees the overview of methods segmented into two groups, the first one based on patent citations (Chapter 3.1) and the second one segmented into keywords/classification codes (Chapter 3.2). Finally, the final subchapter presents the identified gap in research and the research questions (Chapter 3.3).

3.1. Patent Citation Based Methods

The citation information contained in patents captures the direct relationships between patents and clarifies technological antecedents and descendants [3]. At its foundation, citation based analysis assumes the view that process of technological development is cumulative, with each inventor benefiting from the work of previous inventors and subsequently influencing the work of future inventors [3]. These directional relationships can illustrate the flow of knowledge from patent to patent, demonstrating the spillover of technology through clearly separated entities. The types of citations being analyzed differ among researchers, the most common being direct citations followed by co-citations and bibliographic coupling [90][109]. Direct citations are defined as an antecedent –descendent link, and co-citations are defined as linking patents being cited by the same patent. Bibliographic coupling is defined as linking patents that cited the same paper. Shibata [110] did a comparative study focused on investigating how these three types of citation networks performed with the goal of detecting emerging research fronts. When analyzing the citation relationships of a large number of patents, networks can be created where nodes represent patents and edges represent citations. A graphical representation of these networks is shown in Figure 4.

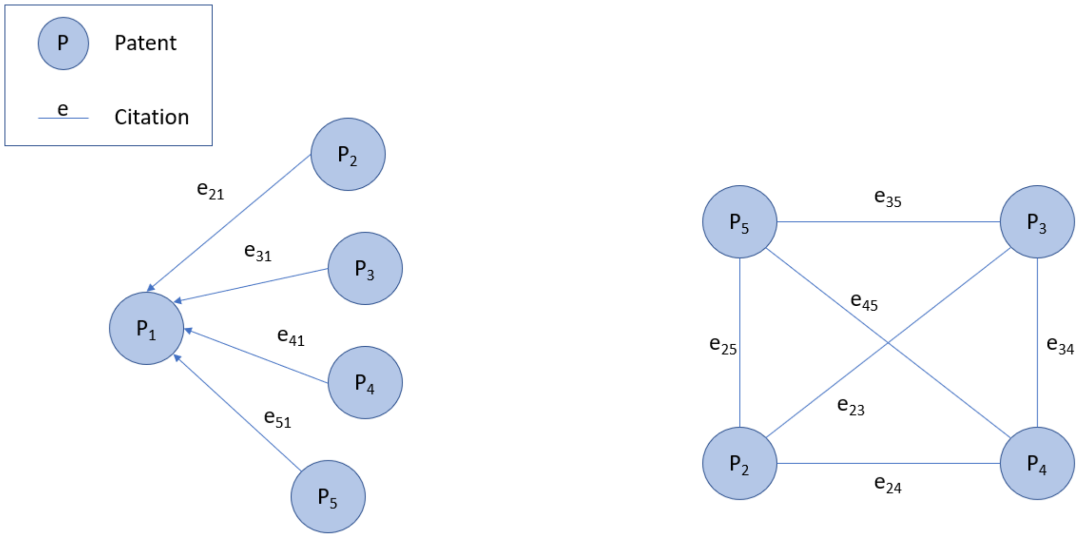

Figure 4 Visualization of direct citation network (left) and co-citation network (right)

If patent P_1 cites patent P2, then there is a direct citation between these two patents. If patents P2 and P3 are cited by patent P1, there is a co-citation between patents P2 and P3. If patents P2 and P3 cite patent P1, there is bibliographic coupling between patent P2 and P3. Patent citations can have two directions: Forward and backward. Forward citations are citations received by later patents as prior art and are often considered indicators of an invention's technological impacts. This relationship is illustrated in Figure 5. Therefore, a larger number of forward citations may indicate a higher value of a patent [111]. Backward citations refer to patents that influenced the citing patent and are therefore cited as prior art.

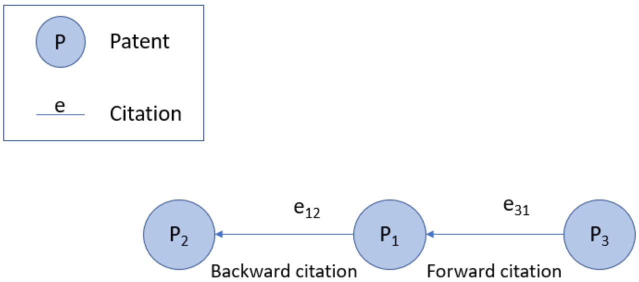

Figure 5 Visualisation of forward and backward patent citations

It is worth noting that, when evaluating a technology field by the number of citations alone, the younger patents i.e. patents published more recently, have a smaller chance of being cited than the older ones, so this type of citation analysis may not be useful for younger technology fields [112]. Moreover, certain other drawbacks of citation-based analysis have been identified. First, citation analysis only indicates individual links between two particular patents, making it difficult to understand the overall relationship among all the patents. Second, because citation analysis considers only citing/cited information, the scope of analysis and the depth of potential information are limited. Finally, it cannot consider the internal relationship between patents, taking only the existence or frequency of citations into account [102].

Identifying technology trajectories

A large volume of research studying patent citations focuses on the flow of knowledge and the creation of technology trajectories, i.e. the branches in the evolution of a technology. Duguet et al. [85] demonstrate the association of patent citations with the knowledge flow between new technologies. Fontana et al. [113] analyze the structure of connectivity of patent citation networks with the goal of reconstructing the main technological trajectories in a studied technology field. Huang et al. [114] combine tech mining, used to capture crucial atributes of technology, with the results of a patent citation network analysis to study the evolutionary trajectory of a new and emerging technology. Kim et al. [94] investigate the structure of the evolution of knowledge spillovers across technological domains measured by patent citations and explore the capability of link prediction for forecasting the evolution of cross-domain patent networks. You et al. [115] propose a forecasting method for technology development based on a two-level model of knowledge transfer among patents and patent subclasses, based on a two-level patent citation network. Patent subclasses with a higher development potential are identified, and the correlation between technology development opportunity and the topological structure of the patent citation network is discussed.

Identifying technology convergence

Other researchers focus more on exploring technology convergence, combining two or more elements of technology in order to create a novel system with radically improved functions. Kim et al. [116] attempt to understand the components of the convergence of technology by performing an analysis of citation links between patents in a technology network, using patent codes in order to identify key technologies which played a crucial role among the groups of convergence technologies and predict potential future technology convergence. Ko et al. [112] present a procedural method to analyze trends of industry-wide technology convergences based on knowledge flows of patents. A technological knowledge flow matrix is constructed, representing knowledge flow among technology classes. This is then extended to an industry wide knowledge flow matrix exploring the concordance between technology classes and industrial sectors. A visual map is created showing trends of industrial technology convergence. Park et al. [14] propose a quantitative future-oriented approach to technological opportunity discovery for technology convergence based on predicting potential technological knowledge flows between heterogeneous fields. These technological knowledge flows are predicted by a link prediction method applied to a directed network, with the network representing patent – citation relationships.

Network analysis

Lee et al. [117] use network properties (degree centrality, betweenness centrality and closeness centrality) to model global technology evolution. Moreover, a distance-based patent citation map is constructed by calculating relative distances and positions of patents in the patent citation network. Chen et al. [118] track the growth of communities detected within patent networks and evaluate which network properties predict the long-term growth of the communities. A network is constructed from patents belonging to a technology field, and a static clustering algorithm is applied to snapshots at certain time intervals. A dynamic tracking algorithm is applied in order to link discovered communities between successive years. A correlation between proposed metrics and growth rate is then checked. Takano et al. [119] propose a method for analysing technology fields by generating a cluster solution that includes unconnected and connected components of a direct citation network. They additionally analyse the change of these clusters by adding unconnected patents to the citation network. Ha et al. [120] use patent citations to identify core patents, patents that are especially influential in a technology field, combining it with data mining to capture keywords. Chang et al. [121] create a patent citation matrix and establish an indicator to calculate the lineal linkage between two patents with the goal of determining core patents (here referred to as basic patents) in a technology field. The hierarchical cluster analysis is used to classify these basic patents. Erdi et al. [86] also view the network of patents connected by citations as an evolving graph representing the innovation process. They identify clusters of patents representing technological branches and give predictions about the temporal changes of the structure of these clusters. Hung et al. [122] examine a patent citation network of radio frequency identification patents finding that a patent citation network can be considered a “small world” network, characterised by a high degree of clustering and short path length between any two nodes. Moreover, they find the patent citation network follows a power-law connectivity distribution.

Other patent citation based methods

In a more traditional context of studying technology evolution, researchers use patent attributes as a basis for determining the life cycle stage of a technology. Fallah et al. [84] use the cumulative number of forward citations to measure the rate of technological innovation, suggesting a modified S – curve to model the technologies growth. Altuntas [56] use patent citation analysis for prioritizing the portfolio of investment projects. Patent data is used to construct an average initial direct-relation matrix, using citations between technology classes and assuming that the number of citations shows the degree of influence between the technologies.

Marco et al. [123] use hazard estimation as means to separate patent quality from citation inflation, an attribute of patents where younger patents receive fewer citations than older ones. Kim et al. [94] propose a forecasting methodology for multi-technology convergence based on a patent-citation analysis, a dependency-structure matrix, and a neural network analysis.

Table 4 provides an overview of the papers presented in this chapter covering the methods for exploring and analysing technology based on patent citations. The first column lists the author of the paper, the second lists the method used in the paper and the third lists the results of the method used.

Table 4 Overview and comparison of patent citation methods

Author	Method	Research aim
Naoki Shibata, Yuya Kajikawa, Yoshiyuki Takeda, 2009 [90]	Comparing three types of citation networks	Detect emerging research front
Duguet & MacGarvie, 2007 [85]	Comparing firms' citations with survey results	Validity of interpreting citations as a measure of knowledge flow
Fontana, Nuvolari, & Verspagen, 2009 [113]	Identifying key inventions; Analyze structure of patent citation network Reconstruct main technological trajectories	Identify pattern of technical change
Huang, Zhu, Guo, Porter, & Zhu, 2014 [114]	Finding main paths of patent citation network	Identification of technology trajectories
E. Kim, Cho, & Kim, 2014 [116]	Network analysis of technology components Identify key technologies Betweenness centrality	Understanding characteristics of technology convergence
Ko et al., 2014 [112]	Technological knowledge flow matrix	Trends of industry-wide technology fusion

Park & Yoon, 2018 [14]	Link prediction in a directed network	Discovering opportunities for convergence
J. Kim & Lee, 2017	Text mining Association rule mining Link prediction	Identifying areas for concentric diversification
You et al., 2017 [115]	Two level citation network Time series modelling	Forecasting technology development trends
P. C. Lee, Su, & Wu, 2010 [117]	Network properties (degree Centrality, Betweenness Centrality, Closeness Centrality) Distance-based patent citation map	Global technology evolution
R. Chen, Park, & Smith, 2014 [118]	Community detection Tracking community growth over time Evaluating network properties	Predict long – term growth of communities
Takano, Mejia, & Kajikawa, 2016 [119]	Unconnected component inclusion technique	Increased understanding of technology landscape
Ha, Liu, Cho, & Kim, 2015 [120]	Patent citation analysis; Identifying core patents; Data mining keywords	Extracting core technologies
Chang, Lai, & Chang, 2009 [121]	Relationship of patents withing patent citation matrix	Establishing indicators for finding basic patents; Classification of basic patents
Fallah et al., 2009 [84]	Forward citations analysis Growth curve	Rate of technological innovation
Altuntas & Dereli, 2015 [56]	DEMATEL method	Portfolio prioritisation of investment projects
Marco, 2007 [123]	Parametric and non-parametric hazard estimation	Exploration of heterogeneity in the rate of patent citation
J. Kim & Lee, 2017 [94]	Patent – citation analysis; Dependency-structure matrix; Neural – network analysis;	Forecasting multi-technology convergence.
Erdi et al. 2012, [86]	Network clustering	Identification of patent clusters; Predicting changes to these clusters
Huang et al. [122]	Network analysis	Small world characterisation of citation network; Power-law connectivity distribution

3.2. Patent Keyword/Text and Classification Code Based Methods

A potential drawback of citation-based methods is that while they provide insight into the relationships between patents and the resulting knowledge flows, they provide little insight into the contents of patents, i.e. the knowledge contained within patents. A key distinction of the methods covered in this chapter is that they use morphological patent analysis to analyze the content of patents from free text by extracting keywords. So, while citation-based methods take a macro look at a technology field, since they are able to identify important and interrelated patents, keyword analysis is a micro approach, able to understand individual patents. Since the original patent documents are expressed in a natural language format, they have to be transformed into structured data. The process of keyword extraction is applied to identify keywords [102]. These keywords can then be used to construct a keyword network and gain a different perspective on a technology field than that provided by a citation network [124].

Kim et al. [125] used text mining to create an integrated patent – product database from US patents. Association rule mining is used to construct a product ecology network, and link prediction is used to identify potential areas for concentric diversification. Lee et al. [126] also apply the association rule and link prediction to IPCs related to triadic patents with the goal of predicting technology convergence. Future convergence patterns are predicted by applying a link prediction method to the IPC co-occurrence network. Feng et al. [127] combine text mining and morphological analysis in order to gain insight into both current and future patent technology. The morphological structure is created using a keyword selection method combining high-frequency keywords, amendatory mutual information measurement and artificial selection. Altuntas et al. [40] use IPC codes as proxies for examining technology scope, using the total number of different IPC codes as a measure of a technologies expansion potential and patent power where higher patent power leads to a higher spillover of technology between different fields.

A specialized form of keyword base patent analysis are patent maps. Patent maps are “patent information collected for a specific purpose of use, and assembled, analyzed and depicted in a visual form of presentation such as a chart, graph or table” [128]. A patent map has the following features:

- It is based on patent information
- It has a clear purpose of use
- It consists of appropriate patent information for the purpose of use
- It contains organized patent information
- It presents information visually

Trappey et al. [129] combine patent content clustering with technology life cycle forecasting to evaluate possible market opportunities. Key phrases are extracted from which a key phrases correlation matrix is derived. This matrix is used as the input for the clustering algorithm (K –

means). Lee et al. [130] use principle component analysis to reduce the number of keywords extracted from patents and create a two–dimensional patent map. From the map, patent vacuums (areas in which patents have not been granted) are identified. Yoon et al. [131] also identify patent vacuums as well as patent hot spots (areas in which patents actively appear) by constructing dynamic patent maps, based on subject-action-object (SAO) structures extracted using NLP from patent text, showing technological competition trends. Jeong et al. [132] propose identifying patent vacuums using a method for identifying essential patents (patents indispensable in producing a product) by using a patent map based on generative topographic mapping (GTM). Son et al. [133] also propose a generative topographic map (GTM) based patent map for identifying patent vacuums.

Joung et al. [134] use text mining tools to identify technical keywords and construct a technical keyword–context matrix. Hierarchical clustering is then applied, enabling the monitoring of emerging technologies by identifying clusters of technical keywords. Suh et al. [135] evaluate emerging technologies for services by proposing a keyword-based three-dimensional patent map. Five values of the keyword are calculated and combined in order to get the priority value of an emerging technology. Chen et al. [136] divide a technology field into communities and track the evolving trajectories of these communities through a visualization where each community is drawn as a function of its size, average age and time. From these trajectories, the structure of a technology can be investigated as well as emerging subjects. Kim et al. [137] cluster patent documents based on keywords collected from the patents from a technology field. A semantic network of keywords is then constructed based on the clustering results and a patent map is created by rearranging each keyword node of the semantic network according to its earliest filing date and frequency in patent documents. This enables the understanding of advances of emerging technologies and forecasting future trends. Yoon et al. [138] propose a self – organizing feature map based patent map that visualizes the complex relationships among patents and the pattern of technological advancement. They propose three types of patent maps (vacuum map, claim point map and portfolio maps) with the goal of monitoring technological change, developing new products and managing intellectual property. Huang et al. [139] apply co – classification analysis of patents to reveal the technical evolution process of a technical field, extract patterns and topics using co-word analysis and employ main path analysis to discover significant clues about technology hotspots. Zhang et al. [140] propose a unified framework, integrating different types of patent information, to generate a technology evolution tree for a given topic or patent classification code. This generated tree allows a variety of patent-related analyses, such as identifying prior art and detecting technology gaps. Wu et al. [141] use a self–organizing map to cluster patents into different quality groups based on previously defined quality indicators. Kernel principal analysis and the support vector machine are then used to enhance the patent quality classification model. Kim et al. [142] generate a patent development map by identifying the technological taxonomies of patents and visualizing the development paths among patents through sensuosity analyses based on semantic similarities. Yang et al. [143] use an NLP parser to extract conceptual graphs from patent claims using anchored relaxation labelling. In a later work [144], they use text mining (finite state machines, part-of speech tags) to convert a patent claim into a formally defined conceptual graph. The purpose of clustering is to segment elements into groups where elements inside a group are more similar to elements in the same group then elements within other groups. Yoon et al [102] extract keywords from patents and create an incidence matrix. The incidence matrix is used to generate a patent network. A series of quantitative measures is then applied on this network with the goal if identifying influential patens. Tseng et al. [145] describe a series of text mining techniques for the creation of patents maps, proposing a novel approach for verifying the success of information extracted via text mining. Daim et al. [146] integrate bibliometrics and text analysis with other forecasting tools (scenario planning, growth curves and analogies) with the goal of forecasting emerging technologies.

Table 5 provides an overview of the research presented in this chapter covering the methods for exploring and analysing technology based on text mining. The first column lists the author of the paper, the second lists the method used in the research and the third lists the results of the method used.

Table 5 Overview of text mining methods

Author	Method	Research Aim
Kim et al. 2017 [125]	Association rule mining; Link prediction	Identification of areas for concentric diversification
Lee et al. 2015 [126]	Association rule; Link prediction	Predict pattern of technology convergence
Feng et al. 2012 [127]	Morphological analysis	Insight in patented technology
Altuntas et al.	IPC codes	Patent power
Trappey et al. 2011 [129]	Patent content clustering; Technology life cycle forecasting	Forecast possible market opportunities
Lee et al. 2009 [130]	Text mining; Principal component analysis	Identifying vacuum from patent map
Yoon et al. 2013 [131]	Subject – action – object based content analysis	Identifying vacuums; Identifying hot spots
Jeong et al. 2013 [132]	Generative topographic mapping	Find candidates for essential patents
Son et al. 2011 [133]	Generative topographic mapping-based patent map	Identify vacuums
Joung et al. 2017 [134]	Technical keyword – context matrix; Hierarchical clustering;	Monitoring emerging technologies
Suh et al. 2209 [135]	Topic map of keywords; Clustering	Service – oriented technology roadmap
Chen et al. 2012 [136]	Network snapshots; Girvan – Newman Clustering	Visualisation of community evolution trajectories
Kim et al. 2008 [137]	K-means clustering, Semantic network of keywords	Understanding advances of emerging technologies
Yoon et al. 2002 [138]	Principle component analysis; Self-organizing feature map	Monitoring technological change; Developing new products; Managing intellectual property;
Huang et al. 2017 [139]	CO-classification analysis	Technical evolution process; Emerging and key technical fields Topical analysis
Zhang et al. 2014 [140]	Steiner tree	Technology evolution tree for a topic or classification code
Wu et al. 2016 [141]	Self-organizing maps; Kernel principal component analysis; Support vector machine	Forecasting patent quality
Kim et al. 2016 [142]	Semantic patent topic analysis	Patent development map generation
Yang et al. 2008 [143]	Anchored relaxation labelling	Conceptual graph
Yang et al. 2012 [144]	Finite state machines; Part- of-speech tags; Dependency tree	Conceptual graph
Yoon et al. 2004 [102]	Network analysis	Identifying trends; Identifying venues for product development;
Tseng et al. 2007[145]	Text mining	Create patent maps for topic analysis
Daim, Rueda, Martin, & Gerdsri, 2006 [146]	Bibliometric analysis Growth curves System Dynamic	Forecasting emerging technology;

3.3. Research Gap and Research Questions

Based on the overview of the theoretical background presented in chapter two and the overview of research presented in this chapter, research gaps are identified with the goal of defining the direction of this research and facilitate the illumination of research questions to be answered during the course of this research. Because this research uses patent as the primary source of data i.e. uses patent as proxies for technical invention, the exploration of research gaps focused on the drawbacks of patent-based methods for describing the life cycle stages of a technology and patent-based methods for exploring patterns in knowledge flow within a technology domain during its life cycle. Consequently, in the overview of research, priority was given to models which use patents as a primary data source and study technology change. The identified gaps are presented in the following text and, based on the identified gaps, research questions that will be answered by the end of this thesis are defined.

While it is demonstrated that there is a significant volume of research exploring the use of patents to study technology change, most of this research, in a technology evolution context, focuses on studying technology trajectories and convergence [113][114][116][112] However, there is limited research combining insights from research based on other resources (i.e. paper citations) and applied to patent networks. As patents are presently considered the most reliable structured records on inventive activity, this lack of research is considered a noticeable gap as studying patent citations as technological indicators might provide insight into the development of a technology domain. The review of the literature shows that the majority of patent analysis methods focus on exploring technology development trajectories by examining the direct patent citations. While this approach provides insight into the generational flow of knowledge, it provides little insight into how existing patents might combine and co-contribute to a future patent in the form of co-citations.

This lack of diversity in patent-based approaches can further be observed when exploring how patents are used in exploring the life cycle stages of a technology domain. It is noticeable that the vast majority of adopted patent-based methods for technology life cycle analysis base themselves on models derivative of the basic S – Curve model, i.e. the cumulative value of some patent attribute value [84][147][24][40]. While these methods provide some insight into the life cycle stages of the examined technology, they provide little understanding of the underlying dynamics of patent attributes and how they correlate to the life cycle phase.

Finally, a review of literature related to forecasting the development of technology identifies a wide array of quantitative and qualitative approaches. Narrowing the focus on patent-based quantitative methods for exploring the future development of technology, it can be seen that the majority of these methods are based on expanding the same S – Curves used to describe the life cycle of a technology domain and extrapolating future trends. While other quantitative methods exist, primarily causal models covered in Chapter 2.3.4, these do not use patent data as a primary data source. Moreover, they give limited insight into future knowledge flows. Consequently, this lack of methods that use patents, especially patent co-citations, to provide insight into future knowledge flows is identified as a gap in the field.

Based on these identified gaps, the following research questions are defined:

RQ 1 Can the dynamics of patent citation creation be used to determine the life cycle stages of a technology domain?

RQ 2 Can examining the occurrence of patent co-citations provide insight into patterns of knowledge flow within a technology domain?

As this research uses link prediction to predict the occurrence of patent co-citations, the following additional research questions are defined:

RQ 3 Can examining the occurrence of patent co-citations be used to identify which parts of a technology's life cycle contribute the most to future inventions?

RQ 4 When are the predicted co - citations, representing knowledge flow, created?

The answers to these research questions will be used to assess the validity of the hypothesis outlined in Chapter 1.1. More specifically, research question 1 contributes to answering the first part of the thesis hypothesis, namely that it is possible to model the dynamics of a technology domains development based on the existing record of technical invention. Research questions 2, 3 and 4 contribute to answering the second part of the thesis hypothesis, namely that insight can be gained into the potential future directions of technology development based on the existing records of technical inventions.

4. THEORETICAL FRAMEWORK FOR EXPLORING THE EVOLUTION OF TECHNICAL INVENTION WITHIN A TECHNOLOGY DOMAIN

This chapter aims to present the theories, presumptions and concepts this thesis uses to explore the problems presented in Chapter 1 and answer the research questions stated in Chapter 3. Insights from the literature have been synthesized into a unified framework, combining knowledge from different fields. The choice of the particular theories used is elaborated, as is their relation to the gap in the literature and their relation to the stated aims and hypothesis

As has been demonstrated in Chapter 2 and Chapter 3, covering the review of literature, multiple theories and concepts exist related to the evolution of technology as well approaches attempting to explore the future innovation-driven development of technology. Because of the extensive breadth of approaches describing the evolution of technology, it is essential to contextualize the presented research within the confines of a single theoretical framework as a springboard for further research. Consequently, this chapter aims to present the most relevant theory from each field studied in this thesis, namely those studying patents and technology, studying the evolution of technology, and those studying technology forecasting. The outline and synthesis of the theories on which this research is based allows the introduction of certain presumptions that represent these theories condensed to their simplest expression. These presumptions, stated at the end of each subchapter, represent the core concepts that this research builds upon, using them to help achieve the research aims and solve research questions.

To recap the space where this research sits, as well as the aims of the research, this research focuses on exploring how knowledge flows within a technology domain relate to the evolutionary stages of a technology domain using formalized records of technical invention as proxies for technology. This research aims to create a methodology for conducting an analysis of a technology's life cycle in a novel way, using formalized records of inventions. Moreover, based on the results of this analysis, an additional aim is to gain insight into previously undiscovered patterns governing the evolution of technology. These patterns can be used for exploring the intuition governing both mature and disruptive technologies and comparing the findings, using these findings to predict the future evolution of a technology domain. In order to accomplish this, the theories related to the fields of study this research draws upon are defined in Chapters 4.1-4.3. The first subsection, 4.1, presents the theoretical arguments for choosing patents as proxies for technologies to be used in this research, making comparison between patents and alternative data sources that might be used as proxies for technologies. The arguments for choosing patents are presented, as are the advantages and disadvantages of using patents as the primary data source. The second subsection, 4.2, presents the theory of technology evolution, which will be used and expanded on in this research. Some key features of this theory are presented, as well as some key existing research. The drawbacks of the existing research are presented and the contribution to the theory is stated. This is followed by the third subsection (Chapter 4.3), dealing with technology forecasting, which outlines which attribute of a technology's evolution will be forecasted and presents the theoretical context of forecasting this attribute. Finally, the fourth subsection (Chapter 4.4) outlines the integration of the three propositions into a single unified theoretical framework, which the presented research will be based on.

4.1. Patents and Technology

This research uses patents as the primary source of data. More specifically, patent applications are compiled into datasets and specific metadata is then extracted from the patent applications to be used for further analysis. The reason patents were primarily considered as the primary source of data is because they are created from a technology/financial perspective rather than a general public/end-user/consumer perspective [44], making them suitable to be used in a technology management context. Nevertheless, several other data sources were considered as alternative proxies of technology, as a supplement to patent applications, namely scientific publications and web search queries. A comparison of these three sources is shown in Table 6.

Table 6 Comparison of patents and alternative data sources

Characteristic	Patents	Scientific Publications	Web
Availability	High	Medium	High
Economic barrier	Low	High	Low
Language barrier	Low	Low	Medium
Structured	High	Low	Low
Level of knowledge	Expert level	Expert level	Highly variable
Publish requirements	High	High	Low
Document detail	High	Medium	Low
Knowledge maturity	Medium	Very young	Unreliable

Seven characteristics of sources were considered when determining the one most suitable for our research, based on the research of Albert [148]. The primary characteristic taken into account was openness; whether a barrier existed limiting the availability of the data. As one of the aims of this research was to make it as open as possible, a high barrier to acquiring data was considered a drawback. Three factors influence openness:

- Availability,
- Economic barrier,
- Language Barrier.

Availability explores whether an accessible repository of data exists, which enables for a relatively simple way of searching, filtering and sorting of data as well as its retrieval. The Economic barrier characteristic explores whether the repository is free to use or behind a paywall. The Language barrier explores whether a large part of the data is accessible in a globally common language (in our case, the English language).

Another important characteristic was the structuredness of the data source. The structuredness of a source can range from low (continuous data without structure) to very high (clearly defined content; extensive metadata). Because of the large amount of raw data being analyzed, a highly structured data source is preferable to a lower structured one as a highly structured data source allows for more straightforward automation of a large dataset analysis. "Level of knowledge" clarifies the level of knowledge the document contains and is directly linked to the level of knowledge the author of the document needs in order to publish it. This creates a condition that disqualifies potential data sources with a high number of noisy data, i.e., low-quality data. Publishing requirements specify the level of institutionalised gatekeeping preventing the publishing of the data source. Document detail is ranked by how detailed an explanation of the technology a document contains. Knowledge maturity is ranked by the approximate maturity of the technology contained in the document.

Of the three data sources considered, web search queries were disqualified almost immediately. The extremely low publishing requirements means there is almost no institutional quality control of the available data, meaning the level of noise in any created dataset would be very high. Moreover, the unstructured nature of this type of data makes any large-scale dataset creation difficult.

Scientific literature showed to be a more promising source of data, however several drawbacks were identified. Primarily, a large volume of research is not publicly accessible but exist behind a paywall, requiring special access to the publishers' database, reducing the accessibility of the data source. Second, while scientific literature has some structure, it can differ significantly depending on the publisher. Moreover, a piece of scientific literature mostly consists of free text with limited metadata increasing the difficulty of extracting information. Finally, the maturity of technology being described is very young, often at the theoretical or early conceptual phases. All of these reasons disqualified scientific literature as a viable data source in this research.

Finally, the use of patent applications as the sole source of data was confirmed. Patent applications have significant advantages making them a viable data source for this research. First, they are publicly available, meaning the level of accessibility is very high. While specialized services exist, which offer advanced tools for manipulating patent databases, raw patent applications are open and easily available. Second, patent applications are highly structured documents with clearly marked subsections and a uniform structure. Therefore, it is reasonably easy to extract metadata from patents. Since it is general practice to apply patents in multiple jurisdictions, it is a reasonable assumption that the vast majority of inventions are patented in the USPTO, meaning they are available in the English language. Based on these reasons, patens are chosen as the primary source of data in this research.

As has been mentioned in the review of the literature (Chapters 2 and 3), patents have a history of being used as proxies for technology and knowledge [88] [30][149][2][117][150]. The underlying assumption behind this use is that most research and development activities generate inventions, which are then protected by innovators as their intellectual property in the form of a patent. Consequently, patents can be considered proxies of invention [88]. According to the work of Yoon et al. [18] patents can be considered as containers of knowledge elements, while Phelps et al. [89] call them "discrete knowledge artefacts", alongside other knowledge artefact papers and products.

Patent citations are the primary metadata contained in patent applications that are used in this research. To reiterate, the citation information of patents capture the directed relationships between patents and define technological antecedents and descendants [3]. At its foundation, citation-based analysis presumes a view of the process of technological development that is cumulative, where each inventor benefits from the work of previous inventors and subsequently influences the work of future inventors [3]. These directional relationships can show the flow of information from patent to patent, demonstrating the spillover of technology through clearly separated entities. In the context of knowledge management, a patent citing another patent implies that the cited patent contains a piece of knowledge that the citing patent builds upon [86]. The types of citations being analyzed differ among researchers, the most common being direct citations, co-citations and bibliographic coupling [49]. Direct citations are defined as an antecedent –descendent link. Co-citations are defined as linking patents being cited by the same patent. Bibliographic couplings are defined as linking patents that cited the same paper.

Several drawbacks of using patents as a data source should be noted. First, there is a delay between the time a new patent is filed and the time a patent is granted due to the process of examining a patent application. This delay is between 12 and 18 months which is considered high by some researchers [44]. Second, patents are notoriously difficult to create datasets from, as individual patent applications can be intentionally worded in an unclear way in order to reduce their discoverability. However, with the increase of the number of patents contained in a dataset, the omission of a small number of patents does not influence the quality of the datasets in a significant way. Nevertheless, a tested methodology for retrieving patents related to a technology domain should be used when creating datasets. Finally, not all inventions are patented. However, as these inventions are not publicly disclosed, their influence on the evolution of technology is limited. Moreover, as was argued when discussing incomplete datasets, the number of inventions not patented is small enough that it should not influence the quality of a dataset in a meaningful way.

To summarize, patents are structured and formalized records of invention, which are free to access and have a history of being used as proxies for technologies. This makes them in line with the aims and objectives stated in Chapter 1 and confirms their potential to contribute to answering the research question stated in Chapter 3.3. Their highly uniform structure makes them the most suitable record for conducting analysis on a large scale.

Presumption 1:

Patents are records of technical invention and are a viable proxy for technology. Consequently, they can be used to explore the evolution of a technology domain.

4.2. Technology evolution

Considering the different approaches to describing the evolution of a technology and the mechanisms governing it, as demonstrated in the review of the literature (Chapter 2.1), it is crucial to define the theory and approaches to exploring a technology's evolution on which this research builds upon. In doing so, the most prominent theories and theoretical models will be compared and then contextualized within the scope of this research. It should be noted that these theoretical models do not specify indicators and instead resort to latent exogenous variables. Concepts from the chosen theory used in this research are identified and presented, contextualizing the presented research within an existing concept used to study the evolution of technology. This provides a measuring model of latent exogenous variables enabling a quantitative insight analysis of a technology domains evolution.

Table 7 provides a summary of the key concepts from the field of technology evolution as well as their key characteristics, providing a summary of the theories and concepts outlined in Chapter 2.1. The text following the table presents the synthesis of these discrete theories and their implementation into the theoretical framework.

Table 7 Key concepts of technology evolution

Concept	Characteristics
Incremental evolution	Gradual accumulation of small variations over time that yield a novel innovation [4]
Radical evolution	Radical departure from the norm of incremental improvement [46] [29]
Disruptive	Significantly alters the way that consumers, industries of businesses operate [9]

A difficulty arises when attempting to adopt a particular theory of technology evolution as there is not a clear demarcation separating different theories. This is demonstrated by the fact that while certain traditional theories contain discrete concepts, the majority of the state of the art approaches to modelling the evolution of technology combine the concepts presented in Table 7. For example, while a researcher will claim that technology evolution is an exclusively incremental linear process, they will concede that this linearity is interrupted by sudden development surges [22]. However, they will not call these surges “radical” or “disruptive”, merely an increase of the speed of linear evolution.

Nevertheless, the research presented in this thesis subscribes to the theory of technology evolution most accepted at the present time, which states that a technologies evolution consists of a series of incremental improvements occasionally marked by radical innovations [8][55][27][60]. Moreover, disruptive innovations are rare but possible and a separate phenomenon from radical innovations, so the exploration of disruptive technologies should be incorporated into the research.

A concept familiar to most theories of technology evolution, and one widely accepted by researchers, is the S – curve based maturity assessment model, which is used to model the life cycle of a technology, segmenting it into four stages [24][23][151]. The S – curve model has a history of being used for exploring the evolution of technology and is based on the insight that some aspects of a technologies life cycle follow an S-shaped pattern [24]. The technology S-curve is a visualization showing the progress of technology performance regarding some transient metric, whether the R&D – effort and expenses [63] or simply as a function of time [151]. The life cycle of a technology is commonly segmented into four stages: emergence, growth, maturation and stagnation. The emergence and growth stages are often grouped together as a single stage, especially when the research is focused on the later stages of a technologies life cycle [40][152]. In either case, these stages are artificial periods, transient in nature, introduced in order to ease the interpretation of different states of technology maturity and take appropriate action. The boundaries between these stages are not clear and often a certain amount of overlap exists. However, they are determined with a significant deal of accuracy by simple grouping periods of a technology’s life cycle requiring the same action. Table 8 shows a summary of the four life cycle stages and the characteristics of each stage. Moreover, it was noticed during the review of literature that the nomenclature marking life cycle stages is not uniform, and often varied depending on the researches. Therefore, Table 8 also provides several common aliases to the life cycle stage names used in this literature.

Table 8 Overview of the life cycle stages of a technology domain

Life cycle stage	Alternative name	Characteristics
Emergence	Invention, Initiation	Pacing technology with low competitive impact and low integration in products [24][44]
Growth	Innovation	Pacing technologies turn into key technologies; integrated into products; maintain high competitive impact [24][44]
Maturation	Maturity, Saturation	Loss of competitive impact; base technology; might be replaced by new technology[24][44]
Stagnation	Decline, Saturation	Reduced utility of technology, imminent replacement likely[24][44]

Figure 6 shows a generalised illustration of an S – curve showing the performance of a technology over time. An overview of the attributes that can be used as performance indicators is presented in Table 9. The values on the abscissa are some measure of time, most commonly years [153]. The location of the four life cycle stages on the S – curve are marked, as described by previous research [24][151][153].

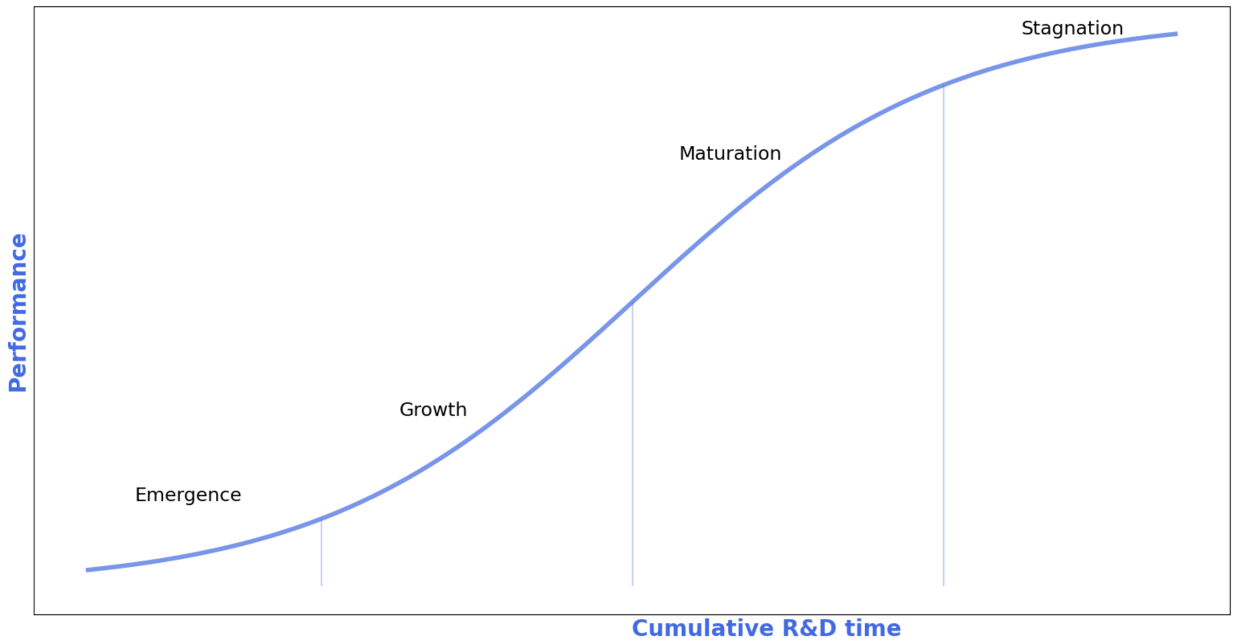

Figure 6 Generalised illustration of an S- Curve model of a TLC[24]

Table 9 presents an overview of some of the indicators used to create an S – Curve model of the life cycle of a technology domain based on patent data as presented by Gao et al. [24].

Table 9 Overview of patent based indicators [24]

No.	Indicator	Indication description
1	Application	Number of patents by application year
2	Priority	Number of patents by priority year
3	Corporate	Number of corporates by priority year
4	Non – corporate	Number of non – corporates by priority year
5	Inventor	Number of inventors by priority year
6	Literature citation	Number of backward citations to literature by priority year
7	Patent citation	Number of backward citations to patents by priority year
8	IPC	Number of IPCs (4 – digit) by priority year
9	IPC top 5	Number of patents of top 5 IPCs by priority year
10	IPC top 10	Number of patents of top 10 IPCs by priority year
11	MC	Number of Manual Codes (MCs) by priority year
12	MC top 5	Number of patents of top 5 MCs by priority year
13	MC top 10	Number of patents of top 10 MCs by priority year

Each of the indicators presented in Table 9 has a history of being used with varying levels of success to explore the TLC. The relatively large number of patent-based indicators provide certain advantages for researchers using patents to study the life cycle of a technology. There is an inherent modularity due to having the option to use multiple indicators, making this approach agnostic regarding the larger framework used. For example, a framework that is based on the number of patent applications by year can implement the corresponding S – curve model as a part of the framework. In the same way, a framework based on patent citation can implement an S -Curve model based on patent citations.

The use of the S- curve model of the technology life cycle in this research is suitable for a number of reasons. Primarily, the presented research is based on exploring the evolution of technology using patent citations. As patent citations have a history of being used to create S – Curve models in past research [24][40], as seen in Table 9, it is assumed that the existing research can be expanded within this research. Moreover, the cumulative nature of S- Curve based models makes them compatible with the theory of technological evolution, which views technology evolution as a series of incremental improvements marked by radical leaps. These discontinuities in development can often be identified by discontinuities within the S – Curve [154].

One of the benefits of using the S – Curve to model the life cycle of a technology is that it enables the identification of the implementation potential of a technology. As is shown in the research presented by Albert [44], the life cycle stages of a technology can be connected with its strategic relevance categories, namely Base Technologies, Key Technologies and Pacing Technologies. The strategic relevance of a technology for an industry is the importance of that technology compared to other technologies for commercialization in that industry [148].

The relationship between the life cycle stages identified on the S –curve and the strategic relevance of a technology is illustrated in Figure 7.

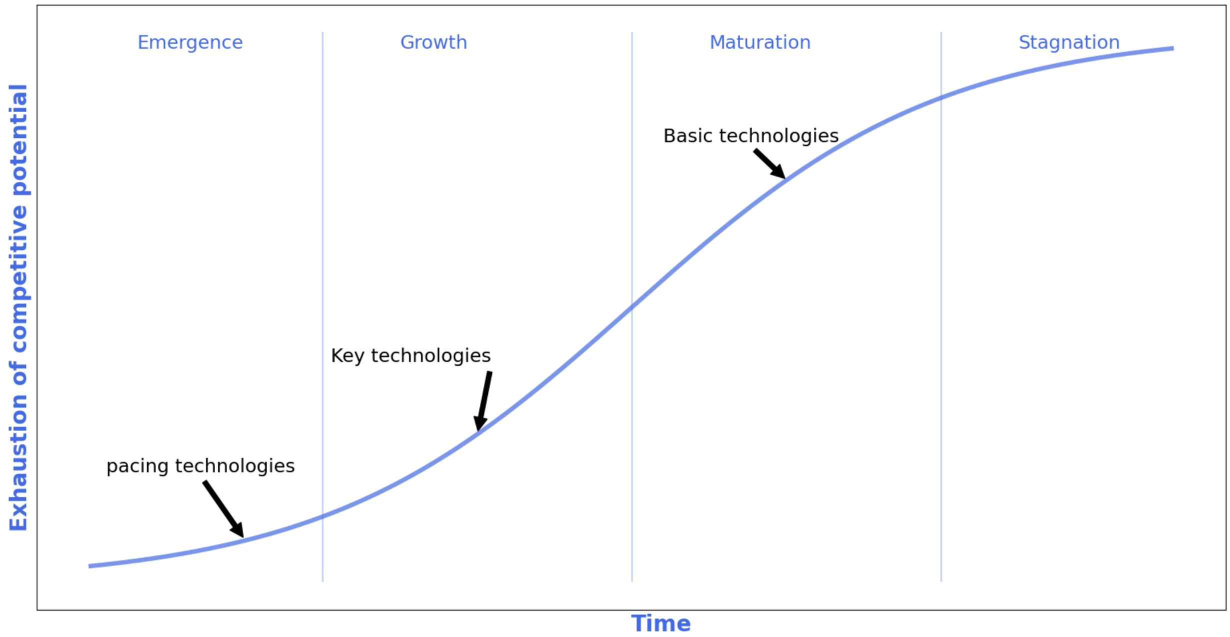

Figure 7 Relationship between the Technology Life Cycle Stages and Strategic Relevance [148]

Table 10 lists the categories of technology strategic relevance as well as the primary characteristics of each category.

Table 10 Technology strategic relevance categories and their definitions

Strategic Relevance Categories	Definition
Basic Technologies	*Technologies that a firm has to master in order to participate in an industry; the technology on which most products in an industry are based on* *Low strategic value*[155]
Key Technologies	*Have a major impact on the competitiveness of a firm; competitors have not yet mastered them* *Influence critical performance parameters* *High strategic value*[155]
Pacing Technologies	*Still being developed; early maturity state* *Good chance of becoming key technologies*[155] *High strategic value*

Identifying the strategic relevance of a technology is useful in an industrial context. By analyzing a technologies position on the S – curve, it is possible to determine if a technology is worth investing in, as the strategic relevance of a technology depends on the exhaustion of its competitive potential in a particular industry [148]. For example, a technology in the maturation or saturation stage of its life cycle is most likely a basic technology. As such, it is not worth investing in as it there are limited opportunities for their use to gain a competitive advantage [23]. Consequently, its strategic relevance is low and fewer resources should be allocated in its development [40]. However, a technology at the growth stage of its life cycle is potentially a key technology, meaning it has a higher chance of successful commercialisation. This makes it a technology of high strategic value and consequently makes it a viable target of resource allocation [23].

To conclude this subchapter, this thesis subscribes to the theory of technology evolution, which states that technology evolves in incremental steps with occasional radical improvements and disruptions. The concept of technologies following four stages during their life cycle is adopted and built upon, as is the concept that certain attributes of technologies follow an S-shaped curve during their life cycle. This is in line with the aims of this research, one of which is introducing new ways of exploring the life cycle stages of a technology and determining its implementation potential. This concept organically builds upon the concepts introduced in Chapter 4.1, dealing with patents as proxies for technologies, as the two have a history of working in synergy as a means for exploring a technologies evolution and determining its life cycle stages.

Some drawbacks of using the S – curve model should be noted. One often mentioned drawback is that the S-curve model is based on technology – specific performance indicators, as opposed to standardized variables (those which remain the same, independent of the subject of observation). This increases the difficulty of comparing technologies that are not directly connected or do not share the same measurable attributes. Using patent applications as the basis for S – curve models solves this problem, as patent applications are proven to be consistent with the metadata they contain. These metadata can be used as standardized variables, making it possible to develop technology-agnostic models.

Presumption 2:

Technologies follow a 4-stage life cycle, with each stage having certain predetermined characteristics. The technology life cycle can be approximated by an S-curve which can be used to identify the life cycle stages of the technology. Finally, there is a correlation between the life cycle stages of a technology with the strategic relevance of that technology. Consequently, identifying the life cycle stages of a technology can provide insight into the implementation potential of technologies.

4.3. Technology forecasting

In this research, one of the aims is to predict the potential future directions of the development of a technology. More precisely, a focus is placed on exploring how knowledge will flow within a technology domain, using patent citations as a measure of knowledge flow. We base our approach on the fact that patent citations are associated with the knowledge flow of new technologies [85].

The concept of using patent citations as a basis for exploring knowledge flow is based on the existing practice of researchers using citation analysis to study the flow of knowledge between different knowledge artefacts. While the citations between research papers were traditionally being explored [5][156][12][157], citations between patents have also been used to study knowledge diffusion [158][159][110][160]. However, while research exists exploring the use of patent citations as a way of studying changes in knowledge domains as well as research that uses patent citations to study technology trajectories and convergence, limited research exists which combines insights gained from examining paper citation with patent analysis. This is a noticeable gap as patents are, as previously stated, considered to be among the most reliable structured records on inventive activity, covering a wide range of fields of innovation. Therefore, studying patent citations as technological indicators might provide insights into the development of a technology domain. The majority of patent analysis methods covered in the review of literature focus on using direct patent citations to explore technology trajectories. While this approach does provide insight into the general flow of knowledge (i.e. the flow of knowledge from older to newer patents), it provides little insight into how existing patents might co – contribute to a future patent in the form of co-citations.

Three types of patent citations were considered for this research: direct citations, bibliographic coupling and co-citations. An overview of the three instances of patent citations is provided in Table 11 and the rationale for choosing a particular one for exploring the knowledge flow within a technology domain is presented in the text following the table.

Table 11 Overview of the three instances of patent citation [86][158][161]

Citation type	Positive	Negative
Direct citation	Shows knowledge flow between ascendant and descendant patent	It is not suitable for exploring the co-contribution of technology
Co-Citation	Shows co-contribution of knowledge	Provides limited direct knowledge flow over generations
Bibliographic coupling	Shows co-authorship of document	Author dependent as opposed to technology dependent

In this research, the focus is on the second way of representing patent citations, in the form of a patent co-citation network, when attempting to predict the future flow of knowledge within a technology domain, assuming these co-citations represent the flow of knowledge, i.e. co-contributions of existing knowledge to new innovations. It is assumed that, by predicting the occurrence of new links in this network, we predict existing knowledge's co-contribution to future inventions. The dynamics of a patent co-citation network are thus explored, studying whether patent co-citations can be predicted, as well as when these predicted links occur and which patents contribute the most to future co-citations. While existing research focuses on predicting the creation of new links in a patent citation network, it mainly focuses on exploring knowledge flow between different technology domains [15][126].

To conclude this chapter, in this research patent co-citations are used to represent the collective contribution of existing knowledge contained in patents, to future inventions. By successfully predicting links in a graph created from patent co-citation, the future flow of knowledge can be predicted.

Presumption 3:

Patent co-citations represent co-contributions of knowledge to new inventions. A network can be created from these co-citations where nodes represent patents and edges represent co-citations. This network represents the knowledge co-contribution within a technology domain. Moreover, by predicting the creation of new edges within this network, insight can be made into the future co-contribution of existing knowledge to future inventions.

4.4. Unified theoretical framework

Based on the theories and concepts outlined in this chapter, a unified theoretical framework can be constructed. The constructed framework is to be used as the basis for the research presented in this thesis. A positive attribute of the theories and concepts making up the theoretical framework is their high level of compatibility, consequently simplifying the creation of a joined framework. There is an organic synergy within the framework, starting with dataset creation which is built upon by the analysis of a technology domain and ends with forecasting. Based on the literature review and the contextualization of the existing research within the space of this research, several theoretical presumptions are made upon which this research builds. These presumptions form the base of the theoretical framework.

Figure 8 shows a visualization of the unified theoretical framework and the relationship of the component theories. Figure 8 consists of blocks representing different theories within the framework, sorting them by their contribution to the research methodology and illustrating their relation to the research questions and stage within the research (Data Gathering, Data Analysis I, Data Analysis II). Moreover, it illustrates their interdependence as flows of data and results**.** The unified theoretical framework is based on the notion that patents can be used as proxies for technologies (Presumption I). Consequently, a significant emphasis is placed on a method for the creation of patent datasets that accurately represent a chosen technology domain. Moreover, emphasis is placed on understanding the nature of patents and the information they contain and identifying the exact relevant patent metadata to be used as the basis for further research. The second part of the theoretical framework, building on the first part, focuses on analysis, exploring the evolution of technical invention within a technology domain. This stage draws upon the standard four stage technology life cycle concept often explored by applying an S – Curve model to some technology attribute (Presumption II). The presented research expands on this field by introducing a novel way of exploring the technologies life cycle by using a dynamic analysis of patent metadata to determine the life cycle stages of a technology. This stage of the theoretical framework contributes to answering research question 1. Finally, building on the previous two parts of the theoretical framework, an attempt is made to predict the future development of an aspect of a technology's evolution. To be more precise, this stage focuses on predicting future flows of knowledge within a technology domain based on patent metadata and draws upon existing research focusing on exploring knowledge flow by analysing citations (Presumption III). This stage of the theoretical framework uses the results of the dataset creation stage of the theoretical framework as well as the results of the analysis conducted in the second stage of the theoretical framework in order to contribute to answering research questions 2, 3 and 4.

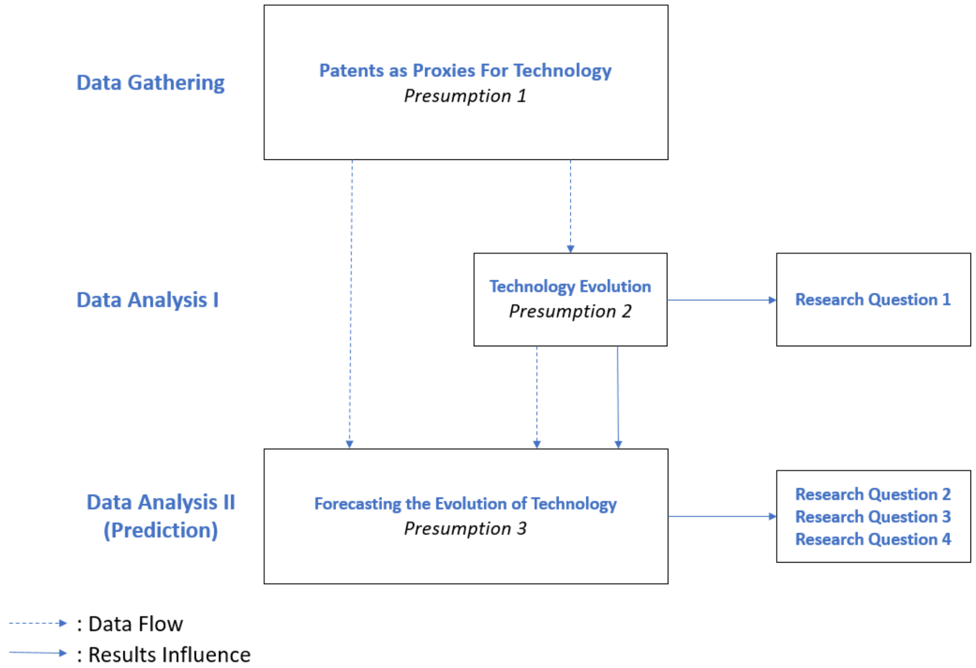

Figure 8 Visualization of the Unified Theoretical Framework

Combining the theories from the three fields of study, summarized as Presumption I, II and II, allows for creating a unified theoretical framework, laying the foundation which will support the analysis and interpretation of results and help make broader generalizations.

A few shortcomings of the proposed framework should be addressed. These are primarily manifested in the validation phase. In order to properly validate the framework, it should be applied to a variety of datasets representing numerous different types of technologies. However, that kind of validation would be out of scope of this thesis. A more detailed overview of the limitations of this research is presented in Chapter 9.1.

5. DESIGN OF THE EMPIRICAL PART OF RESEARCH

This chapter presents the design of the second descriptive study conducted to help answer the research questions and verify the hypothesis. An outline of the data collection and dataset creation methods is provided, as are the methods and tools used to analyse the collected data.

5.1. Introduction

The empirical research conducted in this research consists of 5 stages illustrated in Figure 9 and correlates with stage 5 of the research methodology outlined in Chapter 1.2, namely the second descriptive study. Each stage of the empirical research applies a presumption defined in the theoretical framework presented in Chapter 4.4 to a technology domain with the goal of answering the research questions defined in Chapter 3.3 and verifying the hypotheses outlined in Chapter 1.1.

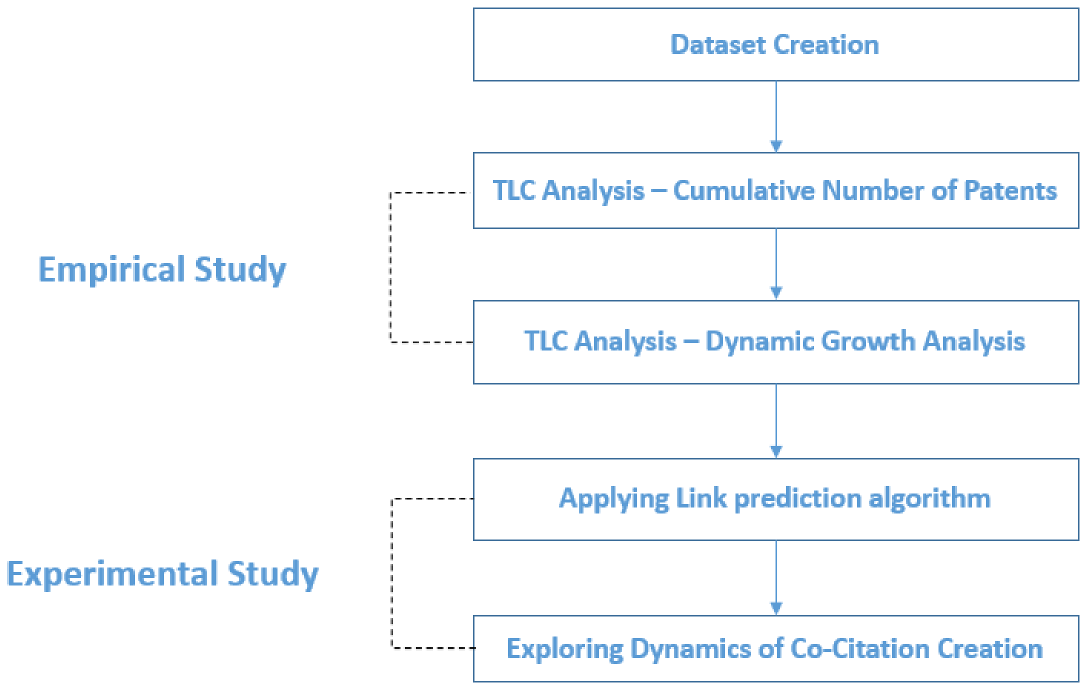

Figure 9 Overview of the second descriptive study stages

The descriptive study II, used to answer the thesis research questions and test the thesis hypothesis, consists of two parts, i.e. two studies. In order to synthesise the results of each study, the hierarchy-of-hypotheses (HOH) approach of synthesis is used, where the general hypothesis is mapped in relation to sub-hypotheses defined in each study [162], connecting them in a hierarchically nested fashion. The hierarchy-of-hypotheses is presented in Figure 10, illustrating the relationship between the two sub-studies, their respective sub-hypotheses, the thesis research questions and the central thesis hypothesis.

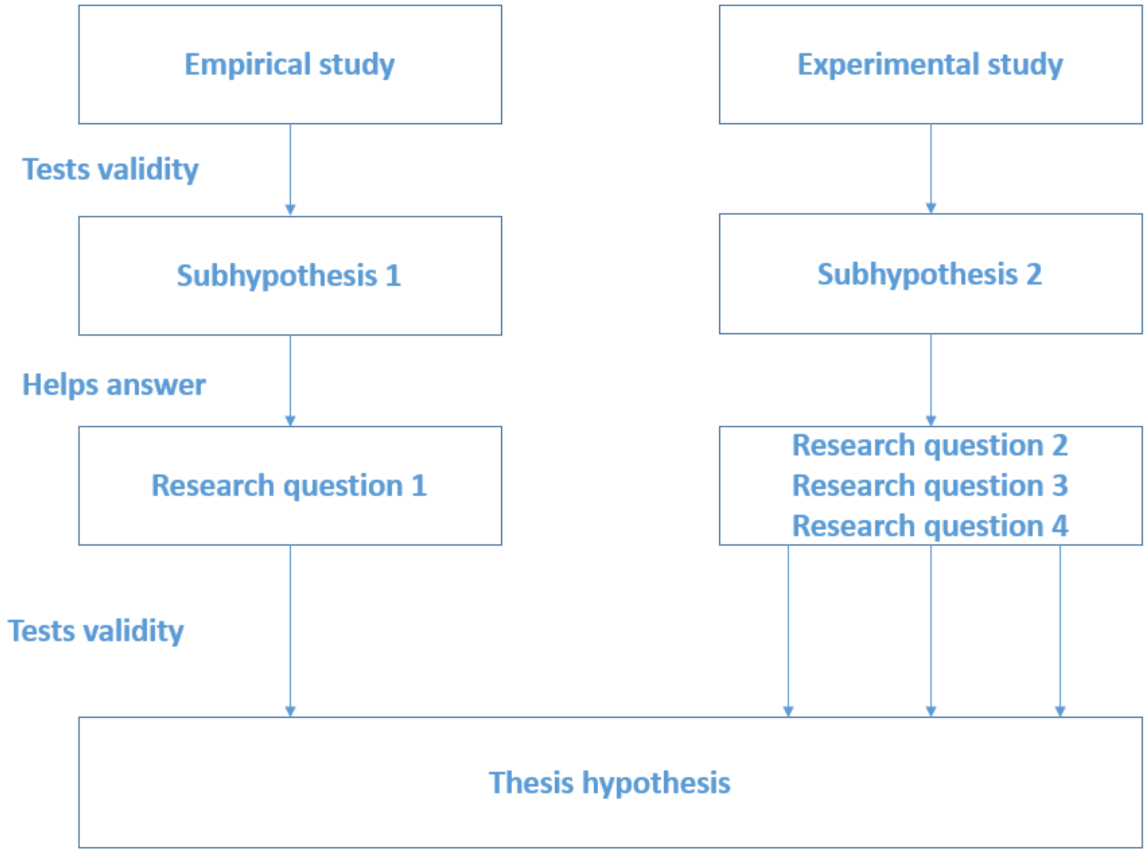

Figure 10 Hierarchy of hypotheses in this research (after Ryo et al. [162]*)*

Each of the two substudies explores a different phenomena related to the evolution of a technology domain, with the goal of validifying its specific subhyptothesis. The results of these studies help answer the research questions which in turn help verify the thesis hypothesis.

The first sub-study is a purely empirical study, tasked with observing and identifying the life cycle stages of a technology domain and testing the hypothesis that the life cycle stages of a technology domain can be identified by conducting a dynamic growth analysis of a patent citation network. This step attempts to answer research question 1 as outlined in Chapter 3.3 and contribute to testing the governing hypothesis of this thesis. Table 12 provides a summary of the phenomenon being studied in this part of the empirical study and states the hypothesis being tested.

Table 12 Subhypothesis being tested by empirical study

Phenomenon Explored	Subhypothesis
Technology Life Cycle Stages	The growth analysis of a patent citation network can be used to identify the life cycle stages of a technology domain

This empirical study observes the results of a dynamic growth analysis of a patent citation network superimposing the results to those of an established technology life cycle analysis methods proven to be accurate. A more detailed overview of this empirical study is presented in Chapter 5.3.

The second substudy of the empirical study consists of an experimental study testing the hypothesis that the dynamics of knowledge flow within a patent co-citation network can be modelled and predicted using an appropriate link prediction algorithm. This sub-study contributes to answering research questions 2, 3 and 4.

Table 13 Subhypothesis being tested by experimental study

Phenomenon Explored	Subhypothesis
Patent co-citations	A link prediction algorithm can be used to describe the intuition governing the growth of a patent co-citation network and predict future co-citations.

The experimental part of the study consists of two experiments, both based on the application of link prediction algorithms to a training dataset (The creation of these datasets is presented in Chapter 5.4). The first experiment sees the size of the training dataset constant while the applied link prediction algorithm is a variable. The goal of varying the applied link predication algorithm is to identify the link prediction algorithm best describing the underlying intuition behind the network growth. This part of the study attempts to answer research question 2. The second one sees the link prediction algorithm constant, while the training dataset changes. The goal of varying the training dataset is studying the dynamics of patent co-citation creation, exploring which training dataset contributes the most to the precision of the link predictions algorithm as well as when the predicted links occur. This part of the study attempts to answer research questions 3 and 4. In both cases, the measures of success used are the precision of the link prediction algorithm and the area under the receiver operator characteristic curve (AUC). An expended overview of both of these metrices is presented in Chapter 5.4.

Table 14 provides an overview of both experiments, highlighting what is constant and what is varied in each experiment, as well as the measures of success and aims of the experiment.

Table 14 Overview of Experiments

	Variable	Constant	Measure of success	Aim
Experiment 1	Link prediction algorithm	Size of the training dataset	The precision of results; AUC	Identification of most precise link prediction algorithm
Experiment 2	Size of the training dataset	Link prediction algorithm	The precision of results; AUC	Identification of the training dataset providing the most precise results

A more detailed plan of the experimental study is presented in Chapter 5.4.

As outlined in Figure 12, the empirical research starts with the creation of a dataset consisting of patents representing the technology domain being studied. A life cycle analysis of the technology domain is conducted by applying one of the existing methods from the literature presented in Chapter 3, followed by a method introduced in this research. The technology domain life cycle analysis is based on the part of the theoretical framework synthesizing knowledge from the Technology Evolution field. Finally, an attempt is made to forecast the future dynamics of knowledge flow within the chosen technology domain by choosing a link prediction algorithm representing the underlying intuition governing the evolution of the chosen aspect of technology and applying it to the created dataset.

Table 15 provides an overview of the design of the empirical research, emphasising how each stage of the empirical research builds upon the previously defined presumptions, as well as what research question each stage attempts to answer.

Table 15 Overview of the contribution of each stage

Stage	Presumption	Research Question
Dataset Creation	Presumption 1	/
Technology Domain Life Cycle Analysis – Empirical Study	Presumption 2	RQ 1
Patent co-citation analysis Experimental study –	Presumption 3	RQ 2 RQ 3 RQ4

5.2. Dataset creation

The first stage of the empirical research consists of creating a dataset comprised of patents representing the chosen technology domain. The dataset creation method used in this research is a modified version of the method devised by Benson and Magee [163]. Table 16 provides an overview of the steps for creating the dataset.

Table 16 Method for the Retrieval of Patents Related to the Chosen Technology Domain

Step	Detail of Step
Step 1. Choosing a technology domain	Identifying a technology domain; Determining whether if it is within set constraints
Step 2. Identifying keywords	Determining keywords based on expert insight; Review of written records describing the technology
Step 3. Initial search of patent database	Input of keywords; Set constraints: • Grouped by simple families • Sorted by date of application
Step 4. Identifying classification codes	Based on the review of written records; Based on analysis of patents retrieved in Step 3
Step 5. Classification code based filtering	Filtering out of patent classified using codes identified in Step 4
Step 6. Manual review	Manual review of the created dataset; Manual review of patents which were filtered out

The first step of the dataset creation methodology consists of choosing a technology domain appropriate for this research. In choosing an appropriate technology domain for the empirical research, specific requirements had to be met:

- The examined technology domain represents an engineering field;
- The technology must be from a field that has a practice of patenting inventions;
- The life cycle stage and the nature of the examined technology are known based on export knowledge or prior work.

The first requirement ensures that the examined technology domain represents an engineering field as this is the field of study of this thesis. The second requirement stems from the fact that not all fields have the practice of patenting inventions. Moreover, some fields deliberately do not patent their inventions, strategically protecting their core technology [76]. While the act of choosing the appropriate technology domain might seem trivial, both an under-constrained or over-constrained technology domain, i.e. a technology domain with a too large or too small number of patents, would not be suitable for the empirical research as the number of relevant patents would be too large, in the case of an under-constrained technology domain. Therefore, the relevant patents would be too diverse to provide any meaningful insight in the context of this research.

Similarly, an under-constrained technology domain would contain too few relevant patents making gaining any meaningful insight from the analysis impossible. Finally, the third requirement ensures that the examined technology domain's life cycle stage and nature are already known. This knowledge can come from expert knowledge in the form of white papers, reports or interviews with experts from the fields, or scientific papers and books exploring the technology domain. The importance of this is twofold. First, prior knowledge of the technologies technology life cycle stage allows for the easier contextualisation of the results of this research. Second, this research aims to study both mature and emerging technology domains, exploring the particularities of each type of technology and comparing the results. Consequently, prior knowledge of the type of technology is required in order to choose appropriate technology domains representing both types of technologies.

The second step of dataset creation consists of identifying relevant keywords which accurately represent the chosen technology domain. The process of choosing appropriate keywords varies in its complexity depending on the technology domain being studied. Certain technology domains can be described by a simple set of keywords. However, technology domains consisting of somewhat “niche” technologies require a degree of input from domain experts familiar with the intricacies of the studied technology. In either case, some prior domain knowledge is a requirement to properly define the keywords describing the technology domain.

These keywords are then used for the initial search of a patent database for relevant patents (Step 3) [163][124][164]. The success of this step dramatically depends on the quality of the patent database being used. More accurately, it greatly depends on how the database parses queries and searches for patents matching the selected criteria. Support for Boolean operators (“AND”, “OR”, “NOT”…) is preferable as it enables creating more complex keyword combinations [163][76][88]. Consequently, in this research, a database supporting Boolean operators was used as the primary data source.

The results of the initial keyword-based search often contain a large number of patents, a significant percentage of which are “false positives,” i.e. patents not related to the technology domain. This is because databases often parse not only patents titles but entire applications, which may contain technology domain keywords in their claims but are not related to the technology domain. Moreover, at this stage of the dataset creation, it is better to err by retrieving patents not relevant to the technology domain than by not retrieving patents relevant to the technology domain. Because of the retrieval of irrelevant patents, a filtering of the results must be made. This filtering process consists of using classification codes used to classify the patents in the dataset as the basis on which irrelevant patents are removed from the dataset.

In Step 4 of the dataset creation, the classification codes used to classify the patents within the dataset are collected and ranked based on the frequency of their occurrence in the dataset and the most frequently used classification codes are identified. All of the collected classification codes are then manually examined, and their relevance as being representative of the technology domain is checked based on the information from Table 2 and Figure 3. Moreover, classification codes used to classify patents in the dataset which are not representative of the technology domain are also identified. A filtering process is then conducted (Step 5). Patents collected in the initial search based on keywords, but not classified using the appropriate classification codes for the technology domain, are identified and removed from the dataset.

Finally, in Step 6, a manual curation of the created dataset is made. Patents not relevant to the technology domain that were not filtered based on their classification codes in Step 5 are identified and removed manually. After the filtering process, the entire dataset is manually examined with the goal of identifying any unforeseen errors.

In order to prepare the collected dataset for the empirical studies, some further processing of the collected data has to be performed. The dataset is split into two separate datasets, the first one containing all of the metadata contained in the retrieved patent applications. These metadata include, but are not limited to, classification codes, applicant, date of application, abstract and jurisdiction. The second data set consists of only the backward citations of the retrieved patents. This split of the retrieved patents is illustrated in Figure 11. This additional pre-processing step of creating two datasets is the consequence of the heavy reliance of this research on using patent citations as the basis for analysis, which requires patent citation information in a separate dataset, making subsequent steps of the empirical research easier to implement.

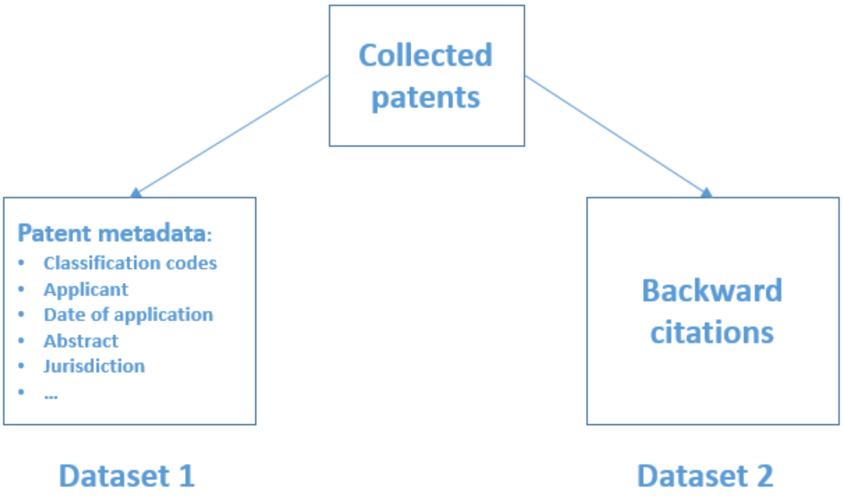

Figure 11 Illustration of the splitting of the dataset containing the collected patents

Several potential drawbacks of this approach should be noted. In addition to the known drawbacks incumbent to using patents as a source of data, this methodology has some additional potential drawbacks. Primarily, a danger always exists of certain patents being overlooked or added into the dataset by error. While the methodology used is reasonably accurate, errors are unavoidable. However, the number of patents added in error is deemed inconsequential compared to the size of the entire data set. Therefore, it is assumed that any errors have a negligible effect on the accuracy of the analysis.

To conclude this stage of the overview of empirical research, it should be emphasised that an important attribute of the dataset creation methodology is that it was devised with the guiding principle of using only free and open tools and databases. Consequently, both the patent database used as the primary source of patent data as well as the tools used for the processing of data and creation of datasets are all free to use. The databases used were the lens.org [165] database as well as google patents [166], while the scripting language Python was used for web scraping and data processing.

5.3. Empirical study - Technology Life Cycle Analysis

The second stage of the empirical research, as outlined in Figure 9, involves conducting a life cycle analysis of the selected technology domain.

As the review of literature shows, a number of methods for exploring the life cycle stages of a technology exist. In this research, we aim to contribute to the family of methods for determining the life cycle stages of a technology based on patent information by introducing a new method based on the dynamic growth analysis of a backward patent citation network. More specifically, it is explored whether the dynamic growth analysis of a patent citation network, created from patents from a technology domain, can be used to identify the life cycle stages of that technology domain. To reiterate, a backward citation network is a graph where nodes represent patents and edges represent citations, i.e. it is a network showing the relationships between antecedent and descendant patents. Figure 12 illustrates the workflow of this stage of the empirical research. As mentioned at the end of Chapter 5.2., the retrieved patents dataset is split into two datasets, the patent citation dataset and the patent metadata dataset. Each of these datasets is then used to conduct a technology life cycle analysis, with one analysis being based on one of the established methods covered in Chapter 4.2. and the other analysis being introduced in this research.

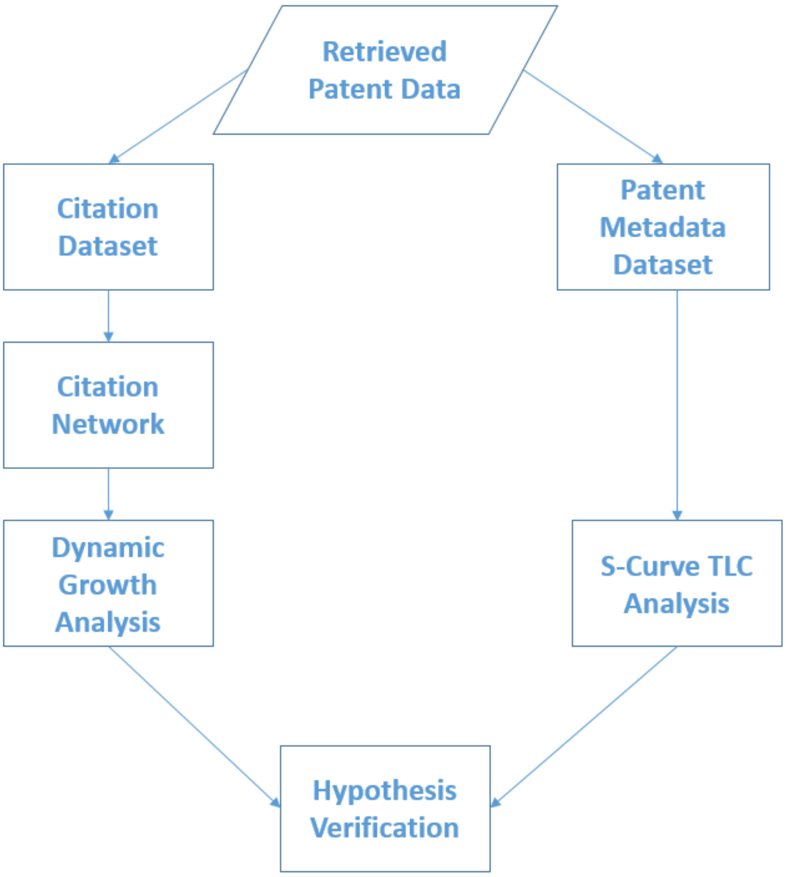

Figure 12 Flowchart illustrating the first empirical study

TLC analysis – Cumulative number of patents

From the patent metadata dataset, a technology life cycle analysis is conducted. This is conducted using one of the established patent based life cycle analysis methods outlined by Gao et al. [24]. More precisely, the method based on the cumulative number of patents applied over time, falling into the S-shape group of models, is used to generate a visualization of this growth [167] as was demonstrated in Chapter 3 and Chapter 4.2, the visualisation of the technology life cycle of a mature technology usually takes the form of an S-shaped curve and can be used to identify the technologies life cycle stages. Figure 6 shows a generalized representation of this S – Curve.

The reasoning behind conducting two life cycle analyses is that the established one serves as a verification and control method, the results of which will be compared to the results of the life cycle analysis method introduced in this thesis. This particular method, based on the cumulative number of patents, has been chosen as the control method for several reasons: a) multiple studies have demonstrated it as accurate for describing the life cycle stages of a technology [24][40][40][168], b) it is intuitive and straightforward to understand, c) it relies solely on the metadata contained in patent applications.

TLC analysis – Dynamic growth analysis

The second technology life cycle analysis is conducted on the patent citation dataset by applying a dynamic growth analysis on a network created from the patent citation dataset. This approach, used to conduct a dynamic network analysis, is adapted from the work of Štorga et al. [169]. The patent citation network is generated and continuously recalculated whenever a new patent, and its citations, are added. This allows for a visualization of the network's growth over time, illustrating the dynamics and evolution of the new citation network as new patents are added. From examining the continued growth of the network, a chart can be made illustrating the dynamics of growth. A generalization of this illustration is shown in Figure 13.

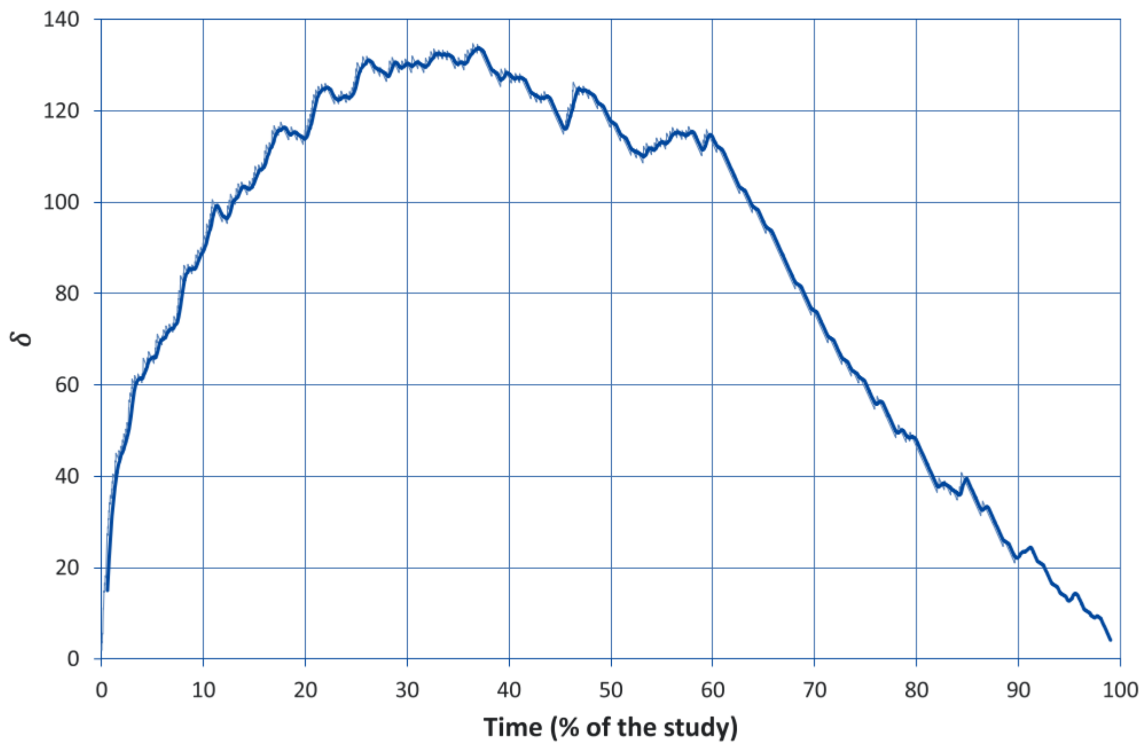

Figure 13 Generalization of a dynamic growth analysis graph [169]

As is seen in Figure 13, the abscissa in the graph represents a time step in the growth of the network. The ordinate, denoted by δ, represents the network's rate of growth.

The algorithm of the growth analysis, as outlined by Cash et al. [170], is as follows:

- For each step *i* in which a node was added or a nodal degree has increased, with respect to the corresponding total number of edges m or the total number of nodes n, the actual growth δ is measured with the following expression:

$$\delta(i) = m(i) - n(i) \qquad \textit{Eq. 1}$$

- The measure δ_e takes into account the total size of the network at the end of the study, which is then averaged over all steps i. This is performed with respect to the total number of steps $i=p$ constituting the session as well as the total number of edges m_p, and the total number of nodes n_p. Thus, for each step *i* the δ_e is defined by the following expression:

$$\delta_e(i) = i \times (m_p - n_p) \qquad \textit{Eq. 2}$$

- Finally, the relative network growth indicator δ_q per step i is defined with respect of measures as given by equations (1) and (2):

$$\delta_q(i) = \delta_e(i) - \delta(i) \qquad \textit{Eq. 3}$$

The results of the growth analysis provide insight into the network formation and growth dynamics, enabling the identification of different growth phases. In a generalized dynamic growth analysis of a network, a positive trend in the growth analysis graph corresponds to a period where more nodes than edges are added to the network. A negative trend corresponds to a period where more edges than nodes are being added. In the context of the research presented in this thesis, a positive trend means more new patents than new citations are being added to the network, while a negative trend means the opposite.

Once both the dynamic growth analysis and the S-Curve analysis have been conducted, the hypothesis that the dynamics of patent citations can be used to determine the life cycle phases of a technology is tested (Table 12).

5.4. Experimental study – Link prediction

In the final stage of the empirical research, an attempt is made to understand the dynamics of knowledge flow within a technology domain. The assumption is that future knowledge flow may be predicted by understanding the underlying intuition governing the dynamics of knowledge flow within a technology domain. An experimental study is conducted, applying link prediction algorithms to a patent co-citation network with the goal of assessing the precision of the algorithm in predicting future patent co-citations. The goal of this experiment is to identify the link prediction algorithm which best describes the dynamic of the patent – co-citation network growth and as well as explore the dynamic of the creation of the created links. The practice of using subsets is a common practice in research focused on comparing link prediction algorithms [15], as is the method for the evaluation of link prediction success [171][172][173]. Figure 14 illustrates the workflow used to conduct this stage of the empirical research.

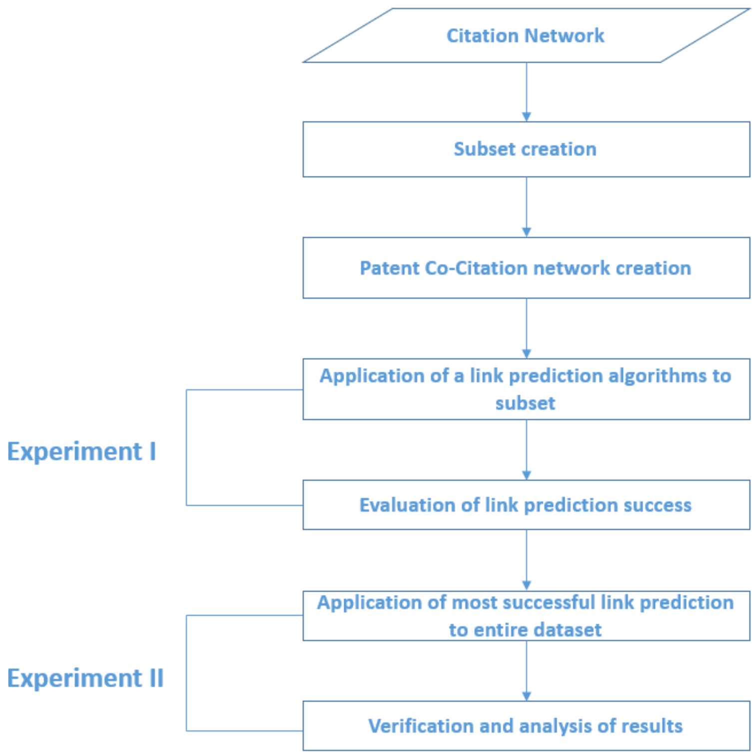

Figure 14 Flowchart of the second experimental study

Since the experiment is based on prediction in a time series problem, the dataset used to create the patent citation network in the previous stage of the empirical research must be divided into a training and testing set [174]. In order to prepare the training and testing sets for the experiment, more specifically, the application of the link prediction algorithm, the patent citation networks created from both sets must be converted into patent co-citation networks. Then, four link prediction algorithms are applied to a small subset of the training dataset, and the success of each algorithm is compared. The algorithm showing the highest level of success is chosen to be used in further research and applied to the rest of the training subsets. After a more detailed analysis is performed using the chosen algorithm, a detailed analysis and interpretation of results are made.

Subset and co-citation network creation

The forecasting part of this experimental study explores the dynamics of knowledge flow within a technology domain. More precisely, it explores how knowledge diffuses and combines to create new knowledge. These combinations of existing knowledge can be represented and visualized as a co-citation network, where nodes represent patents and edges represent links between co-cited patents.

A link prediction algorithm is used to predict future co-citations. However, a co-citation network must be created before a link prediction algorithm can be applied. This network is created by converting the existing patent citation network from the previous stage of the empirical research into a co-citation network. Using the existing patent citation network provides several advantages. An obvious one is that converting an existing dataset eliminates the need for another data retrieval procedure. However, another advantage is that the patent citation network is a time series when written in a tabular form. This enables the creation of co-citation network snapshots of different times in the technology's life cycle. More accurately, it simplifies the creation of multiple training and testing subsets, therefore facilitating the ability to conduct a large amount of analysis.

The process of converting the patent citation network into the patent co-citation network is done by applying an algorithm adapted from the work of Štorga et al. [169], and is demonstrated in the patent co-citation network creation algorithm, outlined in Table 17:

Table 17 Patent co-citation network creation algorithm in continuous time

```
Algorithm: Patent co-citationon network creation
    Input: record set S, edge label l_e
    Output: patent label set ∑, network G = G(t) represented over t
 t ← 0;
 foreach record r ∈ S do
     l_v ← Read (r);
     if l_v ∉ ∑ then
         add l_v into ∑
         add node v to node set V;
         map label to node l_v → v;
 fi
 citation ← Read (r);
 P ← ρ(l_v, V, citation);
 For each node u ∈ P do
         Establish edge e = {v, u};
         Add edge e to edge set E;
          Map label to edge l_e → e;
 od
 t ← t + 1
 od
```

In order to assess the performance of the link prediction algorithms, historical network data is used. Since link prediction is a time-related activity, the dataset must be sequenced in a chronological order to separate the data. The patent citation dataset is segregated into a training dataset, marked $T_{t,t1}$ (V_1, E_1), consisting of a set of nodes V_1 and edges E_1 from time t to time t_1, and testing dataset, marked $V_{t1,t2}$ (V_2, E_2) consisting of a set of nodes E_2 and edges L_2 from time t_1 to time t_2. The split point is determined by the technology life cycle analysis results from the previous stage of empirical research (Chapter 5.3.) For the mature technology domain, the point of the split is chosen at the time where the technology life cycle transitions from the growth stage to the maturation stage. The training and the testing sets are segmented into further subsets, marked T_n and V_n (Figure 15), the largest subsets consisting of the entire training/testing set with each subsequent subset progressively smaller in size. For an emerging technology, it is assumed that the results of the technology life cycle analysis will not show clear distinctions between life cycle stages. Therefore, the split between training and testing datasets is done as a sliding window, moving the split point in regular intervals Figure 16.

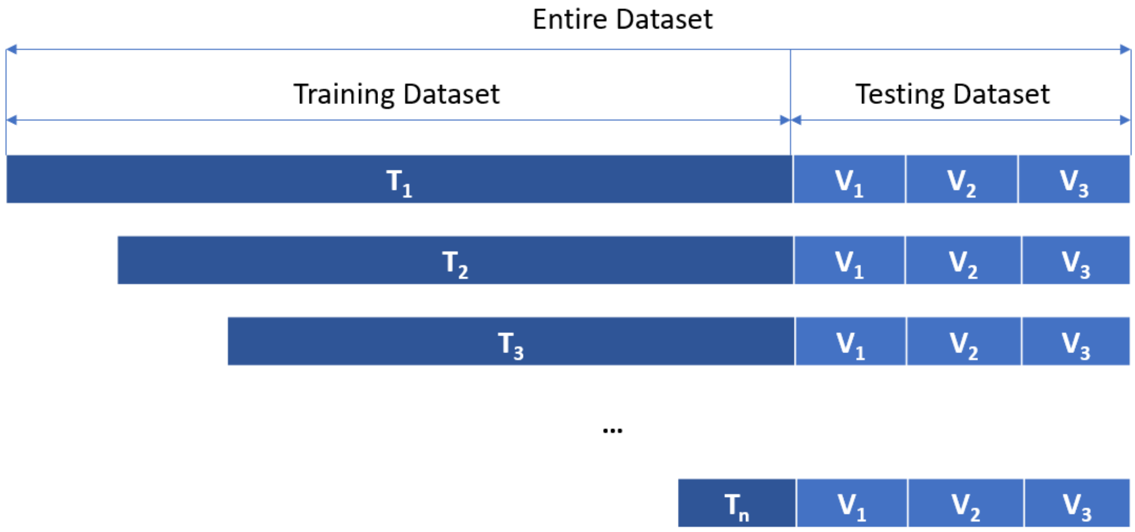

Figure 15 Visualisation of training and testing dataset and subset creation (mature technology domain)

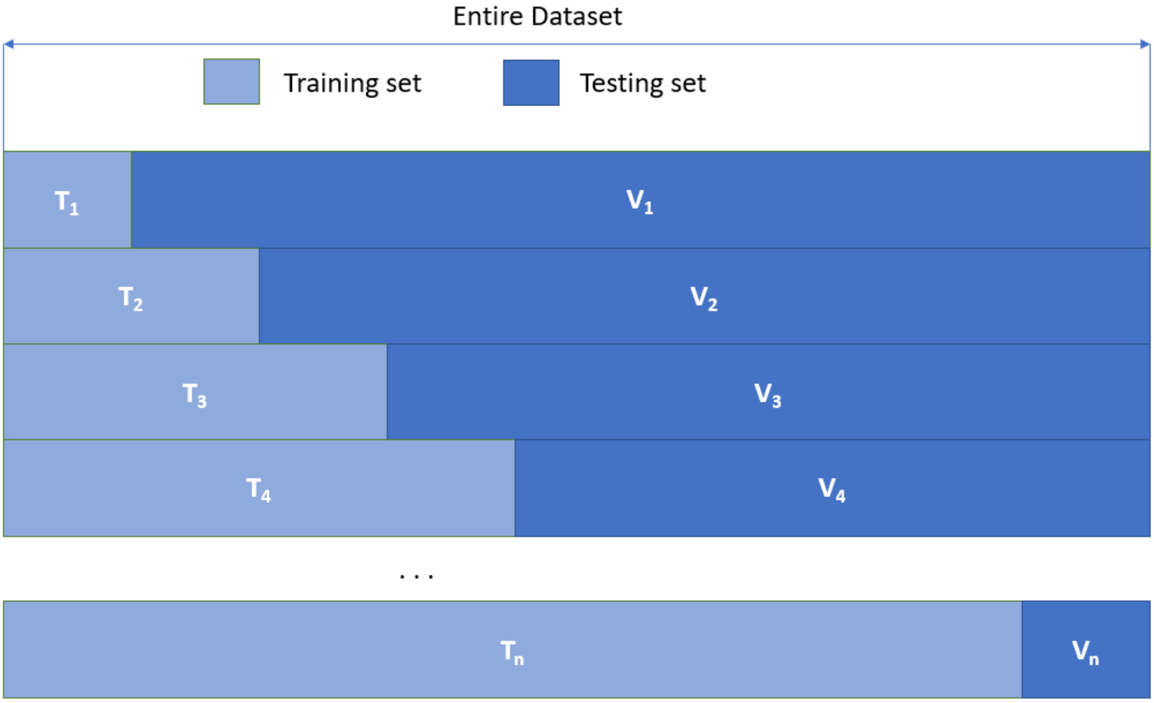

Figure 16 Visualisation of training and testing dataset and subset creation (emerging technology domain)

The segmenting of both the training and testing datasets into subsets allows us to study the dynamics of patent co-citation network growth in more detail. The segmenting of the training dataset enables us to study which part of the training dataset contributes the most to the precision of the link prediction algorithm. By applying the link prediction algorithm to each subset of the training dataset and comparing the respective precisions, a deeper understanding can be obtained of how each segment contributes to the precision of the algorithm. Translated into the context of studying knowledge flow, the influence of older vs newer knowledge on future inventions can be studied. The segmenting of the testing datasets into subsets enables us to study the time frame in which the predicted links will occur. Translated into the context of studying knowledge flow, it provides insight into how soon the predicted knowledge will occur.

From each of the created subsets, a patent co-citation network is created. The created co-citation network is a dynamic network consisting of a set of edges, denoted *u*, and a set of vertices denoted *v*, which grows and changes over time with the addition of new nodes and edges.

Link prediction

It is assumed that the growth of the patent co-citation network follows some pattern that can be identified. Consequently, by identifying the pattern by which the patent co-citation network grows, it is assumed that the future growth can be predicted. Link prediction algorithms (often called similarity measure algorithms) are used as a tool for identifying this pattern. A computational problem underlying network evolution as defined by Liben-Nowell and Kleinberg [175] is studied: "Given a snapshot of a network at time *t*, we seek to accurately predict the edges that will be added to the network during the interval from time *t* to a given future time *t'*." In the context of this research, the question is reframed and becomes: "Can we accurately predict the edges being added to a patent co-citation network, as well as the time when the predicted edges will be added?". The link prediction problem is therefore the attempt of inferencing which new interactions are likely to occur in a network, given its snapshot [175]. The creation of a network and the adding of new edges between nodes are both dynamic events based on local interactions among nodes that shape the evolution of the network [176]. Therefore, link prediction algorithms aim to predict the likelihood of a link occurring between two nodes in a complex network. According to Zhou et al. [177], the link prediction problem can be stated as:

Consider an undirected simple graph G (V, E), where V is the set of nodes and E is the set of edges. For each pair of nodes, $u, v \epsilon V, e = (u, v) \epsilon E$ *represents a link between the two nodes. For every possible pair of nodes, a link prediction algorithm assigns a score, Sim(u,v), as a* *measure of similarity between the two nodes. The non-existent links should be sorted in a descending order according to their scores, with the links on top most likely to be established.*

Generally speaking, link prediction problems occur in 3 instances, as demonstrated in Figure 17: a) only the addition of new links within a network, b) only the removal of existing links from a network, and c) the simultaneous addition and removal of links at the same.

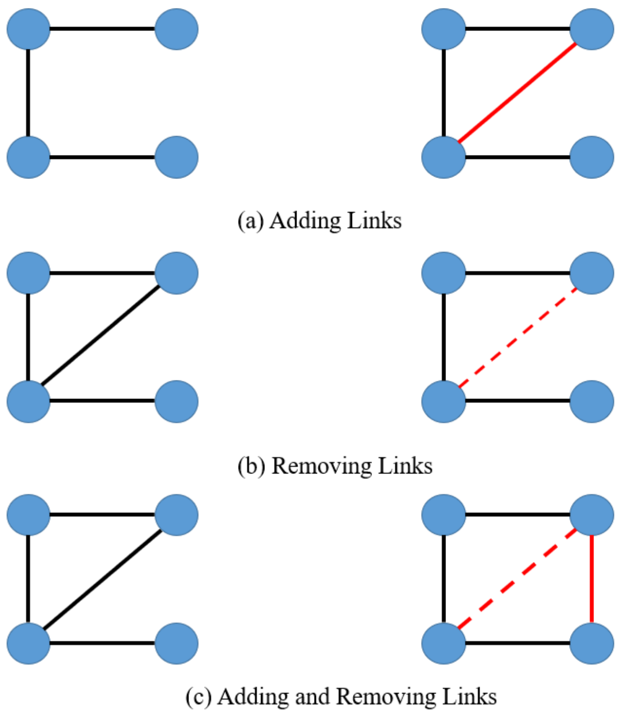

Figure 17 Instances of link prediction problems

In this research, the first type of link prediction problem is studied. Since the patent citations in patent applications are immutable, existing links cannot be removed from a co-citation network. Therefore, only additions can be made to a snapshot of a co-citation network. A set *U* is defined, containing all possible links within a network. Quantitatively, the number of links contained in set *U*, *n(U)*, can be defined as

$$n(U) = \frac{|V| \times (|V| - 1)}{2} \qquad \textit{Eq. 4}$$

where $|V|$ is the number of nodes in set *V*. Then, a set of non-existent links can be defined as the difference between all possible links *U* and the existing links in a set *E*. Quantitively, this set can be defined as:

$$NL = U - E \qquad \textit{Eq. 5}$$

Finally, it is assumed that there are some missing links (or links that will appear in the future) in the set *NL*. However, these missing links are not known as they have yet to occur. Therefore, to test the link prediction algorithm set *E* is divided into two subsets, a training one and a testing one E^P, where the missing links are the links appearing in set *E* but not appearing is set E^T:

$$ML = E - E^T \qquad \textit{Eq. 6}$$

The goal of link prediction is to predict the links in set *ML* as successfully as possible.

The output of a link prediction algorithm is a tuple consisting of three columns; the first two columns containing the ID of the nodes being connected by the predicted link and the third column containing the similarity measure, i.e. likelihood that the predicted link between the two nodes will occur. In principle, this tuple represents an ordered list of all non – observed links in the network or equivalently given each non-observed link a similarity measure score to quantify its existence likelihood [171]. Table 18 shows a generalisation of the output provided by the link prediction algorithm. The predicted links are sorted by the value of the similarity measure. For most methods, a higher similarity measure value means a higher likelihood that the predicted link will appear in the future.

Table 18 Generalisation of Link Prediction Output

Node 1	**Node 2**	**Similarity Measure**
Id..1	Id..2	Sim_value 1
Id..1	Id..3	Sim_value 2
Id..2	Id..4	Sim_value 3
....	...	…
Id..x	Id..y	Sim_ Value n

This similarity measure ranges from 0 (no likelihood of a link occurring) to a maximal value, the maximal value varying from case to case and also depending on the link prediction algorithm being used. It should be noted that, when applying a link prediction algorithm to a network, the resulting tuple contains the similarity score for every possible combination of nodes (complete graph). Consequently, this tuple can grow quite large, especially in more extensive networks. Note that a complete graph has a number of edges equal to Eq. 4, where

$|V|$ is the number of nodes in set V, making the handling of the tuple quite unpractical. Therefore, a cut-off value is applied to the returned tuple, filtering out all of the predicted links beneath a certain threshold.

Two standard metrics exist to quantify the accuracy of a link prediction algorithm [171]. These are the area under the receiver operating characteristic curve (AUC) [178] and precision [179][180][181]. The main difference between these two metrics is that the AUC evaluates the algorithms' performance according to the entire list of predicted links, while the precision metric only focuses on the links with a precision above the cut-off values. A more detailed overview of these two metrics is as follows:

a) AUC: Provided the similarity values of all non – observed links, the AUC value can be interpreted as the probability that a randomly chosen missing link is given a higher similarity value than a randomly chosen non-existent link. In an algorithmic implementation, the score of each non-observed link is calculated. Then, at each time a missing link and a non-existent link are randomly picked to compare their score. If among n independent comparisons, there are n' instances of the missing link having a higher score and n'' times they have the same score, the AUC value is [171]:

$$\text{AUC} = \frac{n' + 0.5n''}{n} \qquad \textit{Eq. 7}$$

b) Precision: Precision is defined as the ratio of relevant items selected to the number of items selected. If, after applying the cut-off rate to the output of the link prediction algorithm, N_s predicted links remain, among which N_{rs} predicted link are correct (i.e. they occur in the testing set), then the precision of the algorithm is defined as [179]:

$$\text{P} = \frac{N_{rs}}{N_s} \qquad \textit{Eq. 8}$$

In order to deepen the understanding of the dynamics of patent co-citation network growth, several different link prediction algorithms are compared in order to determine which one describes the dynamics of network growth with the highest precision. As the compared link prediction algorithms all have a different underlying intuition, it is hypothesized that the results of the initial analysis will show one link prediction algorithm preforming significantly better than the others, meaning it is the best at describing the underlying intuition of network growth dynamics.

Four link prediction algorithms are compared in this research with the goal of identifying the link prediction algorithm most successful in predicting the missing link in a patent co-citation network. An overview of these algorithms is provided in Table 19.

Table 19 Overview of link prediction algorithms used

Algorithm name	Similarity measure definition	Underlying intuition
Resource allocation index [177]	$Sim(u,v) = \sum_{w \in \Gamma(u) \cap \Gamma(v)} \frac{1}{\lvert\Gamma(w)\rvert}$	Common neighbours
Jaccard coefficient	$Sim(u,v) = \frac{\lvert\Gamma(u) \cap \Gamma(v)\rvert}{\lvert\Gamma(u) \cup \Gamma(v)\rvert}$	Common neighbours, similarity of sample sets
Adamic – Adar index [182]	$Sim(u,v) = \sum_{z \in \Gamma(u) \cap \Gamma(v)} \frac{1}{\log\lvert\Gamma(z)\rvert}$	Low-degree neighbour is more likely to indicate a future connection then a high-degree one
Preferential attachment [183]	$Sim(u,v) = \lvert\Gamma(u)\rvert \cdot \lvert\Gamma(v)\rvert$	Rich – get – richer

In the empirical study covering the mature technology domain, each of the link prediction algorithms presented in Table 19 is applied to a subset of the training set closest to the point where the dataset was split into the training and testing set (Figure 15). To reiterate, this split point roughly coincides with the transition period between two technology life cycle stages, namely growth and maturation. The link prediction algorithm showing the highest precision is then applied to the rest subsets of the training set.

In the empirical study covering the emerging technology domain, the four link prediction algorithms outlined in Table 19 are applied to the training and testing subsets as outlined in Figure 16. Because of the unpredictable nature of the emerging technology domain, demonstrated in Chapter 2.1.3, it is assumed that there will be no clear delamination line between life cycle stages. Consequently, the link prediction algorithms are applied to a wider spectrum of training subsets in relation to the mature technology domain.

Identifying the appropriate link prediction algorithm which can be used to describe the growth dynamic of a patent co – citation network enables us to gain insight into the underlying intuition governing the diffusion of knowledge within a technology domain i.e. how existing knowledge combines to create new knowledge artefacts.

6. EMPIRICAL STUDY – MATURE TECHNOLOGY DOMAIN

This chapter presents the results of the first empirical study focusing on exploring a mature technology domain. An overview of the selected technology is provided, as is the method for dataset creation. A life cycle analysis is conducted, and the results are compared to previous research. After the life cycle analysis, a link prediction algorithm is chosen and applied to a patent co-citation network with the goal of exploring the underlying intuition of network growth and predicting the future development of the technology.

6.1. Technology background – Car Headlights

The first empirical study presented in this thesis covers the technology domain of car headlights. This is a technology that has seen a continuous improvement since its inception in the 1880s. Because of its long history of being implemented in a commercial product, in the context of this thesis it is used as an example of a mature technology, i.e. a technology in the mature phase of its life cycle. This evaluation of the technologies life cycle stage is based on both expert knowledge and a review of the literature covering innovations from the automotive sector [184][185].

While the earliest headlights were fuelled by acetylene or oil, modern incarnations use electricity as an energy source, with the most proliferated technology in use being tungsten-halogen headlights. More advanced models do exist, including high-intensity discharge (HID) headlights, often called “xenon headlights” because they are filled with xenon gas. At present, the state of the art headlight technology used in vehicles are headlights that use light-emitting diodes (LED) as a light source. Finally, laser-based high beam lamps are being introduced in some premium vehicle models. Table 20 shows an overview of key technologies in the car headlights technology domain and their approximate time of commercial introduction [184][185].

Table 20 Key years in car headlights technology evolution

Technology	Approximate time of technology's introduction
Acetylene/Oil	1880
Electric headlights	1900
Sealed Beam (tungsten)	1940
Tungsten + halogen	1960
High-intensity discharge (HID)	Mid 1990s
Light-emitting diodes (LED)	Early 2000s
Lasers	The mid-2010s

6.2. Dataset creation

A dataset consisting of patents from the car headlights technology domain is created as outlined in the dataset creation part of the chapter outlining the design of empirical research (Chapter 5.2).

Data retrieval

The patents relevant to the technology domain were retrieved using a method based on the one outlined by Benson et al. [163]. The dataset creation process started by searching the chosen patent database using the keyword pair "car headlights". An analysis was conducted on the preliminary results and the most frequent CPC codes relevant to the field used to classify the patents were identified. Table 21 shows the most frequent identified CPC codes as well as their definitions. A manual examination of the definitions of the identified CPC codes confirms their relevance to the examined technology domain.

Table 21 The most frequent relevant CPC codes in the retrieved dataset

CPC Code	Definition
B60Q1	Arrangements or adaptations of optical signalling or lighting devices
F21S4	Lighting devices or systems using a string or strip of light sources

The identification of these classification codes enables the identification of patents retrieved based on the keyword search but not relevant to the studied technology domain. Patents not containing the identified classification codes were then filtered out, increasing the relevance of the patent dataset. Moreover, patents classified with classification codes unrelated to the technology domain were also filtered out and removed from the dataset. It should be noted that the patents filtered out were manually reviewed to ensure that no relevant patents were filtered out by mistake.

After the cleaning of the dataset, 14 114 relevant patents remain. The oldest patent has the application year of 1902 while the youngest patent has the application year 2017. The dataset is stored in the form of a table, where the patents are listed sequentially in chronological order, from oldest to youngest. This table contains all of the metadata related to the patents except backward and forward citations. A separate table is created containing all of the retrieved patents as well as their backward citations. This split of data into two datasets is covered in Chapter 5.2, Figure 11.

6.3. Empirical study - Technology Life Cycle analysis

In this step, a life cycle analysis of the studied technology domain is conducted, based on the method outlined in Chapter 5.3. First, a technology life cycle analysis is conducted using an established method based on plotting the cumulative number of patent applications over time, and the technology domains life cycle stages are identified. Then, a dynamic growth analysis is conducted of the patent citation network. By superimposing the results from the two analyses, the dynamics of patent citation network growth can be contextualized within the phases of a technologies life cycle. These activities correspond to the stage of Empirical research presented in Figure 9 marked Empirical study.

Life cycle analysis based on the cumulative number of patent application

In determining the life cycle phase of the technology domain, we first employ one of the established methods for patent based life cycle analysis, presented in Chapter 4.2. This method uses the cumulative number of patent applications to plot a curve visualizing this change over time [24]. The data contained in the table with patent metadata are analyzed and the results of the analysis are visualized in Figure 18.

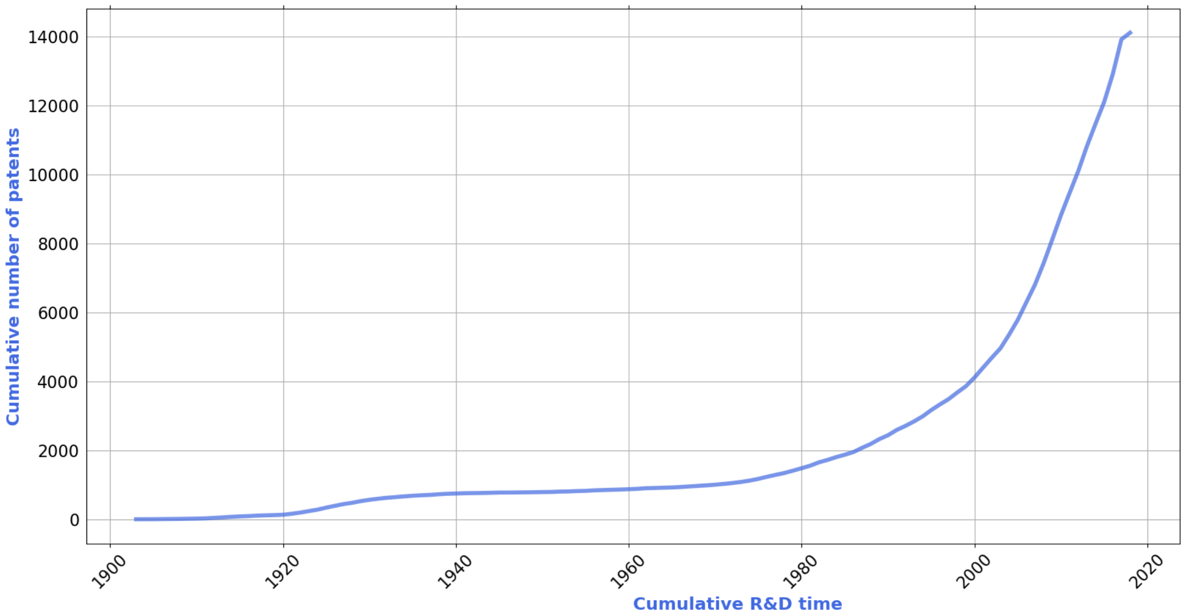

Figure 18 S - Curve of the car headlight technology domain based on the cumulative number of patents

Figure 18 shows the cumulative number of applied patents on the Y-axis and the application year on the X-axis. It is important to note the age of the examined technology domain, with the first patent occurring in 1899. While there was an increase in the number of applied patents from that time, Figure 18 makes it noticeable that the growth of the cumulative number of patents was relatively slow, at least in the first 70 years of the technologies life cycle. This is primarily due to the fact that inventions were very rarely patented until the final quarter of the twentieth century [186]. This is immediately observable by examining the cumulative number of patent applications from 1900 to 1980, with the cumulative number never exceeding 2000.

There is a noticeable change in the following period, with the period from 1980 to 2020 seeing an increase of the number of patent applications by a factor of 7.

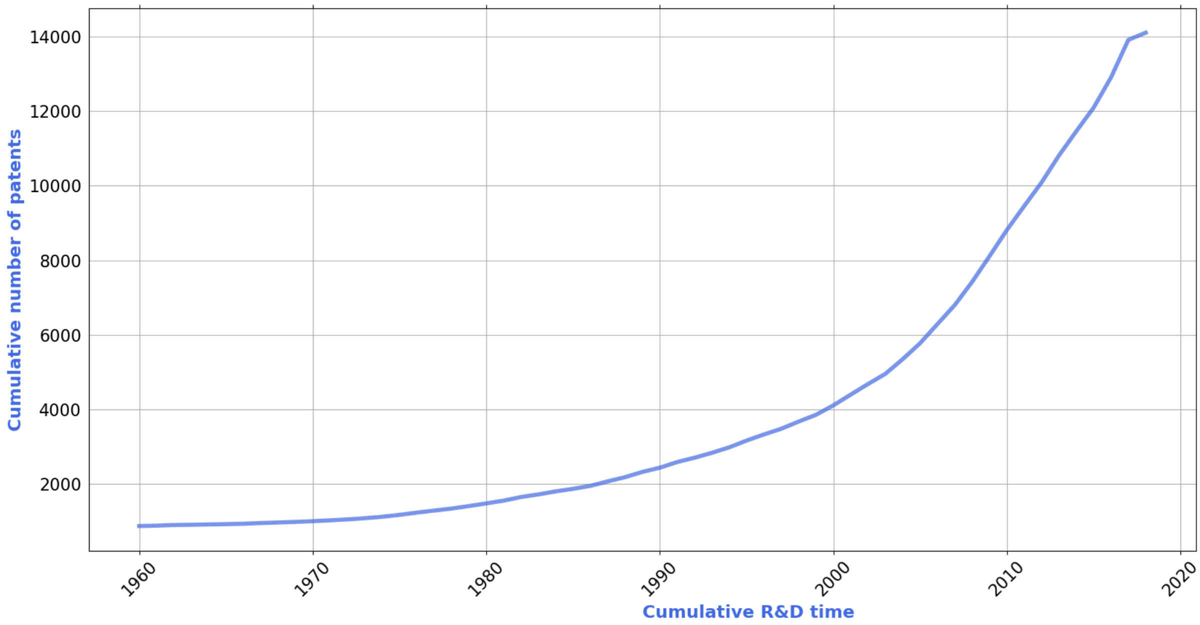

Figure 19 Curve of the car headlight technology domain based on the cumulative number of patents (focused)

Figure 19 also shows the cumulative number of patent applications over time but focuses on the period between 1960 and 2020. This chart provides a more relevant insight into the life cycle stages of the examined technology domain, focusing on the period of a contemporary level of patent activity. The curve in Figure 19 describes the growth of the technology domain, providing insight into its maturity. There is a slow growth of the curve until roughly 1985 indicating an initiation phase. This is followed by an increase in the curves growth rate, peaking at about 2005 and then starting to slow down. Even though the number cumulative number of patents continues to increase, the decrease of the rate of growth signals that the technology domain is losing momentum [76]. This loss of momentum, roughly starting at around the year 2012, corresponds to the end of the growth stage and transition to the maturation stage.

Life cycle analysis based on the dynamic growth analysis

Using the method outlined in Chapter 5.3, a dynamic growth analysis of a patent citation network is conducted based on table containing backward patent citations with the goal of exploring whether a correlation exists between the dynamics of patent citation network growth and the stages of a technologies life cycle.

The patent citation network consists of nodes representing patents and edges representing direct citations. The network is generated and continuously recalculated after the addition of a new node and link. The whole patent citation network, at the end of its growth, consists of 53.910 nodes and 73.946 edges.

The growth analysis provides insight into the network formation and growth dynamics, enabling the identification of network's different growth phases. A positive trend in the growth analysis corresponds to a phase where more nodes than edges are added to the network. In the context of this research, this means that more new patents than citations are added to the network. A negative trend corresponds to a phase where more edges than nodes are being added i.e. more citations than patents are being added.

Figure 20 shows the results of the dynamic growth analysis of the patent citation network. A clear period of positive growth can be identified, followed by a shift to a negative trend. It should be noted that in this case, the step of the study represents the cumulative number of patents in the technology domain.

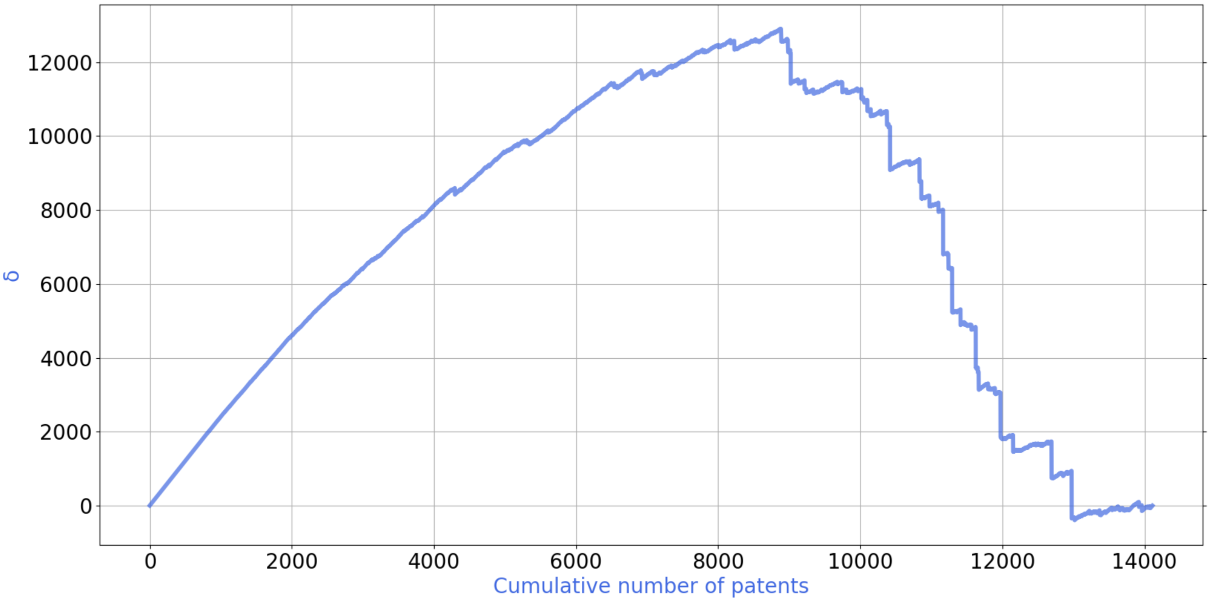

Figure 20 Technology life cycle of examined technology based on the patent citation network analysis

Figure 21 shows the results of the dynamic growth analysis of the patent citation network superimposed with the results of the S – Curve based life cycle analysis. The life cycle stages identified by the S – Curve growth are marked on the graph as well as the introduction time of significant technologies in the car headlight technology domain based on table one.

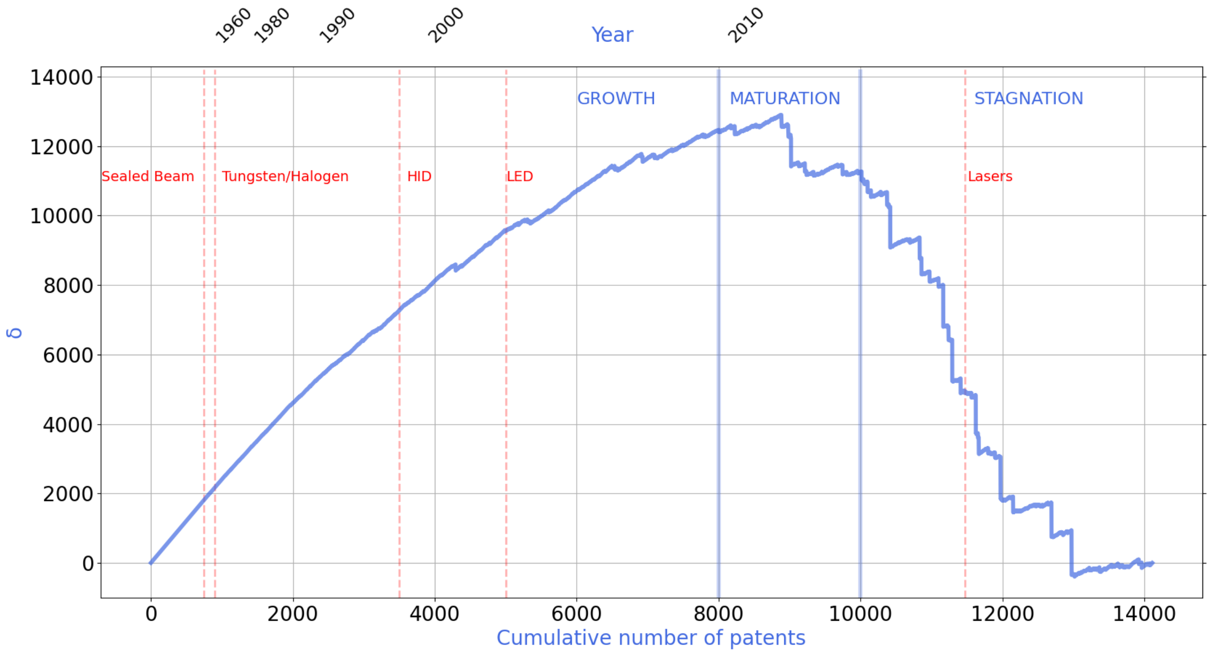

Figure 21 Growth analysis superimposed with results from S – Curve analysis

Consequently, Figure 21 provides context to the results of the analysis of the dynamic of patent citation network growth. During the growth life cycle stage, more new patents than citations are introduced into the network resulting in a positive growth of the graph. During the maturation phase, a reduction in the slope of the curve can be observed, followed by a decrease in the slope resulting in a negative trend. This means the number of new patents in the maturation stage is roughly the same as the number of new citations, indicating that a negative trend will follow. Finally, the saturation phase sees a sharp decrease in the slope of the curve resulting in an even more pronounced negative trend. This means the saturation phase sees a significantly higher number of new citations than new patents. These results are in concordance with the most widely accepted theories of technology evolution [16][28][64]. During the growth phase of a technology domain, more novel inventions are being introduced, resulting in a higher number of patents. As a technology domain matures and starts to stagnate, new inventions are rarer and are often the results of the combinations of existing technologies, resulting in a relatively higher number of citations related to new patents.

The first quantitative analysis presented in this subchapter, conducting a life cycle analysis based on the cumulative number of applied patents, confirms the findings of previous research stating that the number of cumulative patent applications plotted over time follows an S – Curve with discernible life cycle phases [40][24].

The second quantitative analysis conducts a dynamic growth analysis on a patent citation network. While citation data has been used to determine the life cycle stage of a technology, it was generally based on examining the number of forward and backward citations or other cumulative metrics of patent metadata [24][84][113]. Therefore, the presented approach represents a novel approach to exploring a technologies life cycle by exploring the correlation between the life cycle stages of a technology and the dynamic of patent citation network growth, taking into account both the change of the number of citations and patents.

6.4. Experimental study – Link prediction

In this stage of the empirical study, link prediction algorithms are applied to a patent co-citation network with the goal of identifying the link prediction algorithms showing the highest success in predicting missing links. The patent co-citation network is created based on the algorithm outlined in Chapter 5.4. The dataset is segmented into two sets: a training set, on which the link prediction algorithms will be applied, and a testing set, which will be used to evaluate the precision of the link prediction algorithm. The time where the subset will be split is determined based on the results of the technology life cycle analysis. Furthermore, both the training and testing sets are segmented into smaller subsets in order to provide further insight into which period of a technologies life cycle contributes the most to the accuracy of the link prediction algorithm as well as when the predicted missing links occur.

The end of the growth stage/beginning of the maturation stage, which roughly corresponds to step 9000 of the study, is chosen as the split point between the testing and verification subsets. Based on the results of the technology life cycle analysis (Chapter 6.3.), it is observed that a change in the life cycle stage from growth to maturation corresponds with a change in the dynamic of patent citation network growth. To clarify, this period sees the start of the decline of growth, meaning more new edges than nodes are added to the network, which makes it a promising period to apply a link prediction algorithm. Moreover, to better understand the dynamics of the growth of the technology domain, both the training and testing data sets are split into smaller subsets, in line with Figure 15. The largest training subset consists of the entire training set, with each subsequent subset progressively smaller in size, resulting in nine training subsets in total. The testing subset was split into three subsets, each one progressively larger than the previous one. This progressive increase is assumed to be adequate to contribute to identifying the time frame in which the successfully predicted links occur. Both the training subsets and testing subsets were coded for easier subsequent referencing (Table 22 and Table 23).

Table 22 Overview of training subsets with matching codes

Step of Study	0-9000	1000-9000	2000-9000	3000-9000	4000-9000	5000-9000	6000-9000	7000-9000	8000-9000
Code	T1a	T2a	T3a	T4a	T5a	T6a	T7a	T8a	T9a

Table 23 Overview of testing subsets with matching codes

Step of Study	9000-11000	9000-13000	9000-14113
Code	V1a	V2a	V3a

A patent co-citation network is created from each of the subsets using the algorithm outlined in Chapter 5.4 (Table 17). The four link prediction algorithms presented in Table 19 are applied to a patent co-citation network created from subset T8a representing the co-citations created from patents applied shortly before the peak of the growth stage. The four link prediction algorithms are tested as predictors for future links in a patent co-citation network. The reason four different link prediction algorithms are considered is because each describes a different underlying intuition of network growth.

The four link prediction algorithms are tested by applying each one to a small subset of the training dataset, T8a. The reasoning for choosing this particular subset was based on the intuitive hypothesis that the period encompassing the end of the growth and start of the maturation phase marks a period of a reduced pace of innovation [16][31][64], manifested in a patent co-citation network as a period where the number of newly created edges is higher than newly created nodes, making it a promising starting point for the application of a link prediction algorithm.

The precision of each link prediction algorithm, calculated as presented in *Eq. 8.*, as well as the AUC, presented in *Eq. 7*, are shown in Table 24.

Table 24 Precision of link prediction algorithms applied to T8 (20% cutoff)

Algorithm	Precision	AUC
Resource Allocation Index	0.52	0.64
Jaccard	0.2	0.41
Adamic – Adar	0.4	0.78
Preferential Attachment	1	n/a

Based on the results shown in Table 24, it is clear the preferential attachment link prediction algorithm shows the highest precision when applied to the co-citation network created from subset T8a. Note that because of a precision of 1, the AUC can not be calculated. Based on these results, the Preferential Attachment link prediction algorithm is chosen to be applied to the rest of the training subsets. To reiterate, the Preferential Attachment link prediction algorithm is a greedy algorithm devised based on the insight that, in most real networks, new edges are not created randomly but have a higher probability of connecting nodes that have a higher degree (higher number of connections) [176]. Similarly, in the context of this research, it is reasonable to assume that patents having a higher number of citations also have a higher chance of being cited again. The class of system might then be described by a model based on a preferential attachment mechanism, more commonly known as a rich – get–richer mechanism.

To reiterate, the preferential attachment score of node u and node v is defined as:

$$Sim(u,v) = |\Gamma(u)| \cdot |\Gamma(v)| \qquad \text{Eq. 9}$$

where $\Gamma(u)$ denotes the set of neighbours of node u and $\Gamma(v)$ denotes the set of neighbours of node v. The basic premise is that the probability that a new edge involves node x is proportional to $|\Gamma(x)|$, the current number of neighbours of x.

The Preferential Attachment link prediction algorithm is applied to the rest patent co-citation networks created from the training dataset subsets (T1a-T9a, Table 22). As mentioned in the chapter outlining the design of empirical research (Chapter 5.4), link prediction algorithms output very large datasets containing the similarity scores of all possible links that can occur within the network the algorithm was applied to. As a large percentage of those scores equals zero, or are very low, a cut-off value is determined in order to verify the predicted links in a practical way [171]. In this research, the cut-off values are determined as percentages of the maximum calculated similarity value. For example, a cut-off rate of 80% would filter out all predicted links with a similarity measure value less than 80% of the maximum predicted similarity measure values.

The results of the Preferential Attachment link prediction algorithm can be seen in Table 25 - Table 27. The link prediction algorithm was the most precise when applied to these three subsets (T7a-T9a). As the results of the preferential attachment link prediction algorithm applied to subsets T1a-T6a all showed a lower level of precision compared to T7a-T9a, only the results based on subsets T7a-T9a are presented in this chapter while the rest of the results can be found in the Appendix. The first subset showing noticeably high precision is T7a, demonstrating a high precision when the cut-off value is set to 90%, the precision then dropping when the cut-off value is reduced (Table 25). However, the subset T8a demonstrates a noticeable increase in precision, reaching a value of 1 in the top 10% and top 20% of predicted links, and dropping to 0.95 in the top 30% and top 40% of predicted links (Table 26). The precision decreases drastically in the T9a subset, never rising over 0.77 (Table 27), which is an intuitive result since this training subset is half the size of the other training subsets, implying the sample size is not large enough for the link prediction algorithm to make accurate predictions. Based on these results, it can be deduced that, in order to get accurate results, the preferential attachment link prediction algorithm needs to be applied to only the patents close to the end of the growth stage and the start of the maturation stage. This is a significant finding because the link prediction process is very resource and time intensive. Consequently, any reduction in the testing dataset size dramatically improves the performance of the algorithm.

Table 25 Link prediction results for T7a

Percentage of predicted links being verified	Similarity measure value	Number of predicted links	Number of correctly predicted links	Precision	AUC
Top 10 %	247 295	12	11	0.91	0.81
Top 20 %	219 820	60	56	0.93	0.61
Top 30 %	192 342	588	541	0.92	0.59
Top 40 %	164 865	1 045	918	0.87	0.61

Table 26 Link prediction results for T8a

Percentage of predicted links being verified	Similarity measure value	Number of predicted links	Number of correctly predicted links	Precision	AUC
Top 10 %	199 836	7	7	1	1
Top 20 %	177 632	38	38	1	1
Top 30 %	155 428	294	282	0.95	0.55
Top 40 %	133 224	2 483	2 381	0.95	0.62

Table 27 Link prediction results for T9a

Percentage of predicted links being verified	Similarity measure value	Number of predicted links	Number of correctly predicted links	Precision	AUC
Top 10 %	99 014	633	490	0.77	0.87
Top 20 %	88 012	2 409	1 058	0.43	0.58
Top 30 %	77 011	8 732	1 980	0.22	0.82
Top 40 %	66 009	51 281	2 149	0.04	0.93

The results from Table 25 - Table 27 are summarised and presented in Figure 22, showing how the precision of the preferential attachment link prediction algorithm changes applied to the different training datasets and in relation to the change of the cut-off value.

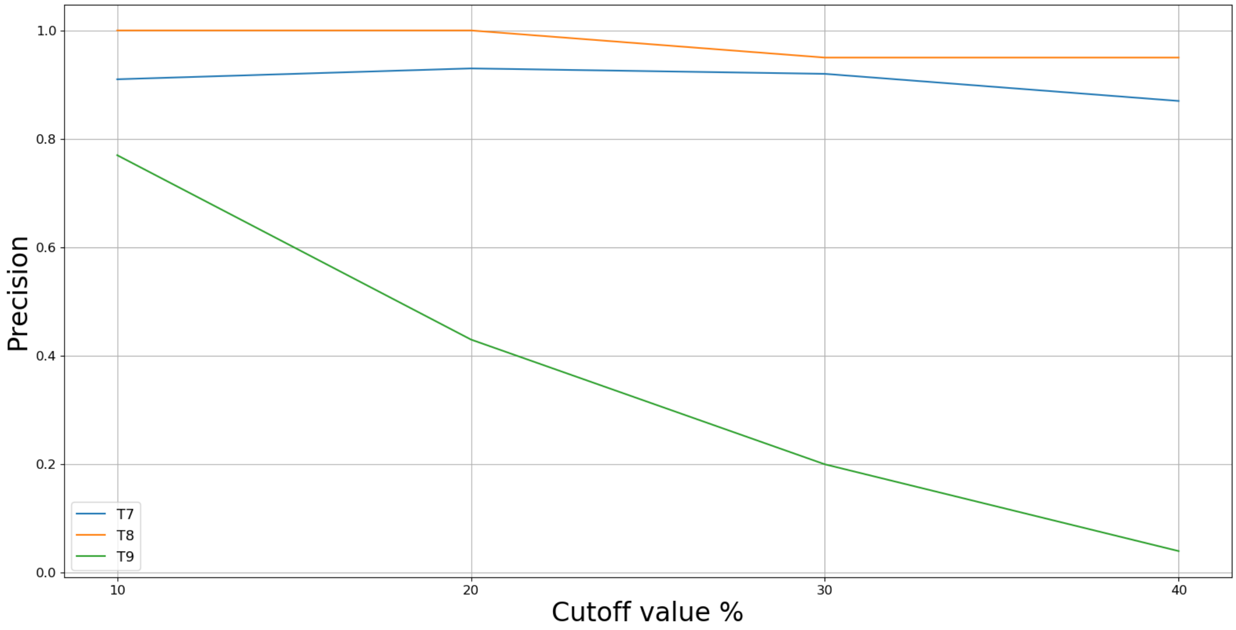

Figure 22 Relationship between the cut-off value, precision and training dataset size

Table 28 - Table 30 show the approximate time frame it takes for the predicted links to occur after the time point where the prediction algorithm was applied (the end of the growth stage). The occurrence of the predicted links is compared with the testing subsets outlined in Table 23. Based on the results, most of the predicted links are created in the near to mid future, while a small number of predicted links is created at a later time.

Table 28 Occurrence time of predicted links for subset T7a

Percentage of predicted links being verified	Number of correct predictions in V1	Number of correct predictions in V2	Number of correct predictions in V3
Top 10 %	6	6	6
Top 20%	11	13	19
Top 30%	25	31	50
Top 40%	421	435	663

Table 29 Occurrence time of predicted links for subset T8a

Percentage of predicted links being verified	Number of correct predictions in V1	Number of correct predictions in V2	Number of correct predictions in V3
Top 10 %	7	7	7
Top 20%	32	38	38
Top 30%	251	282	282
Top 40%	2 164	2 379	2 381

Table 30 Occurrence time of predicted link for subset T9a

Percentage of predicted links being verified	Number of correct predictions in V1	Number of correct predictions in V2	Number of correct predictions in V3
Top 10 %	156	489	490
Top 20%	348	1 056	1 058
Top 30%	863	1 975	1 980
Top 40%	902	2 140	2 149

The results from Table 28 - Table 30 are illustrated in Figure 23 Occurrence of predicted links, a percent stacked bar chart displaying the distribution of the predicted links within the testing subsets V1a – V3a aggregated by training subset and cut-off value. For example, it is observable that in the case of applying the link prediction algorithm to dataset T7a and setting the cut–off value to the top 20% of results, roughly 60% of the predicted links occur in the V1a testing set, 10% in the V2a testing set and 30% occur in the V3a testing set.

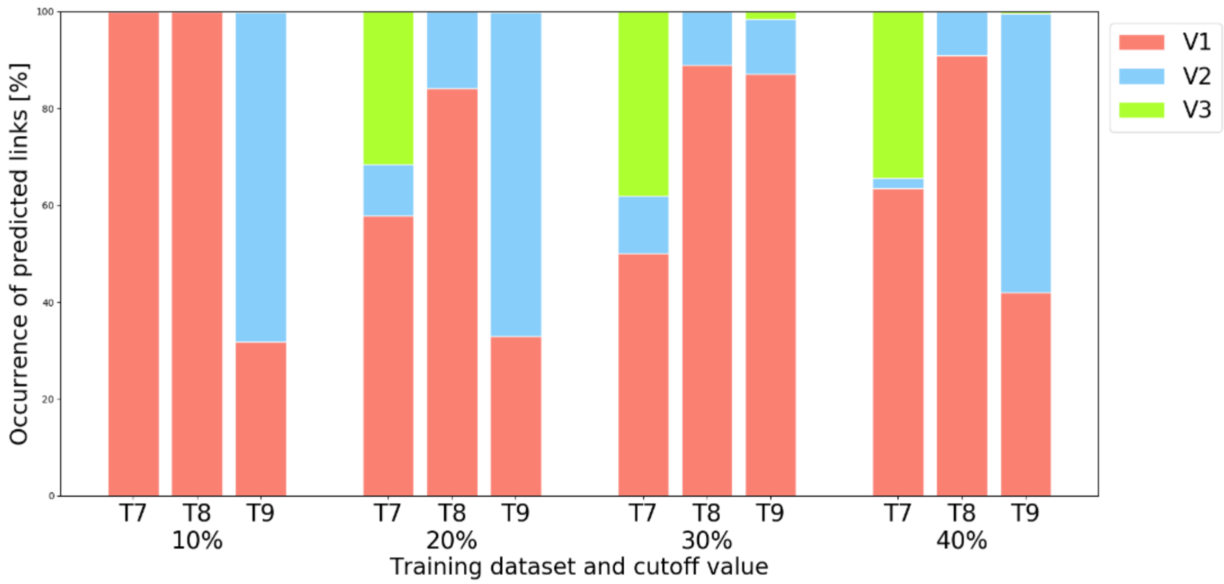

Figure 23 Occurrence of predicted links

6.5. Highlights of the results of the first empirical study

In this case study, the evolution of a mature technology domain is studied, and the ability to predict its future development is explored. The technologies maturity was determined based on a review of relevant literature and verified by talking with experts from the automotive field. The knowledge that the studied technology is in the mature stage of its life cycle allows for some presumptions based on the insights gathered from Chapter 2.

The first part of the empirical analysis consisted of a life cycle analysis of the studied technology domain. This was done by plotting a curve showing the cumulative number of applied patents over time. The result was an S–shaped curve which is consistent with the insights from the theoretical background, which states that a technology follows an S-curve during its life cycle. As the studied technology domain is in the mature stage of its life cycle, the fact that the plotted curve roughly follows an S-shape aligns with previous knowledge. It should be noted that because of the extremely large timespan over which the cumulative number of patents was plotted (118 years), there was a significant long tail in the plotted curve. This is primarily because the frequency of patenting activity was significantly lower at the time when the technology domain was in its inception. Consequently, the cumulative number of patent applications did not significantly grow for a considerable period in the technology's life cycle, which was noticeable in the visualisation of the cumulative number of patents over time (Figure 18). This was rectified by creating a second visualisation of the cumulative number of patents over time (Figure 19), with all of the patent applications before 1960 removed. The resulting curve presents a visualisation of the technology domain life cycle more relevant to this research and enables the identification of life cycle stages.

The method for analysing the growth dynamic of a patent citation network, introduced in Chapter 5.3, was applied to the patent citation network created from the retrieved patents. The results of the dynamic growth analysis were then compared with those from the technology life cycle analysis based on the cumulative number of patents. Superimposing the results from both analyses, several distinctions can be identified. First, the growth analysis of the citation network based on the studied technology domain consists of two discrete phases. In the first phase, the growth curve follows a generally positive trend, which is followed by the second phase, where the growth curve follows a negative trend (Figure 20). To reiterate, a positive growth trend in the dynamic growth analysis means the number of new patents introduced in the network is greater than the number of citations, which can be interpreted as the average number of citations per patent being very low. A negative growth trend in the dynamic growth analysis means the number of new edges surpasses the number of new nodes, meaning the average number of citations per patent is high. It is important to note that the transition period in the dynamic growth analysis graph, where the growth trend of the curve changes from positive to negative, corresponds to the start of the maturation phase identified based on the results of the technology life cycle analysis visualising the cumulative number of patents over time (Figure 21). Consequently, these results demonstrate the existence of a correlation between the dynamics of the life cycle stage of a technology domain and the dynamic of patent citation network growth.

The first half of a technologies life cycle covering the initiation and growth stage, designated in Figure 21 simply as “growth”, shows a trend where more new patents than citations are being introduced. Viewed in a technology evolution context, this means that the growth stage of a technologies life cycle consists of predominantly original and innovative inventions. This is not to say that these inventions are not based on previous work, simply that the patenting trend shows a higher level of innovations than the second part of the technologies life cycle. The second part of the technologies life cycle, consisting of the maturation and stagnation stage, shows an increase in the number of citations in relation to new inventions. In a technology evolution context, this means that the added inventions show a decreased level of innovation, being based in large part on prior inventions. These findings are consistent with previous research and the theoretical background related to qualitative insights regarding the evolution of technology which state that most technologies start as a series of radical inventions, showing a high degree of innovativeness, manifested in our case as a phase of a relatively low number of citations per patent. As the technology matures, it becomes less innovative and more an incremental improvement of existing inventions. This reduction of innovativeness is manifested as a relative increase in the number of citations per patent. These results provide insights into the nature of the dynamic growth of patent citations over a technologies life cycle as well as affirm the dynamic growth analysis of a citation network as a method for determining a technologies life cycle phase.

Following the results of the dynamic citation network growth analysis, the applicability of using link prediction algorithms to describe the growth of a patent co-citation network, and predict the future development of technology, was explored. The dataset containing patent citation data was segmented into two sets, a training set and a testing set. The time of the split was determined based on the results of the technology life cycle analysis. Both sets were segmented into smaller subsets and patent co-citation networks were created based on these subsets (Table 22 and Table 23). Then, four link prediction algorithms were applied to the subset of the training set close to the transition between the growth and maturation stages with the goal of identifying the link prediction algorithm most successful in predicting missing links i.e. missing patent co-citations. Out of the four applied link prediction algorithms, only the Preferential Attachment algorithm had any meaningful success in predicting missing links promise. This algorithm is a greedy algorithm following a “rich get richer intuition”, meaning nodes with a larger number of edges have a higher probability of gaining even more edges. In the context of this research, this would mean that patents being cited more often have a higher chance of being cited again. Based on the algorithm’s success, it was applied to the co-citation networks created from the other training subsets with the goal of providing additional insights into the dynamics of patent co-citation creation. Table 25 - Table 27, as well as Figure 22, show that newer patents have a greater chance of being co-cited following a rich-get-richer intuition. This build upon the work of Small et al. [158], who studied the citation dynamics of research papers and found that papers that are more recent, or have a higher number of citations, have a greater chance of being cited again. This quantitatively implies that new inventions often contain co-contributions from relatively recent knowledge, i.e. the knowledge flows from newer patents. Moreover, Table 28-Table 30, as well as Figure 23, demonstrate that most of the predicted co-citations occur in the near future. This is again consistent with the findings of Small (Small, 2005), who shows that highly cited papers have a tendency to be cited again in the near future.

7. EMPIRICAL STUDY – EMERGING TECHNOLOGY DOMAIN

This chapter presents the results of the second empirical study focused on exploring the evolution of an emerging technology. An overview of the technology is provided, as is the method for dataset creation. A life cycle analysis is conducted, and the results are compared to previous research. After the life cycle analysis, a link prediction algorithm is chosen and applied to a patent co – citation network with the goal of exploring the underlying intuition of network growth and predicting the future development of the technology.

7.1. Technology background – Analog Neural Networks

The second empirical study presented in this thesis explores the technology domain of Neuromorphic hardware. These are hardware systems that use electronic analogue circuits to mimic the neuro-biological architectures present in the nervous system. Analog Neural Networks implemented in VLSI (Very Large Scale Integration) systems is at present the only feasible alternative to modelling biological neural networks numerically [187]. Dedicated processing units emulate the behaviour of neurons directly in the hardware, and a web of physical interconnections allow for the rapid exchange of information.

While the earliest introduction of the concept of neuromorphic engineering is difficult to identify (source state ranges from the mid-1970s to late 1980s [188]), the first examples of a programmable neural array occurred in 2006, with the majority of advances happening in the last decade. The relatively recent introduction of this technology, as well as its potentially disruptive effect, makes it an example of disruptive technology as defined in Chapter 2.1.3. The rationale for considering this technology domain as emerging and disruptive is reinforced by the existing literature, which views this technology as disruptive [189][190] and white papers and reports [191] written by domain experts who also view this technology as potentially disruptive.

7.2. Dataset creation

The dataset creation step is analogous to the dataset creation process described in Chapter 6.2, as well as in the chapter describing the design of empirical research (Chapter 5.2). However, a few caveats should be mentioned. The relatively recent time of the technology domains introduction means it is represented by a significantly fewer number of patents compared to a mature technology. Furthermore, because of the emerging and disruptive nature of the technology domain, the nomenclature of the field is not yet standardized, meaning different terms are used to describe the same technology. Moreover, these terms are occasionally misused, describing technologies unrelated to the technology domain. This increased the difficulty of creating a representative dataset of relevant patents, consequently increasing the importance of conducting a manual examination of the retrieved patents.

Data Retrieval

The method based on the one outlined in Chapter 5.2 was used to retrieve patents relevant to the studied technology domain. The keyword pair “Neuromorphic Hardware” was used as the basis of the patent database search, followed by an analysis of the preliminary results and identification of the CPC codes most relevant to the field. Table 31 shows the most frequent CPC codes used to classify patents within the retrieved dataset and their definitions.

Table 31 The most frequent CPC codes for the retrieved dataset

CPC Code	Definition
GO6N3	Computer systems based on biological models.
G11C	Information storage based on relative movement between record carrier and transducer.

The retrieved patent dataset was manually filtered, patents not related to the technology domain being filtered out.

The dataset, after the filtering process, consists of 1000 patents. The oldest patent has the application year of 1989, while the youngest patent has the application year of 2019. Two tables are created from the retrieved patents, one containing all of the metadata related to the patents expect citations and the second one containing backward patent citations. This split of data is covered in Chapter 5.2 and illustrated in Figure 11.

7.3. Empirical study - Technology Life Cycle analysis

In this step, a life cycle analysis of the studied technology domain is conducted. As in Chapter 6.3, a technology life cycle analysis is conducted based on the cumulative number of patents applications over time, followed by a dynamic growth analysis of a patent citation network.

TLC analysis based on the cumulative number of patent applications

Figure 24 shows the results of the TLC analysis of the emerging technology based on the cumulative number of patents.

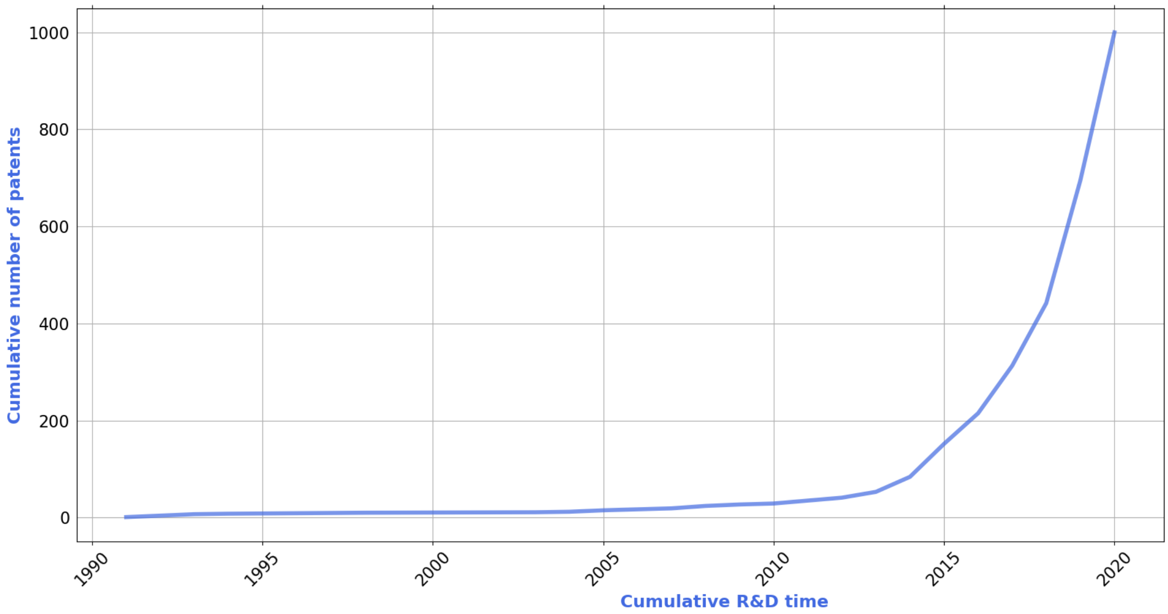

Figure 24 Curve of the Neuromorphic hardware technology domain based on the cumulative number of patents

It is immediately discernible that the TLC curve in Figure 24 does not follow the S-shaped pattern characteristic for mature technologies (Figure 19). In this study of an emerging technology domain, the TLC cure shows a period of almost non-existent growth followed by a sudden increase in the cumulative number of patents. This sudden exponential growth is characteristic for new technologies at the end of the emergence stage and growth stage of their life cycle [24] and reinforces the fact that the observed technology domain is emerging in nature. Unlike the chart in Figure 19 showing the TLC curve of a mature technology domain, Figure 24 does not show clear boundaries between life cycle stages.

Life cycle analysis based on the dynamic growth analysis

Using the method outlined in Chapter 5.3, a dynamic growth analysis is conducted on the patent citation network constructed based on the table containing patent citation data for the studied technology domain. The patent network at the end of the dynamic growth analysis consists of 7280 nodes and 525087 edges.

Figure 25 shows the results of the dynamic growth analysis of the patent citation network.

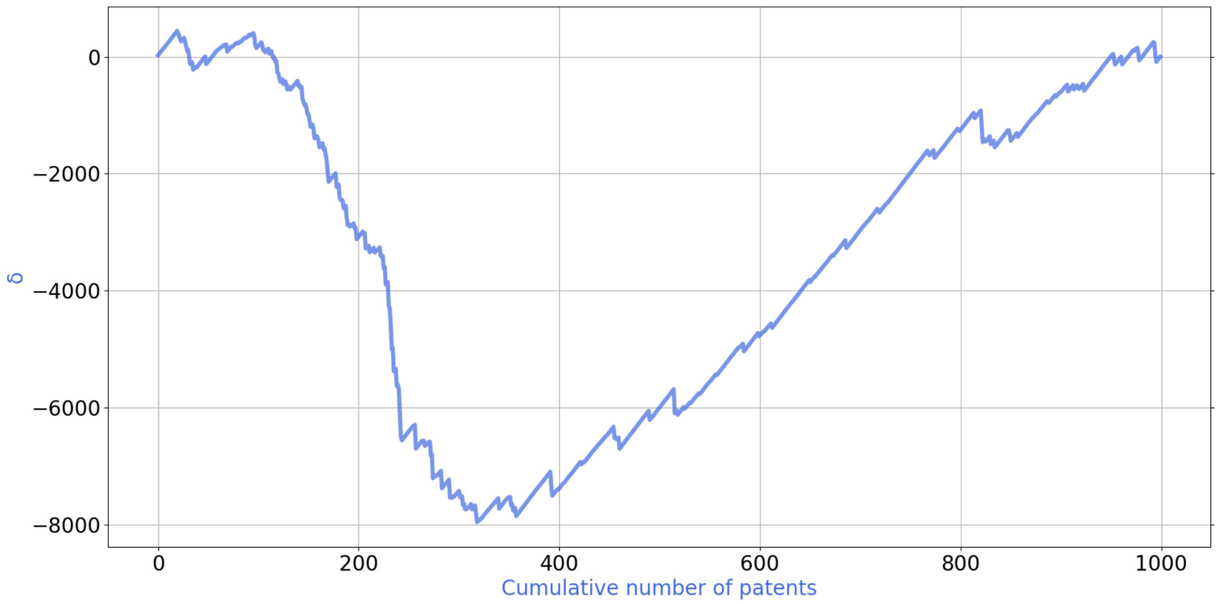

Figure 25 Technology life cycle curve of the examined technology domain based on the patent citation network growth analysis

Examining the results shown in Figure 25, it is worth noting that the curve starts with a noticeable negative trend, followed by a sharp positive trend. A negative trend in the curve means more edges than nodes are being added in this time period. In the context of this research, this means that new inventions in this time period cite a lot of prior inventions. However, an increase in the curve slope illustrates a period of high innovation, as a positive slope of the curve means more new nodes than edges are being added. In the context of this research, this means more inventions than citations are being added, signalling a period where there is an increase in original inventions, i.e. inventions that do not significantly build upon previous work.

Figure 26 shows the results of the S – curve-based life cycle analysis superimposed on the results of the growth analysis. The time identified in Figure 24 as the time at which the cumulative number of patents starts to rise exponentially is marked on the results of the dynamic growth analysis with the year corresponding to the time of the start of the growth.

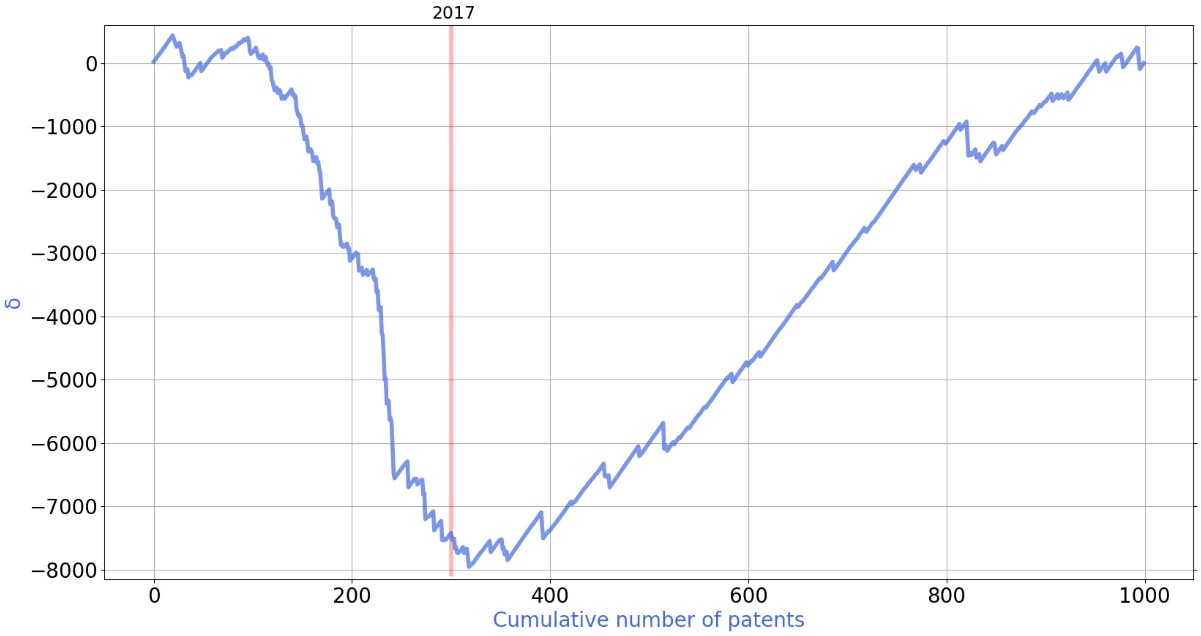

Figure 26 Growth analysis superimposed with results from the S-Curve analysis

Figure 26 provides context to the results of the dynamic growth analysis of the patent citation network. Superimposing Figure 24 and Figure 25 show us that the rapid increase in the cumulative number of patents corresponds with an increase in the ratio of new patents related to new citations. This could mean that this time marks the start of the growth phase of the emerging disruptive technology, which is in line with the theoretical background studying technology evolution presented in Chapter 2.1.3, covering disruptive evolution. There, the evolution of a disruptive technology is characterized by a rapid increase in the number of highly innovative inventions. The rapid increase of the number of inventions is demonstrated in Figure 24, while the high level of the inventiveness of the invention is demonstrated in Figure 25.

Contrasted with the results of the TLC analysis of the mature technology, outlined in Chapter 6.3, Figure 26 does not show the occurrence of key technologies within the technology domain, primarily because of the emerging and disruptive nature of the technology domain which makes the identification of key technologies impossible at this time, i.e. no key technologies are considered "key". Similarly, no life cycle stages are marked in Figure 26, as the sharp increase in the number of patent applications can be interpreted as both a part of the emergence stage as well as a transition between the emergence and growth stage.

7.4. Experimental study - Link prediction

The four link prediction algorithms outlined in Table 19 are applied to the training subsets as outlined in Chapter 5.4. Because the technology domain studied in this chapter is in its infancy, it has no discernible boundaries of technology life cycle stages and consists only of the emergence stage of a technology's life cycle. Consequently, unlike the previous study, presented in Chapter 6.4, which saw the splitting of the dataset into a single training and testing set, roughly at the transition from the growth stage to the stagnation stage, the results of the life cycle analysis shown in Figure 14 do not allow for the clear identification of life cycle stages, as is expected. Therefore, the segmentation of the dataset into a training and testing set is done in line with Figure 16. Nine training sets are created as well as nine testing sets, each training set increasing in size by 100 patents while each testing set decreasing in size by 100 patents. Consequently, the first training subset consists of the first 100 patents in the technology domain, with each subsequent subset increasing in size by an additional 100 patents (Table 32). Analogously, the first testing subset consists of the 900 patents from the technology domain not in the training subset, with each subsequent one decreasing in size by 100 (Table 33). Both the training subsets and testing subsets were coded for easier subsequent referencing (Table 32 and Table 33).

Table 32 Overview of training subsets with matching codes

Step of study	**0-100**	**0-200**	**0-300**	**0-400**	**0-500**	**0-600**	**0-700**	**0-800**	**0-900**
Code	T1b	T2b	T3b	T4b	T5b	T6b	T7b	T8b	T9b

Table 33 Overview of testing subsets with matching codes

STEP OF STUDY	**100-1000**	**200-1000**	**300-1000**	**400-1000**	**500-1000**	**600-1000**	**700-1000**	**800-1000**	**900-1000**
CODE	V1b	V2b	V3b	V4b	V5b	V6b	V7b	V8b	V9b

A patent citation network was created from each of the training and testing sets, and the patent citation networks were then converted into a patent co-citation network using the algorithm outlined in Chapter 5.4.

Application of link prediction algorithm

Each of the four link prediction algorithms was applied to the patent co-citation networks created from the training dataset subsets (T1b-T9b, Table 32), and the results were validated on the patent co-citation network created from the testing subsets (V1b-V9b, Table 33). The summary of the results of each analysis are presented in Figure 27 - Figure 30, showing how the precision of each of the applied link prediction algorithms changes the applied different training datasets and in relation to the change of the cut-off value. Because of the large number of analyses that were made, the tables detailing the results of each analysis aren't presented in this part of the thesis but are presented in Appendix A. Figure 27 - Figure 30 represent an overview of the tables presented in Appendix A.

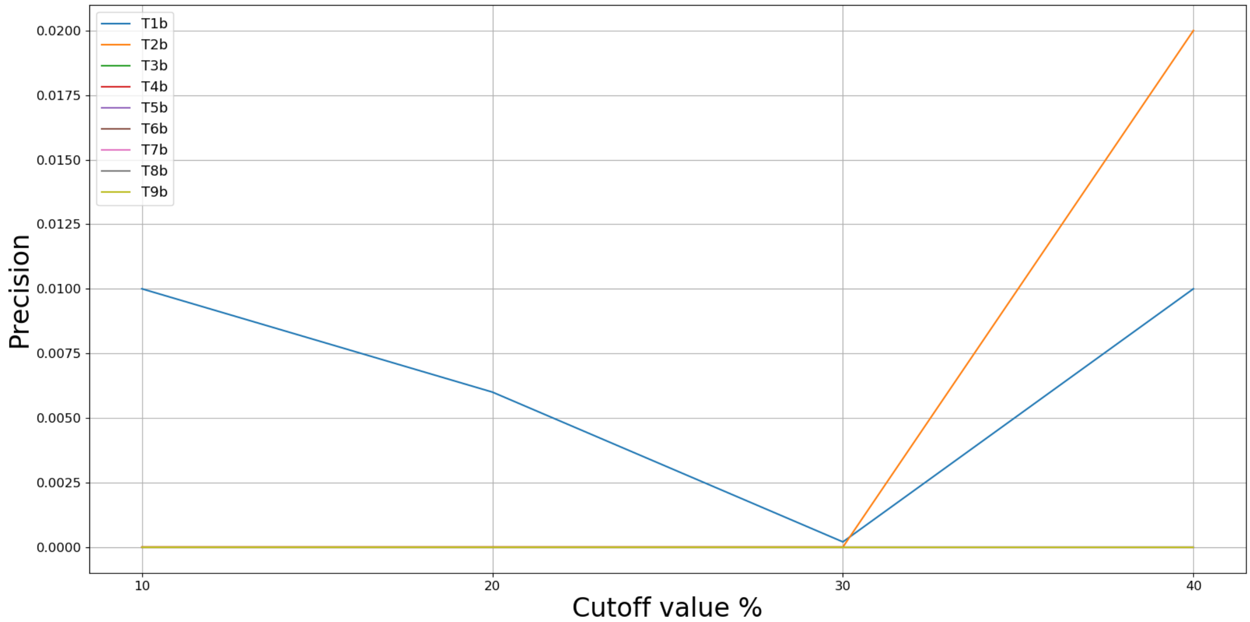

Figure 27 Relationship between the cutoff value, precision and training dataset size (preferential attachment)

The results presented in Figure 27 significantly differ from those of the first empirical study. While the first empirical study saw a pronounced success of the Preferential Attachment link prediction algorithm, the other three link prediction algorithms showing a negligible level of precision, the second empirical study shows a reversal of results. As is observable in Figure 27, in this empirical study, the Preferential Attachment link prediction algorithm showed a minuscule level of precision across all training sets and cut-off values, peaking at a precision of 0.02 when applied to training subset T2b, but consistently staying at 0 for most of the training subsets. A more detailed over of the results of analyses using the Preferential Attachment link prediction algorithm is provided in Appendix A, Table A-34-Table A-42.

Figure 28 shows the results of the Adamic/Adar index link prediction algorithm. Here it can be observed that the algorithm showed a drastic increase in performance compared to the preferential attachment link prediction algorithm, Adamic/Adar index showed a higher precision. However, the precision also peaked when applied to dataset T1b with a maximal precision of 0.69. A steady decrease in the precision of the algorithm can be observed correlated with the increase of the training set size, culminating in a precision of 0 for training sets T6b and larger. In this algorithm, the common neighbour of a node pair with fewer neighbours contributes more to the similarity score than one with a larger number of neighbours. A more detailed overview of the analyses results using the Adamic/Adar link prediction algorithm is provided in Appendix A, Table A-7 - Table A-15.

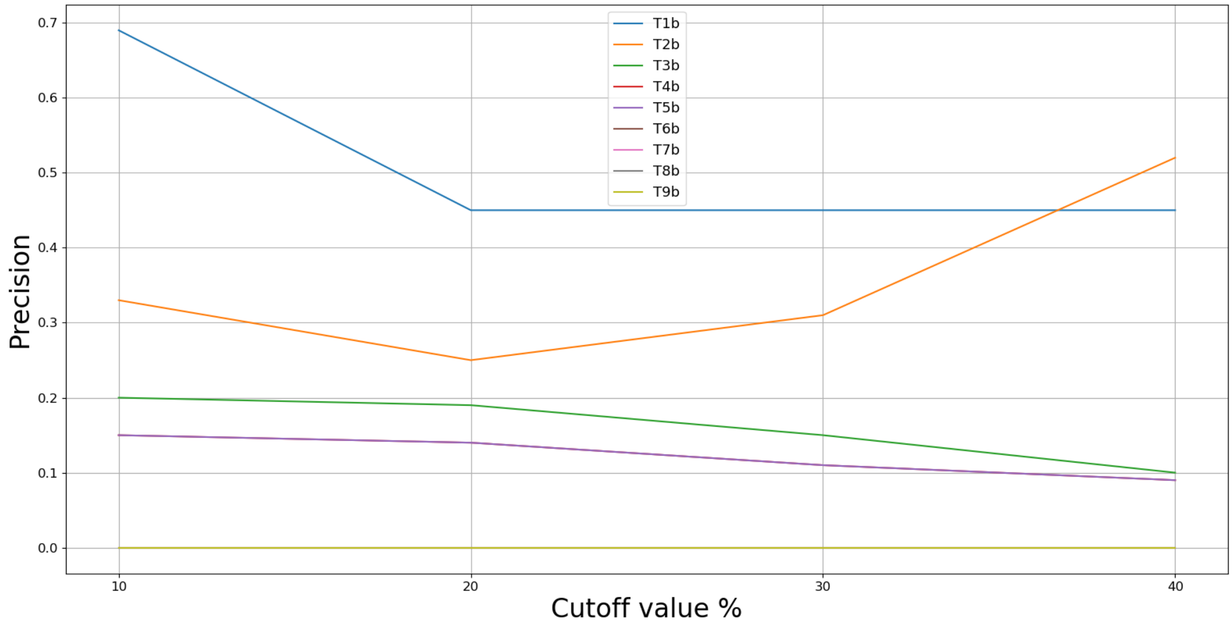

Figure 28 Relationship between the cutoff value, precision and training dataset size (Adamic/Adar index)

The performance of the Jaccard Coefficient link prediction algorithm is illustrated in Figure 29. Similarly, to the results of the Adamic/Adar index link prediction algorithm, the precision of this algorithm peaks at the subset T1b with a maximal precision of 0.48. Also, similarly to the results of Adamic/Adar, this precision decreases correlated to the increase in the size of the training set, staying below a precision of 0.1. Jaccard treats all neighbours of a node as a set and the prediction is done by computing and ranking the similarity of the neighbour set of each node. A more detailed overview of the analyses results using the Jaccard Coefficient link prediction algorithm is provided in Appendix A, Table A-16 - Table A-24.

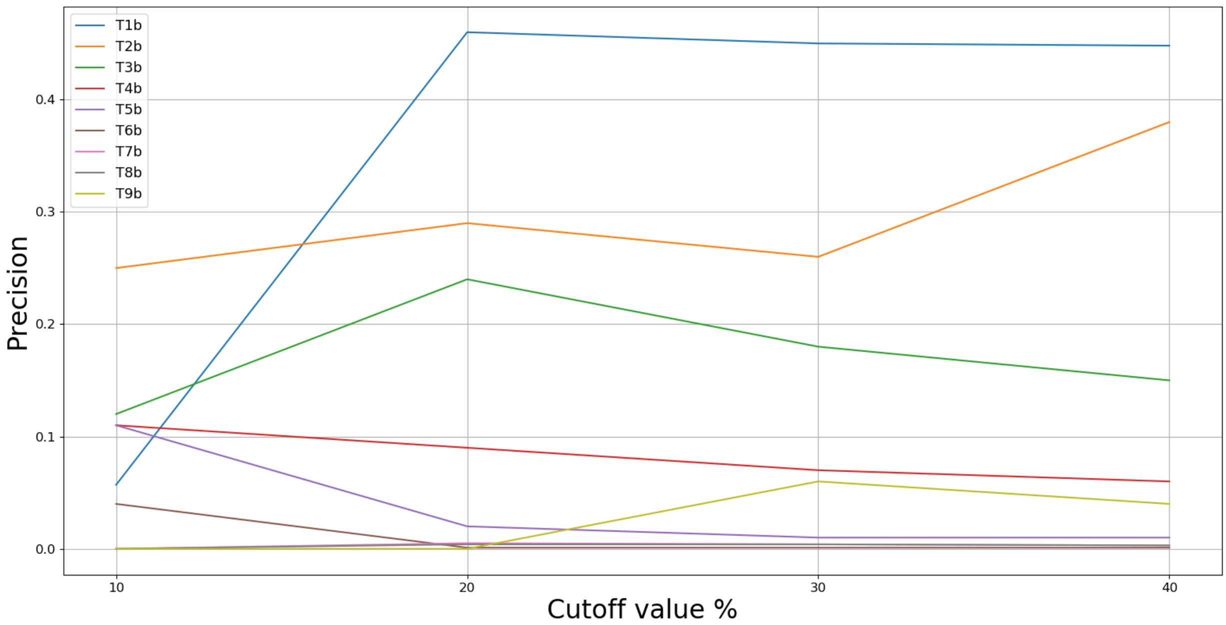

Figure 29 Relationship between the cutoff value, precision and training dataset size (Jaccard)

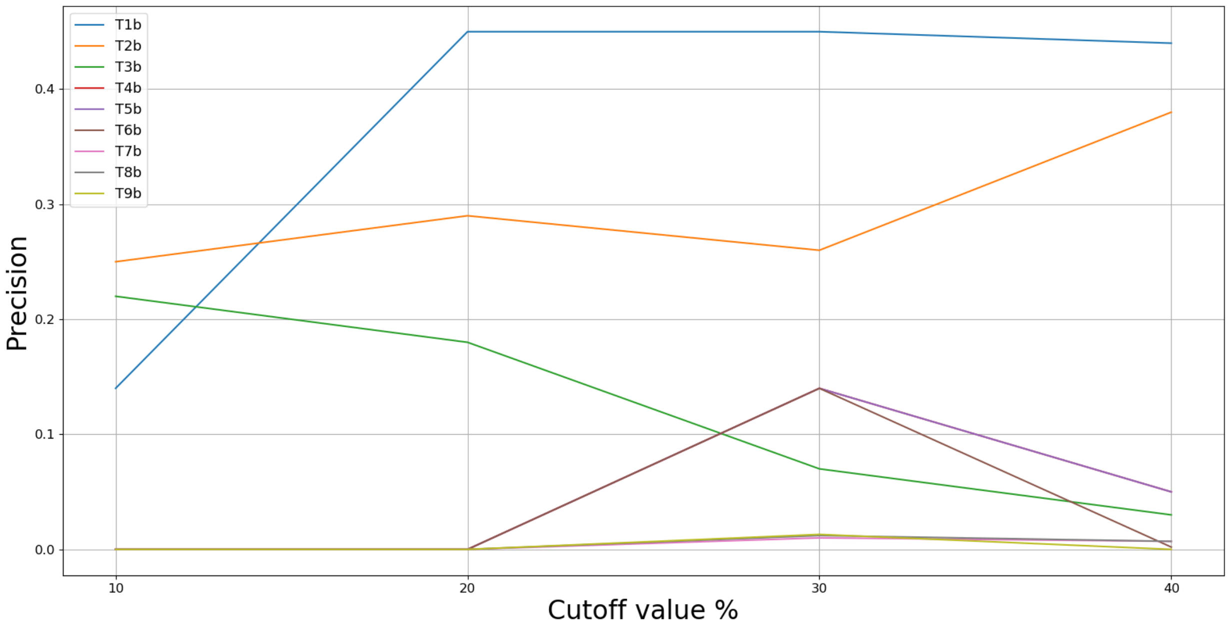

Figure 30 Relationship between the cutoff value, precision and training dataset size (RAI)

Finally, the precision of the Resource Allocation Index link prediction algorithm is presented in Figure 30. The Resource Allocation Index shows moderate precision, peaking when applied to dataset T1b with a maximal precision of 0.44, with the precision decreasing as the size of the training dataset increased. The Resource Allocation Index assumes that a node can send some resource to another node, with their common neighbours playing the roles of transmitters. Each transmitter has a unit of resource, and will averagely distribute it to all neighbours. A more detailed overview of the analyses results using the Resource Allocation Index link prediction algorithm is provided in Appendix A, Table A-34 - Table A-42.

Contrasted to Chapter 6.4, presenting the results of the empirical study of the mature technology, this study doesn't present an overview of the time frame in which the predicted links occur. This is because the extremely low number of successfully predicted links in this study, as well as the low precision of results, mean pursuing further analysis of the results would not contribute to this research in a meaningful way. Further implications of these results are covered in the discussion chapter of this thesis (Chapter 8).

7.5. Highlights of the results of the second empirical study

In this case study, the evolution of an emerging technology is studied. The technology is studied according to the research methodology outlined in Chapter 5, the same as the previous case study. However, some insights gained from the previous case study are applied in this one.

The first part of the empirical analysis consisted of a technology life cycle analysis based on the cumulative number of patents over time. The results of this analysis, shown in Figure 24, show a curve that starts out flat and then rises exponentially. These results are consistent with the theory describing both disruptive technologies and technologies in the emerging stage of their life cycle, which are characterized by a sharp increase in the number of introduced inventions. Moreover, it confirms that a technology life cycle analysis based on the cumulative number of patents can be used to identify technologies in the beginning of their life cycle.

The method for analysing the growth dynamic of a patent citation network, introduced in Chapter 5.3, was applied to the citation network created from the table containing patent citations. The results from the dynamic growth analysis were then compared to those from the technology life cycle analysis based on the cumulative number of patents. Superimposing the results from both analyses, several distinctions can be identified. First, the growth analysis of the citation network based on a disruptive technology follows a "u" shaped pattern, starting with a negative trend followed by a positive growth trend. In the context of this research, a negative trend means the number of citations is greater than the number of new patents, while a positive trend means the opposite. It is important to note that superimposing the start of the exponential growth of the cumulative number of patents with the results of the dynamic growth analysis shows that the shift from a negative growth trend to a positive growth trend roughly corresponds to the start of an exponential growth of the cumulative number of patents. This is a significant insight as it confirms the findings from the first case study, which identified that the growth phase of a technology domain is characterised by a higher ratio of newer inventions to citations. The initial negative growth rate in the dynamic growth analysis is explained by the fact that the first patents considered part of the technology domain still cite knowledge from incumbent technologies. Because the number of patents in this dataset is relatively small compared to that found in mature technologies, this phenomenon is easily identifiable. The technology starts becoming truly disruptive at the point where its cumulative number of patents starts to rise sharply and the ratio of inventions to citations increases.

As outlined in the research methodology chapter (Chapter 5.4), four link prediction algorithms were applied to the training subsets created from the dataset (Table 32). As is demonstrated in Figure 27 - Figure 29, none of the link prediction algorithms achieved the precision the Preferential Attachment link prediction algorithm achieved in the previous study. The Adamic/Adar index showed moderate precision when applied to subset T1b followed by the Resource Allocation index link prediction algorithm and the Jaccard Coefficient link prediction algorithm. The Preferential Attachment link prediction algorithm showed a negligible level of precision in this study. These results are not surprising given the unpredictable nature of disruptive technologies, as covered in Chapter 2.1.3. It should be noted that the Preferential Attachment link prediction algorithm showed low precision when applied to subsets containing patents from the growth phase of the technology life cycle in the first case study.

8. DISCUSSION

The eighth chapter of this thesis provides a discussion of the results of the two studies, contextualising them with the insights obtained from previous research. The goal of the discussion is to relate the findings with the review of literature and research questions, making an argument to confirm the thesis hypothesis and support the overall conclusion.

In this research, formalized records of invention are used as proxies for technologies in order to study the flow of knowledge within a technology domain and its correlation to the life cycle stages of a technology. This is done with the aim of expanding the theoretical fields of study related to the field of knowledge and technology management by gaining new insight into how a technology domain evolves, primarily by enhancing the understanding of how a technology's life cycle can be determined based on formalized records of invention, namely patents, and then by exploring the intuition governing the flow of knowledge elements within a technology domain. Moreover, another aim is to demonstrate that having a deeper understanding of the nature of the phenomenon enables researchers to explore not only the existing knowledge element flow within a technology domain but also predict the future flow of knowledge elements. In a managerial context, this would also reduce uncertainty for different decision makers and stakeholders, as determining the potential direction of the development of a technology domain, can support strategic and long–term planning of the development of products, processes and services as well as broader corporate strategy related to a company's knowledge portfolio

The research questions stated in Chapter 1 are answered and the following chapters will contextualize the results of the conducted empirical studies within each of the thesis research questions. Based on the research question's answers, the hypothesis's validity stated in Chapter 1 is discussed.

8.1. Research Question 1

Can the dynamics of patent citation creation be used to determine the life cycle stages of a technology domain?

Patent applications have a history of being used to determine the life cycle stages of a technology [168][23][24]. In this thesis, the relationship between the dynamics of patent citation creation and the life cycle stages of a technology domain is explored. An attempt is made to find a correlation between the results of a dynamic growth analysis of a citation network of patents representing a technology domain and the technology domains life cycle stages, i.e. if changes in the dynamics of patent citation network growth can be used to identify the life cycle stages of a technology domain. In order to answer the first research question, two empirical studies are conducted, the first one studying a mature technology domain, namely car headlights, and the second empirical study studying a disruptive technology domain, namely neuromorphic hardware.

The dynamic growth analysis of the patent citation network created from the mature technology domain shows three distinct phases (Figure 20). First, the growth curve follows a pronounced positive trend, followed by a period of stagnation from step 8000 to 9000, and finally a pronounced negative trend from step 9000 until the end of study. These results are interesting when superimposed with the results of the life cycle analysis conducted based on plotting the cumulative number of patents over time, one of the more common patent based methods for determining the life cycle stages of a technology used [24]. The visualization of the results based on this method follows an S-shaped curve, which fits with previous research studying the life cycle stages of a technology [24][168] [129] as well as *Presumption II* of the theoretical framework. Based on the results of previous research [24][151], the midpoint of the S – curve roughly represents the period where the technology transitions from the growth stage to the maturation stage. Superimposing the results of the dynamic growth analysis and the analysis based on the cumulative number of patents, it can be observed that the step of study where the midpoint of the S–curve occurs roughly corresponds to the step of study where the trend of growth in the dynamic growth analysis changes from positive to negative (Figure 21). Consequently, it can be stated that the transition of a technology domain from a growth phase to a maturation stage correlates with a drastic change in patent citation patterns.

These results are in line with the theoretical background and previous research. Focusing on the transition between growth and maturation, this period sees the transition from the introduction of predominantly key technologies to the introduction of predominantly pacing technologies [44]. The higher number of new patents in relation to the number of citations observed in the initiation and growth phases is in line with the knowledge that an invention introduced in these stages is more novel and innovative, focusing more on original inventions than those influenced by prior work [71][9]. However, the maturation and stagnation stages see a period in the technologies life cycle where true innovation becomes increasingly rare, and new inventions are mostly recombinations of existing inventions within the field or are simple incremental improvements of existing technology [64][65]. Consequently, observing the shape of the growth curve, a positive trend in the initiation and growth phase followed by a negative trend in the maturation and stagnation phases, confirms the intuition that patents in the first half of a technology domains life cycle contain fewer citations than those in the maturation and stagnation phase.

The results of the dynamic growth analysis of the patent citation network created from the emerging technology domain significantly differ from those created from the mature technology domain. It starts with a pronounced negative slope of the curve, bottoming out at around step 300 of the study. A reversal can then be observed, marked by a curve with a positive slope rising until the end step of the study (step 1000). The results of the life cycle analysis of the emerging technology based on the cumulative number of patents application also differs from those based on the mature technology. Unlike the results of the mature technology, which when visualized followed an S-shaped curve, the results of the emerging technology start with a barely noticeable growth of the curve whose slope starts increasing at roughly step 200 of the study and starts to exponentially rise at step 300. As is noted by Haupt et al. [168], this low number of patent applications is characteristic of the beginning of a new technologies life cycle, as the fundamental scientific and technological problems have yet to be resolved. The sudden increase in the cumulative number of patents is characteristic of the start of the growth stage of a technology domains life cycle [152]. While the results of both analyses significantly differ from the results based on the analysis of mature technology, superimposing the results of the dynamic growth analysis with that based on the cumulative number of patent applications demonstrates allows for an observance of a similar phenomenon as in the previous experiment (Figure 26). More specifically, superimposing the results of the two analyses shows the start of the growth phase, as identified in Figure 24, roughly corresponds with the change of curve slope from negative to positive in the dynamic growth analysis. Observing the dynamics of patent citation growth, a significant difference can be noticed when compared to the dynamics of the mature technology. While the mature technology sees a change in its citation trend at the transition from growth to maturation, here this transition happens at the start of the rapid increase in the number of the cumulative number of patent applications. Contextualized within the theory of technological evolution, this change of citation dynamics is intuitive and in line with the theoretical background [71][9]. The inventions occurring at the very beginning of a technology domain are probably a product of synthesizing knowledge from other technology domains, as is often the case in the introduction of new technologies [64], or they may start as radical changes of incumbent technologies which then veer off and become disruptive [74]. True innovation starts with the exponential rise of the number of inventions within a technology domain, signalling a period of increased creativity and inventiveness. Therefore, this period sees a reduction of the number of citations related to the number of new inventions, as these inventions are presumed to be truly innovative, or in this case, disruptive.

Still, it is a point of discussion as to why this change in the citation trend, occurring in the transition between the emergence and growth stages, is not observable in the growth curve exploring the mature technology domain. Several possible explanations can be offered. The most obvious one is that the extreme difference in the size of the datasets representing the two technology domains causes certain details visible in the analysis of the smaller dataset to be unnoticeable when analyzing the larger one. To be more precise, the dataset describing the emerging technology consists of 1000 patents, while the one describing the mature technology consists of 14 114 patents. Consequently, it is plausible to assume that the resolution of the growth curve describing the mature technology domain is simply too low to show slight changes in the patenting trend occurring at the start of the technology's life cycle. Another possible explanation is that the mature technology domain has such an early time of inception, and the technology so fundamental, that there were no cross-domain citations at its inception.

In either case, the results presented contribute to a clearer understanding of the relationship between the dynamics of patent citations and the life cycle stage of the technology domain by introducing a new method for studying the dynamic growth of a patent citation network. It is undoubtedly shown that a correlation exists between the life cycle stages of a technology domain and the dynamics of patent citation, providing an answer to Research Question 1. A summary of the discussion related to Research Question 1 is provided in Table 34.

Table 34 Summary of discussion related to RQ 1

Study	Implication	Limitation	Recommendation for future work
Mature technology domain	Patent citation growth chart consists of three stages: positive growth, stagnation and negative growth Change of chart slope correspond to change of life cycle stage from growth to maturation	Single study	Study a larger number of diverse technology domains to generalize results.
Emerging technology domain	Change of chart slope correspond to change of life cycle stage from initiation to growth	Unclear if an increase in the cumulative number of patents denotes a change in life cycle stage	Study a larger number of diverse technology domains to generalize results.

Some limitations should be noted. A generalization of these results is limited by the nature of the technologies being studied, a single technology from each category being studied (mature and emerging). Further research should focus on applying the methods presented in this thesis on a wide and diverse number of technologies from different technology fields and different life cycle stages. This would contribute to understanding the particularities of citation patterns of a more comprehensive array of technology and contribute to further understanding the relationship between citation patterns and technology life cycle stages.

8.2. Research Question 2

Can examining the occurrence of patent co-citations provide insight into patterns of knowledge flow within a technology domain?

As to provide an answer to research question 2, four link prediction algorithms were compared by testing their ability to predict missing links in patent co-citation networks created from patents representing a mature and an emerging technology domain. Since each link prediction algorithm uses a different underlying intuition when predicting the occurrence of new links within a network, identifying a link prediction algorithm that is precise in predicting new co-citations within a patent co-citation network consequently contributes to understanding the underlying intuition behind the growth of the network.

Examining the results of the empirical study presented in Chapter 6, studying the mature technology domain, the Preferential Attachment link prediction algorithm was shown to be the most precise out of the four link prediction algorithms that were compared. This algorithm follows a rich – get – richer intuition, meaning nodes with more neighbours have a higher chance of establishing new links. In the context of this research, it would mean that patents that are cited by a larger number of other patents have a greater chance of being cited again. This popularity bias means that highly influential inventions, i.e. inventions that have been recombined a large number of times to create new inventions have a greater chance of contributing to more new inventions, at least at the beginning of the maturation phase. This is in line with the intuitive notion that by the end of the growth phase, established technologies exist within a technology domain and the future development of that technology domain primarily consists of various recombination's of these technologies with incremental improvements [192][59][4]. It is also in line with previous research which studied the diffusion of knowledge using research papers as knowledge containers, finding that highly cited papers have a greater chance of being cited again [158]. Moreover, it confirms the results of Smojver et al. [167], who used a degree distribution analysis of a citation network, created from a mature technology domain, and found that this citation network grows following a popularity bias. The technologies influencing a larger number of subsequent inventions have a larger chance of influencing even more inventions.

Applied to an emerging technology domain, the same link prediction algorithms showed significantly different results. Out of the four link prediction algorithms, the Adamic/Adar index showed the highest level of precision, albeit that precision was moderate to low. This algorithm belongs to the "common neighbours" group of algorithms and follows the intuition that common elements with large neighbourhoods are less significant when predicting a connection between two nodes compared with elements shared between a small number of nodes. Presented more simply, common neighbours of low degrees are taken more seriously following the Adamic/Adar score. The Jaccard's Coefficient Index and the Resource Allocation index also showed low levels of precision, while the Preferential Attachment algorithm showed a negligible level of precision. Focusing on the results of the Adamic/Adar score in the context of this research, the success of this algorithm contributes a clearer understanding of the growth of a co-citation network based on an emerging disruptive technology implying that the likelihood of a patent being co-cited decreases if that patent has a larger number of existing co-citations as opposed to a patent with fewer co-citations. These differ significantly from the results of the study conducted on the mature technology. In fact, these results show that the growth of the co-citation network created from the emerging technology domain follows an intuition that is the opposite of the co-citation network created from the mature technology domain. While the results of the first empirical study imply that the chance of a patent being co–cited increases the more a patent has previously been co–cited, the second empirical study implies the reverse is true for an emerging technology at the beginning of its life cycle. The relative success of Adamic/Adar score implies that the chance of a patent being co-cited decreases as the number of prior co-citations increases, at least in the initial stages of a technology's life cycle. To further contextualize these results within the field of technology evolution, this would mean that new inventions, occurring within the growth stage of a technology domains life cycle, build upon less popular inventions. This insight is in line with the theory describing disruptive technologies as it confirms that these inventions contain a higher level of innovation [36][9], distancing themselves from established technologies, i.e. inventions with a higher number of co-citations, choosing instead inventions with a lower state of diffusion.

Discussing the results of the two empirical studies, it should be noted that while the most successful link prediction algorithm in the first empirical study showed a relatively high level of precision, the most successful link prediction algorithm in the second study showed only a moderate to low level of precision. This should be addressed in future research, as it should focus on applying both link prediction algorithm to a range of diverse technology domains and increase the number of positions within the technology domain life cycle where the link prediction algorithm will be applied. Nevertheless, the high precision of the algorithm in the first empirical study confirms presumption III, outlined in Chapter 4, as it shows that identifying the intuition based on which a co-citation network grows can be used to predict future co-citations.

This application of link prediction algorithms to a patent co-citation network is a novel approach to understanding how these networks grow. It differs from previous citation-based research which, in a technology evolution context, mainly focuses on constructing technology evolution trajectories or pathways [113][114][142] and changes in clusters over time [30]. Moreover, as opposed to prior work, which uses link prediction in the context of knowledge flow between different technology domains [15][126], this paper focuses on a single technology domain at a time and the flow of knowledge within that technology domain. Concerning research question 2, it is shown that an underlying intuition exists behind the growth of a patent co-citation network. This intuition is pronounced at the end of a technology domains growth phase, and it can be stated with a relatively high degree of certainty that the growth of a patent co-citation network follows a popularity bias at this stage. Moreover, it can be stated that the growth of a patent co-citation network at the start of a technology domain life cycle also follows a certain underlying intuition, namely a long tail bias. However, this is less pronounced and cannot be stated with the same confidence level as the first statement related to the end of the growth stage.

A summary of the discussion related to research question 2 is provided in Table 35.

Table 35 Summary of discussion related to RQ 2

Study	Implication	Limitation	Recommendation
Mature technology domain	Co-citations at the end of the growth cycle follow a popularity bias	None	Segment training dataset in a different way and apply link prediction algorithm
Emerging technology domain	Co–citations at the beginning of a technologies life cycle follow a log tail bias	Lower level of precision; questionable reliability of conclusion	Apply the Adamic/Adar algorithm to a number of diverse disruptive technologies

Similar to discussing the first research question, some limitations to the results related to the second research question should be noted. The limitation primarily focuses on the results of the second empirical study, focusing on the disruptive technology domain. As is demonstrated in Figure 27, while the Adamic/Adar link prediction algorithm did prove superior to the other three link prediction algorithms in describing the underlying intuition of co-citation network growth, it did so with a precision that is significantly lower than the results of the first study. While these results imply that an underlying intuition of co-citation network growth does exist, and are in line with insights provided in previous research, they should be confirmed by applying the Adamic/Adar link prediction algorithm in additional studies of disruptive technology domain.

8.3. Research Question 3

Is it possible to identify which parts of a technology's life cycle contribute the most to future inventions?

Based on the data outlined in Table 25- Table 27, presenting the results of the empirical study of the mature technology, it is demonstrated that a co-citation network created from patents at the end of the technology's growth phase and beginning of its maturation phase provides the most accurate results when the link prediction algorithm is applied at the beginning of the maturation phase (see Appendix for additional results). These results provide a new understanding of the dynamic of patent co-citations, as they show newer patents have a greater chance of being co-cited following a rich-get-richer intuition. This is in line with previous research studying the correlation of technology life cycle stages and the theory behind technology life cycle stages, finding that the transition between the growth and maturation phase signals the transition from pacing technologies to key technologies (Figure 7) [24], meaning pacing technologies, which appeared in the growth stage, are integrated into new products and processes becoming key technologies [24][148]. Moreover, the maturity phase primarily consists of incremental innovations [168], meaning most of the inventions created in this phase consists of minuscule improvements to previous inventions created in the growth phase. These results also positively correlate to the findings of Small [158], who studied co-citation patterns of research papers and found that the more recent, or highly cited, research papers have a greater chance of being cited again. This quantitatively implies that new inventions often contain co-contributions from relatively recent knowledge, i.e. the knowledge flows from newer patents.

Based on the data outlined in Figure 27- Figure 30 (and tables provided in Appendix), presenting the results of the experimental study of the emerging technology, it is demonstrated that the link prediction algorithms achieve the highest precision when applied to a co-citation network created from patents at the very start of the technologies life cycle (T1b), after which the precision of results starts to diminish. This is shown to be true for the Adamic/Adar index (Figure 28), Jaccard Coefficient (Figure 29), and the Resource Allocation Index (Figure 30). The Preferential Attachment algorithm showed a negligible level of precision (Figure 27). Consequently, it can be surmised that the earliest patents contribute the most to the precision of the link prediction algorithm for the emerging technology domain. In interpreting these results, it is crucial to reiterate that the link prediction algorithm, most successful in predicting missing links, gives preference to patents with fewer co-citations. The fact that this link prediction algorithm shows the highest precision when applied only at the beginning of the emerging technology domains life cycle is disappointing, as it implies the intuition governing co-citation creation after this period, meaning the start of the growth stage, cannot be predicted. This implication, however, does confirm the unpredictable nature inherent to disruptive inventions [73][74][71]. Moreover, it is in line with previous research, which found that the time of radical innovation, marking the beginning of a technologies life cycle, produces a small number of highly innovative inventions, consisting of novel combinations of existing technologies [168].To conclude, the results of the second empirical study imply that emerging technology domains at the start of their life cycle tend to be influenced by inventions that influenced a smaller number of new inventions in the past. However, the relatively low precision of the link prediction algorithm brings the reliability of these results into question.

Table 36 Summary of discussion related to RQ 3

Study	Implication	Limitation	Recommendation
Mature technology domain	At the end of the growth TLC stage, newer patents contribute more to future inventions than older ones	Single study	Generalise results by conducting multiple studies of different mature technologies
Emerging technology domain	At the beginning of a technologies life cycle, early patents contribute the most to future inventions	The Small size of testing datasets brings the validity of results into question	Conduct the same study of emerging technology domain containing a larger number of patents

The reliability of the results of the second study are impacted by the relatively low precision shown by the link prediction algorithm. Moreover, due to the recent time of the technologies introduction, the size of training datasets is comparatively small, further placing the validity of the results into question. While emerging disruptive inventions are, by their definition, unpredictable, the results presented in this study should be confirmed by applying the same link prediction algorithm to a diverse collection of emerging and disruptive technology domains.

8.4. Research Question 4

When are the predicted link, representing knowledge flow, created?

Exploring the results of the first empirical study (Table 28-Table 30, Figure 23), it is demonstrated that most of the predicted missing link, i.e. patent co-citations occur in the near future after the time when the link prediction algorithm was applied. Similarly to the results related to RQ 3, this period of the technology domains life cycle (the transition for the growth to the maturation stage) sees the transition from pacing to key technologies [24]. As this period marks the slowdown of innovation and the start of increasingly incremental improvements to existing technologies [168], these results confirm the findings of previous research that these incremental improvements closely track the original technology [168][148][3]. These results are also consistent with the findings of Small [158], who demonstrated that highly cited papers have a tendency to be cited again in the near future. Since both papers and patents represent knowledge containers, a confirmation of Small's findings contributes to understanding knowledge flows within a domain of knowledge. In a theoretical context, this provides insight into the dynamics of patent co-citations and, by proxy, of knowledge flow dynamics. It demonstrates the predictability of short-term knowledge flow based on recent patents. This is significant in a forecasting context as it shows the methodology is useful for short term predictions of knowledge flow into new inventions. This could prove a valuable resource in an industrial context, enabling stakeholders to gain insight into the potential short-term technology domain development.

Regarding the second empirical study, exploring the emerging technology domain, the previous part of the discussion has demonstrated that no relevant missing links were predicted, implying a very low reliability of the results of the link prediction. Consequently, contrasted to the study of the mature technology, no analysis of the time frame when the successfully predicted links occur was performed, as this would not contribute to the results of the thesis in a significant way.

Table 37 Summary of discussion related to RQ 4

Study	Implication	Limitation	Recommendation
Mature technology domain	The flow of knowledge element, following a rich-get-richer intuition, occurs in the near future after the application of the link prediction algorithm.	Single study	Generalise results by conducting multiple studies of different mature technologies
Emerging technology domain	Link prediction cannot be used to describe the intuition behind the flow on knowledge elements in an emerging technology	Number of successfully predicted link insufficient to conduct analysis	Conduct additional studies on emerging technologies

8.5. Overview of discussion and relation to the hypothesis

Based on the results of the two empirical studies, as well as the presented discussion related to the research questions, the following conclusions can be made. Based on these conclusions, the hypothesis stated in Chapter 1 will be validated. To reiterate, the thesis hypothesis states that:

The proposed research will verify the hypothesis that, based on the existing records of technical inventions, it is possible to model the dynamics of a technology domains development and gain insights into the potential future directions of technology development.

The discussion of the results related to RQ 1 indicates there is a correlation between the dynamics of patents citations creation and the life cycle stage of a technology domain manifested as a noticeable change in the relation of the number of new patents and the number of new citations. This change is apparent when viewing a graphical representation of the dynamic growth analysis of a patent citation network. In the case of a disruptive technology domain, the sudden increase of the cumulative number of patent applications correlates to a distinct change in the results of the dynamic growth analysis, observable as a reversal from a negative to a positive trend. In the case of a mature technology domain, the transition from the growth to the maturation stage correlates with a change in the results of a dynamic growth analysis, discernible by a sharp transition from a positive to a negative trend.

The discussion of the results related to RQ 2 - 4 contributes to the understanding of the dynamics of patent co-citation creation within a technology domain. The results of the first empirical study demonstrate that the dynamics of the growth of a patent co-citation network, created from patents representing a mature technology domain, follows similar dynamics to those of research paper co-citations [158], indicating knowledge flow follows a similar dynamic in both knowledge containers, namely that patents also have a preferential bias when forming co-citation. Recent patents, and those with a higher number of citations, have a greater chance of being cited again in the near future, at least in the transitional period from the growth stage to the maturation stage. This is an interesting finding as patents, by their nature, contain a different type of knowledge than research papers being proxies for technical invention. Therefore, this is an important contribution to the field of knowledge management as it implies that other methods used to study knowledge flow in paper co-citations might be applied to patent co-citations as well.

The distinction of this research related to prior work should be emphasised. While previous research used patents to explore the evolution of technology by constructing technology evolution trajectories and pathways [113][114][142] and changes in clusters over time [30], this research studies the dynamics of patent citation and co-citation creation, using a dynamic growth analysis to study the dynamics of patent citation and link prediction to gain insight into the dynamics of patent co-citation creation. Moreover, as opposed to prior work, which uses link prediction in the context of knowledge flow between different technology domains [15][126], this research focuses on a single technology domain and the flow of knowledge within it. The introduction of using a dynamic growth analysis of a patent citation network to determine the life cycle stage of a technology domain is a novel approach to studying the life cycle stages of a technology domain, compared to previous research, which primarily focused on using the S-curves to explore the life cycle stages of a technology domain [24]. Moreover, while the vast majority of methods studied only one indicator to determine the life cycle stage (number of patents, number of backward citations, number of classification codes…), this method simultaneously takes into account both the increase in the number of new patents as well as the increase of new citations. Segmenting the training and testing datasets into smaller subsets also provide a novel understanding of the dynamics of co-citation creation, i.e. the co-contribution of existing knowledge to future inventions. The splitting of the testing set into subsets differs from previous work studying the future development of technology, which often do not study the time frame in which the predicted technologies will occur [193][40][115]. The splitting of the training datasets into subsets is also distinct from prior work, such as the work of [15], who separated the training set into subsets but did not explicitly explore which part of the training dataset contributes the most to the predicted links or the work of [126], who created three subsets of the training dataset but did not explore their individual contribution to the link prediction algorithms precision. Works more focused on exploring evolution trajectories [113][114] do not create subsets at all or use roughly created subsets strictly to analyse trends.

As stated in Chapter 1, this research aims to contribute in both a theoretical and practical context. The first theoretical contribution comes from gaining a novel way of exploring the life cycle stages of a technology domain. To be more precise, a novel way of determining the life cycle stages of a technology domain is introduced based on the dynamic growth analysis of a patent citation network. This is noteworthy because it is based on a formalized and open source of knowledge, namely patents, making the method itself both open (being based on a publicly disclosed knowledge source) and easily repeatable (being based of a knowledge source that is structured in its presentation). While methods for determining the life cycle stages of a technology based on patent metadata do exist [40][23][24][168], the method presented in this thesis is novel as it is based on a dynamic analysis of citation network growth, providing a unique insight to researchers. Consequently, a contribution is made in understanding how a technology evolves in a theoretical context as well as in providing an open data-based tool to reduce uncertainty for stakeholders in an industrial context. This is in line with the first expected contribution stated in Chapter 1.3, i.e. the development of a model for quantifying the dynamics of evolution of technical invention and the implementation of technology. The second theoretical contribution comes from providing insight into how knowledge elements flow within a technology domain. By discovering patterns of knowledge diffusion, the underlying intuition of knowledge element flow can be deduced. Consequently, this can be used to predict future knowledge diffusion. Moreover, additional study of the dynamics of knowledge element flow within a technology domain can provide insight into which technologies from a technologies life cycle influence future inventions the most, as well as the approximate time frame of the occurrence of these future inventions. This is in line with the second expected contribution stated in Chapter 1.3 i.e. the development of a tool for simulating the potential direction of the development of a technology.

In a practical context, these results are applicable on both a micro and macro scale. On a micro scale, the methods presented in this research can be used in the ideation stage of product development. Applying the outlined link prediction method to a patent co-citation network created from the technology domain being researched may provide the user with insight into possible co-contributions of patents. This approach expands the research of Youn et al. [194], who viewed inventions as a combinatorial process and patents as carriers of technology. The predicted links within a patent co-citation network could then be considered possible combinations of existing technologies which might contribute to the invention process.

To conclude, the answers to the research questions presented in this chapter confirm the hypothesis stated in Chapter 1 and reiterated at the start of this chapter. It has been demonstrated that formalized records of invention, i.e. patents, can be used to model the dynamics of a technology domains development by providing insight into the dynamics of patents citation creation as well as the underlying intuition governing the creation of patent co-citations. This underlying intuition can be used to predict future patent co-citation, thereby providing insight into the potential future development of a technology domain.

9. CONCLUSION

This chapter presents an overview of the main findings as well as recommendations. Answers to the research questions are provided and their relation to the thesis aims and objectives are explored. The significance and implication of the findings is discussed as well as the contribution of the findings. Finally, some limitations of the study are presented, and recommendations for future research are made.

The research presented in this thesis aimed to create a novel methodology for studying a technology domains life cycle stages using patent citations as the primary source of data. Moreover, it aimed to identify previously undiscovered patterns of knowledge flow within a technology domain and exploring whether these patterns differ between different types of technology domains, namely those representing mature technologies and those representing disruptive technologies, and whether this intuition can be used to predict future knowledge flow within a technology domain. Finally, the dynamic of knowledge flow is studied by exploring which parts of the technology's life cycle contribute the most to the precision of the prediction methodology as well as when the predicted knowledge flows occur.

The results of the first research question show that there is a correlation between the dynamics of patent citation created and the life cycle stages of technology. Presuming that a technology follows a 4-stage life cycle, a study of both a mature technology domain and an emerging technology show that a change in the growth trend of the cumulative number of patents correlates with a change in the dynamics of patent citation growth, presented as a dynamic growth analysis of a patent citation network. Studying a mature technology, this correlation is prominent at the transition from the growth to the maturation stage of the TLC, characterised by a slowing down of the patenting activity in the technology domain. Observing the growth dynamics of patent citation creation sees a variation in the slope of the growth curve from positive to negative, signalling a reversal in the trend of citation activity. Specifically, this transition is made from a time where more patents than citations are introduced to a time where more citations than patents are introduced. Studying a disruptive technology, it is observed that the start of a rapid increase in the number of applied patents also correlates to a change in the growth dynamics of a patent citation network. However, in this case, a reversal of the growth curve is observed to be from negative to positive, signalling an increase in the number of new patents in relation to new citations. These results represent a significant contribution to the body of knowledge as they imply a correlation between the changes in the dynamic of patent citation creation within a technology domain and the transition between the life cycle stages of a technology domain. Consequently, this makes the dynamic growth analysis of a patent citation network representing a technology domain, a viable method for determining the technology domains life cycle stage.

The results of the second research question indicate that the underlying intuition governing the growth of a patent co-citation network can be modelled using the preferential attachment link prediction algorithm, at least when applied to a technology domain at the end of its growth stage and beginning of its maturation stage. However, a study of a disruptive technology domain resulted in less meaningful results, showing only moderate success when the Adamic/Adar Index, Jaccard Coefficient, and Resource Allocation index link prediction algorithms were applied to patent co-citation network representing a disruptive technology domain, and even then, only when applied to the smallest subset of the training dataset. Consequently, it can be deduced that while the underlying intuition of knowledge flow within a mature technology can be identified as following a preferential attachment intuition, no such claim can be made for a disruptive technology domain as the intuition governing the knowledge growth in a disruptive technology domain cannot be determined with a meaningful level of confidence, at least not using link prediction algorithms applied to a patent co-citation network. Nevertheless, the results of the first empirical study imply that the Preferential Attachment link prediction algorithm can be successfully applied to a patent co-citation network in order to predict future knowledge flow, at least when applied to a technology domain at the transition from the growth to the maturation stage of its life cycle. Considering that most technologies follow a standard four-phase life cycle, it is reasonable to expect these findings to be applicable to other technology domains. Furthermore, based on the answers to research questions 3 and 4, the research presented in this thesis contributes to the understanding of knowledge flow in the context of technology domain evolution. More precisely, it is demonstrated that, when applying the preferential attachment link prediction algorithm to a mature technology, a high level of precision can be achieved by taking into account only that patents that occur in a relatively short time before the end of the growth stage of the technology’s life cycle. These findings are significant because they confirm the intuitive claim that younger patents have a greater chance of influencing future technology than older one another contributions in the insight that most of the flow of knowledge occurs in a relatively short time period after the time of the technology domains life cycle where the Preferential Attachment link prediction algorithm was applied.

These findings are noteworthy as they are consistent with the results of previous research studying the dynamic of knowledge flow using scientific papers as knowledge artefacts, indicating that the same dynamic of knowledge flow might apply to the knowledge between other containers of knowledge alongside patents and research papers. The relatively short time between the time of the application of the link prediction algorithm and the occurrence of the predicted knowledge flow demonstrates the usefulness of these results in short term predictions and strategic planning.

Finally, the results of the research in this thesis have several potential applications in a managerial/practical context. The results of the dynamic growth analysis of a patent citation network should prove valuable to stakeholders at all levels whose decisions are influenced by the life cycle stages of a technology, especially those involved in making decisions on a strategic level. A change in the growth trend of the dynamic growth curve correlates with a change in the stage of a technologies life cycle and consequently might signal a change in the strategic relevance of the examined technology. Consequently, insight into the implementation potential of the examined technology is made. Addressing the results of examining the growth of the patent co-citation network, the implication is that a firm's short-term strategic planning should take into account the knowledge contained in the patents relevant to its field, reinforcing the notion that proper knowledge management is invaluable to firms aspiring to produce innovative products. By having insight into future potential co-citations of patents, a firm can leverage its existing patent portfolio or asses the acquisition value of patents or the companies owning them. Furthermore, the presented method might prove helpful in the product design process, facilitating exploratory innovative thinking by providing designers with potential combinations of knowledge previously not considered. Both of these contributions are seen as a potential starting point for further research, focusing on either the strategic application of the presented methods or their application as part of the design process.

9.1. Limitations and Future work

Certain limitations of the presented research should be noted. A limiting factor in this research was the creation of patent collections accurately representing technology domains. While the advantage of using patents as a source of data has been extensively covered in this thesis, the fact remains that there is no tool that is open for use and can effortlessly create representative collections of patents. Therefore, while this research used a modification of existing approaches to creating patent collections, as a creation of an automated tool was out of the scope of this research, future research should focus on creating an innovative and automated tool capable of creating collections of patents accurately representing individual technology domains in conjunction with existing open patent databases.

Moreover, this thesis analyses a single technology domain representing a mature technology and a single technology domain representing an emerging technology. Single case studies are limited in their applicability beyond their respective context and the research presented in this thesis is no exception. Even though two studies were conducted in the course of this thesis, each focuses on two radically different technologies. While the results of both studies are in line with previous research and the theoretical background describing both types of technologies, and while single case studies are not uncommon, future work should be expanded to cover multiple technology domains. The chosen technology domains should be diversified to test whether the results presented in this thesis can be generalised on a wide range of technology domains or are the presented results domain specific or specific to technology domains that share certain traits. In the latter case, the specific traits that make a technology domain viable for the application of the presented methodology should be identified. A specific limitation arises when examining the results of the second empirical study, studying the emerging and disruptive technology domain. While a dynamic growth analysis of the patent citation network did show promising results in line with previous research and the findings of the first empirical study, the application of a link prediction algorithm to the patent co-citation network showed a negligible level of success. While this could be due to the fact that the training sets created from patents representing the emerging technology domain were significantly smaller than those representing a mature technology, consequently not providing the link prediction algorithms with enough data, this is an unlikely culprit for the low accuracy. A more likely reason is the fact that emerging and disruptive technologies are by definition unpredictable, making the failure of the link prediction algorithms to achieve any meaningful precision expected. Future work should nevertheless venture to identify the intuition of co-citation growth in different instances of emerging and disruptive technologies.

While it is demonstrated that future co-citations can be predicted, at least in a mature technology domain, the research presented in this thesis provides no insight into the information contained in the predicted co-cited patents. Consequently, while the results provided in this research contribute to understanding the dynamics of knowledge flow, they do not provide insight into the content of the knowledge. Future work should focus on providing additional context to the predicted co–citations by analysing the content of the examined patents in order to gain a deeper understanding of the technical reason for the creation of the co-citation and examine the specific knowledge element being combined. These combined knowledge elements should be contextualised within the knowledge elements contained in the patent to which the co-cited patents contribute. This could be done by integrating the presented methodology for identifying potential future co-citations with a method that uses natural language processing to analyse the free text contained in a patent [195] or one that analyses the classification code contained in patents in order to gain additional insight into the technologies they represent [192].

Future work should also focus on implementing how unknown and unpredictable factors influence the development of technology, these unknown factors being unavoidable in societal design and can at best be speculated about. The human factor provides a consistent level of uncertainty in any prediction model as well as the influence of emerging disruptive technologies on collective societal characteristics and norms. Incorporating these unknown factors into a future iteration of the framework presented in this thesis should further improve its accuracy.

Finally, while the framework presented in this thesis uses patents as the primary source of data, future work should incorporate knowledge exchange facilitated by open innovation, as this paradigm seems to increasing its importance in both academic research and industrial applications [196]

10. REFERENCES

[1] Colomo-Palacios R, Sánchez-Gordón M, Arias-Aranda D. A critical review on blockchain assessment initiatives: A technology evolution viewpoint. J Softw Evol Process. 2020 Nov 30;32(11):1–11. https://doi.org/10.1002/smr.2272

[2] Pereira S de A, Quoniam L. Intellectual property and patent prospecting as a basis for knowledge and innovation – a study on mobile information technologies and virtual processes of communication and management. RAI Rev Adm e Inovação. Departamento de Administração, Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo – FEA/USP; 2017 Oct;14(4):301–10. https://doi.org/10.1016/j.rai.2017.07.006

[3] Trajtenberg M, Jaffe AB. Patents, Citations, and Innovations [Internet]. Patents, Citations, and Innovations. The MIT Press; 2002. https://doi.org/10.7551/mitpress/5263.001.0001

[4] Ringberg T, Reihlen M, Rydén P. The technology-mindset interactions: Leading to incremental, radical or revolutionary innovations. Ind Mark Manag. Elsevier; 2019 May;79(June):102–13. https://doi.org/10.1016/j.indmarman.2018.06.009

[5] Abramo G, D'Angelo CA, Di Costa F. The role of geographical proximity in knowledge diffusion, measured by citations to scientific literature. J Informetr. Elsevier Ltd; 2020 Feb;14(1):101010. https://doi.org/10.1016/j.joi.2020.101010

[6] Solée R V., Valverde S, Casals MR, Kauffman SA, Farmer D, Eldredge N. The evolutionary ecology of technological innovations. Complexity. 2013 Mar;18(4):15–27. https://doi.org/10.1002/cplx.21436

[7] Fakhar Manesh M, Pellegrini MM, Marzi G, Dabic M. Knowledge Management in the Fourth Industrial Revolution: Mapping the Literature and Scoping Future Avenues. IEEE Trans Eng Manag. 2021;68(1):289–300. https://doi.org/10.1109/TEM.2019.2963489

[8] Munir KA, Jones M. Discontinuity and after: The social dynamics of technology evolution and dominance. Organ Stud. 2004;25(4):561–81. https://doi.org/10.1177/0170840604040676

[9] Christensen CM. The innovator's dilemma. Harvard Business School Press, Boston. 1997. 1–179 p.

[10] Ribeiro LC, Kruss G, Britto G, Bernardes AT, da Motta e Albuquerque E. A methodology for unveiling global innovation networks: patent citations as clues to cross border knowledge flows. Scientometrics. 2014;101(1):61–83. https://doi.org/10.1007/s11192-014-1351-2

[11] Dabić M, Daim TU, Aralica Z, Bayraktaroglu AE. Exploring relationships among internationalization, choice for research and development approach and technology source and resulting innovation intensity: Case of a transition country Croatia. J High Technol Manag Res. 2012 Jan;23(1):15–25. https://doi.org/10.1016/j.hitech.2012.03.002

[12] Corsino M, Mariani M, Torrisi S. Firm strategic behavior and the measurement of knowledge flows with patent citations. Strateg Manag J. 2019;40(7):1040–69. https://doi.org/10.1002/smj.3016

[13] Brennecke J, Rank O. The firm's knowledge network and the transfer of advice among corporate inventors—A multilevel network study. Res Policy. Elsevier B.V.; 2017;46(4):768–83. https://doi.org/10.1016/j.respol.2017.02.002

[14] Park I, Yoon B. Technological opportunity discovery for technological convergence based on the prediction of technology knowledge flow in a citation network. J Informetr. Elsevier Ltd; 2018;12(4):1199–222. https://doi.org/10.1016/j.joi.2018.09.007

[15] Kim J, Magee CL. Dynamic Patterns of Knowledge Flows across Technological Domains: Empirical Results and Link Prediction. SSRN Electron J. 2018; https://doi.org/10.2139/ssrn.2990729

[16] Holahan PJ, Sullivan ZZ, Markham SK. Product development as core competence: How formal product development practices differ for radical, more innovative, and incremental product innovations. J Prod Innov Manag. 2014;31(2):329–45. https://doi.org/10.1111/jpim.12098

[17] Grimaldi M, Cricelli L, di Giovanni M, Rogo F. The patent portfolio value analysis: A new framework to leverage patent information for strategic technology planning. Technol Forecast Soc Change. Elsevier Inc.; 2015;94(1):286–302. https://doi.org/10.1016/j.techfore.2014.10.013

[18] Yoon SJ, Marhold K, Kang J. Linking the firm's knowledge network and subsequent exploratory innovation: a study based on semiconductor industry patent data. Innov

Manag Policy Pract. Routledge; 2017;19(4):463–82. https://doi.org/10.1080/14479338.2017.1358101

[19] Song XM, Montoya-Weiss MM. Critical Development Activities for Really New versus Incremental Products. J Prod Innov Manag. 1998;15(2):124–35. https://doi.org/10.1111/1540-5885.1520124

[20] Roper AT, Cunningham SW, Porter AL, Mason TW, Rossini FA, Banks J. FORECASTING AND MANAGEMENT Second Edition - Index. 2011.

[21] Nagy B, Farmer JD, Bui QM, Trancik JE. Statistical Basis for Predicting Technological Progress. PLoS One. 2013;8(2). https://doi.org/10.1371/journal.pone.0052669

[22] Sood A, James GM, Tellis GJ, Zhu J. Predicting the path of technological innovation: SAW vs. Moore, Bass, Gompertz, and Kryder. Mark Sci. 2012;31(6):964–79. https://doi.org/10.1287/mksc.1120.0739

[23] Chang SH, Fan CY. Identification of the technology life cycle of telematics a patent-based analytical perspective. Technol Forecast Soc Change. Elsevier Inc.; 2016;105:1–10. https://doi.org/10.1016/j.techfore.2016.01.023

[24] Gao L, Porter AL, Wang J, Fang S, Zhang X, Ma T, et al. Technology life cycle analysis method based on patent documents. Technol Forecast Soc Change. Elsevier Inc.; 2013;80(3):398–407. https://doi.org/10.1016/j.techfore.2012.10.003

[25] Basalla G. The evolution of technology. The Evolution of Technology. 2014. https://doi.org/10.1017/CBO9781107049864

[26] Iyer GR, LaPlaca PJ, Sharma A. Innovation and new product introductions in emerging markets: Strategic recommendations for the Indian market. Ind Mark Manag. 2006;35(3):373–82. https://doi.org/10.1016/j.indmarman.2005.02.007

[27] Anderson P, Tushman ML. Technological Discontinuities and Dominant Designs: A Cyclical Model of Technological Change. Adm Sci Q. 1990;35(4):604. https://doi.org/10.2307/2393511

[28] Hoisl K, Stelzer T, Biala S. Forecasting technological discontinuities in the ICT industry. Res Policy. Elsevier B.V.; 2015;44(2):522–32. https://doi.org/10.1016/j.respol.2014.10.004

[29] Abernathy W, Clark K. Innovation: Mapping the winds of creative destruction. Res Policy. 1993;22(2):102. https://doi.org/10.1016/0048-7333(93)90040-o

[30] Érdi P, Makovi K, Somogyvári Z, Strandburg K, Tobochnik J, Volf P, et al. Prediction of emerging technologies based on analysis of the US patent citation network. Scientometrics. 2013;95(1):225–42. https://doi.org/10.1007/s11192-012-0796-4

[31] Fleming L, Sorenson O. Technology as a complex adaptive system: Evidence from patent data. Res Policy. 2001;30(7):1019–39. https://doi.org/10.1016/S0048-7333(00)00135-9

[32] Amy Glasmeier. Technological discontinuities and flexible production networks: The case of Switzerland and the world watch industry. Res Policy. 1991;20(5):469–85.

[33] Habersang S, Küberling-Jost J, Reihlen M, Seckler C. A Process Perspective on Organizational Failure: A Qualitative Meta-Analysis. J Manag Stud. 2019;56(1):19–56. https://doi.org/10.1111/joms.12341

[34] Lamberg JA, Lubinaitė S, Ojala J, Tikkanen H. The curse of agility: The Nokia Corporation and the loss of market dominance in mobile phones, 2003–2013. Bus Hist. Routledge; 2021;63(4):574–605. https://doi.org/10.1080/00076791.2019.1593964

[35] Eriksson N. Conceptual study of a future drone detection system Countering a threat posed by a disruptive technology. 2018;

[36] Beninger S, Robson K. The disruptive potential of drones. Mark Lett. Marketing Letters; 2020;31(4):315–9. https://doi.org/10.1007/s11002-020-09542-8

[37] Frizzo-Barker J, Chow-White PA, Adams PR, Mentanko J, Ha D, Green S. Blockchain as a disruptive technology for business: A systematic review. Int J Inf Manage. Elsevier; 2020;51(April):0–1. https://doi.org/10.1016/j.ijinfomgt.2019.10.014

[38] Frizzo-Barker J, Chow-White PA, Adams PR, Mentanko J, Ha D, Green S. Blockchain as a disruptive technology for business: A systematic review. Int J Inf Manage. Elsevier; 2020;51(April):0–1. https://doi.org/10.1016/j.ijinfomgt.2019.10.014

[39] Giudice M Del, Peruta MR Della, Maggioni V. A model for the diffusion of knowledge sharing technologies inside private transport companies. J Knowl Manag. 2015;19(3):611–25. https://doi.org/10.1108/JKM-02-2015-0047

[40] Altuntas S, Dereli T, Kusiak A. Forecasting technology success based on patent data. Technol Forecast Soc Change. Elsevier Inc.; 2015;96:202–14. https://doi.org/10.1016/j.techfore.2015.03.011

[41] Schwarz JO, Ram C, Rohrbeck R. Combining scenario planning and business wargaming to better anticipate future competitive dynamics. Futures. Elsevier Ltd; 2019;105:133–42. https://doi.org/10.1016/j.futures.2018.10.001

[42] Choi JY, Jeong S, Kim K. A study on diffusion pattern of technology convergence: Patent analysis for Korea. Sustain. 2015;7(9):11546–69. https://doi.org/10.3390/su70911546

[43] Blessing LTM, Chakrabarti A. DRM, a design research methodology. DRM, a Design Research Methodology. 2009. https://doi.org/10.1007/978-1-84882-587-1

[44] Albert T. Measuring Technology Maturity [Internet]. Measuring Technology Maturity. Wiesbaden: Springer Fachmedien Wiesbaden; 2016. 1689–1699 p. https://doi.org/10.1007/978-3-658-12132-7

[45] Bunge M. Technology as Applied Science. Technol Cult. 1966;7(3):329. https://doi.org/10.2307/3101932

[46] Arthur WB. The structure of invention. Res Policy. 2007;36(2):274–87. https://doi.org/10.1016/j.respol.2006.11.005

[47] WIPO. Innovation and Intellectual Property [Internet]. 2022.

[48] Alstott J, Triulzi G, Yan B, Luo J. Mapping technology space by normalizing patent networks. Scientometrics. Springer Netherlands; 2017;110(1):443–79. https://doi.org/10.1007/s11192-016-2107-y

[49] Xu S, Wang X, Liu Z, Luan C. Network structural analysis of technology: A study from patent perspective. J Sci Technol Policy China. 2013;4(3):214–35. https://doi.org/10.1108/JSTPC-01-2013-0001

[50] Choi S, Jun S. Vacant technology forecasting using new Bayesian patent clustering. Technol Anal Strateg Manag. 2014;26(3):241–51. https://doi.org/10.1080/09537325.2013.850477

[51] Boyack KW, Wylie BN, Davidson GS. Domain visualization using VxInsight® for

science and technology management. J Am Soc Inf Sci Technol. 2002;53(9):764–74. https://doi.org/10.1002/asi.10066

[52] Zhang G, McAdams DA, Shankar V, Darani MM. Modeling the evolution of system technology performance when component and system technology performances interact: Commensalism and amensalism. Technol Forecast Soc Change. Elsevier; 2017;125(August):116–24. https://doi.org/10.1016/j.techfore.2017.08.004

[53] Devezas TC. Evolutionary theory of technological change: State-of-the-art and new approaches. Technol Forecast Soc Change. 2005;72(9):1137–52. https://doi.org/10.1016/j.techfore.2004.10.006

[54] Zhang G, McAdams DA, Shankar V, Darani MM. Technology evolution prediction using lotka-volterra equations. J Mech Des Trans ASME. 2018;140(6). https://doi.org/10.1115/1.4039448

[55] Tripsas M. Customer preference discontinuities: a trigger for radical technological change. Manag Decis Econ. 2008 Mar;29(2–3):79–97. https://doi.org/10.1002/mde.1389

[56] Altuntas S, Dereli T. A novel approach based on DEMATEL method and patent citation analysis for prioritizing a portfolio of investment projects. Expert Syst Appl. Elsevier Ltd; 2015;42(3):1003–12. https://doi.org/10.1016/j.eswa.2014.09.018

[57] Garud R, Rappa MA. A Socio-Cognitive Model of Technology Evolution: The Case of Cochlear Implants. Organ Sci. 1994;5(3):344–62. https://doi.org/10.1287/orsc.5.3.344

[58] Steurer J. The Delphi method: An efficient procedure to generate knowledge. Skeletal Radiol. 2011;40(8):959–61. https://doi.org/10.1007/s00256-011-1145-z

[59] Lee J, Berente N. The era of incremental change in the technology innovation life cycle: An analysis of the automotive emission control industry. Res Policy. 2013;42(8):1469–81. https://doi.org/10.1016/j.respol.2013.05.004

[60] Ehrnberg E, Jacobsson S. Indicators of discontinuous technological change: an exploratory study of two discontinuities in the machine tool industry. R D Manag. 1997;27(2):107–26. https://doi.org/10.1111/1467-9310.00048

[61] Abernathy W, Utterback J. Patterns of industrial innovation. Technol Rev. 1978;80(7):40–7.

[62] Utterback JM, Abernathy WJ. A dynamic model of process and product innovation. Omega. 1975;3(6):639–56. https://doi.org/10.1016/0305-0483(75)90068-7

[63] Hacklin F, Raurich V, Marxt C. How incremental innovation becomes disruptive: The case of technology convergence. IEEE Int Eng Manag Conf. 2004;1:32–6. https://doi.org/10.1109/iemc.2004.1407070

[64] Strumsky D, Lobo J, Leeuw S Van Der. Measuring the Relative Importance of Reusing, Recombining and Creating Technologies in the Process of Invention. 2011;

[65] Fleming L. Recombinant uncertainty in technological search. Manage Sci. 2001;47(1):117–32. https://doi.org/10.1287/mnsc.47.1.117.10671

[66] Tushman ML, Anderson P. Technological Discontinuities and Organizational Environments Philip Anderson. Adm Sci Q. 1986;31(3):439–65.

[67] Griffin A, Price RL, Vojak BA, Hoffman N. Serial Innovators' processes: How they overcome barriers to creating radical innovations. Ind Mark Manag. 2014;43(8):1362–71. https://doi.org/10.1016/j.indmarman.2014.08.010

[68] Lucas HC, Goh JM. Disruptive technology: How Kodak missed the digital photography revolution. J Strateg Inf Syst. Elsevier B.V.; 2009;18(1):46–55. https://doi.org/10.1016/j.jsis.2009.01.002

[69] Nemet GF. Demand-pull, technology-push, and government-led incentives for non-incremental technical change. Res Policy. 2009;38(5):700–9. https://doi.org/10.1016/j.respol.2009.01.004

[70] Funk JL. Components, Systems and Technological Discontinuities. Long Range Plann. Elsevier Ltd; 2008 Oct;41(5):555–73. https://doi.org/10.1016/j.lrp.2008.06.001

[71] Taaffe J. Disruptive innovation explored. Total Telecom. 2008;(FEB.):21.

[72] Hopp C, Antons D, Kaminski J, Oliver Salge T. Disruptive Innovation: Conceptual Foundations, Empirical Evidence, and Research Opportunities in the Digital Age. J Prod Innov Manag. 2018;35(3):446–57. https://doi.org/10.1111/jpim.12448

[73] Yu D, Hang CC. A Reflective Review of Disruptive Innovation Theory. Int J Manag Rev. 2010;12(4):435–52. https://doi.org/10.1111/j.1468-2370.2009.00272.x

[74] Dedehayir O, Nokelainen T, Mäkinen SJ. Disruptive innovations in complex product

systems industries: A case study. J Eng Technol Manag - JET-M. 2014;33:174–92. https://doi.org/10.1016/j.jengtecman.2014.06.002

[75] Abdel-Basset M, Chang V, Nabeeh NA. An intelligent framework using disruptive technologies for COVID-19 analysis. Technol Forecast Soc Change. Elsevier Inc.; 2021;163:120431. https://doi.org/10.1016/j.techfore.2020.120431

[76] Porter AL, Cunningham SW. Tech Mining: Exploiting New Technologies for Competitive Advantage. Tech Mining: Exploiting New Technologies for Competitive Advantage. 2005. 1–384 p. https://doi.org/10.1002/0471698466

[77] Mogee ME. Using Patent Data for Technology Analysis and Planning. Res Manag. 1991;34(4):43–9. https://doi.org/10.1080/08956308.1991.11670755

[78] Strumsky D, Lobo J, van der Leeuw S. Using patent technology codes to study technological change. Econ Innov New Technol. 2012;21(3):267–86. https://doi.org/10.1080/10438599.2011.578709

[79] Held P, Salotto J. Cooperative patent classification (CPC). 2011.

[80] Palumbo M. Commentary: Cooperative Patent Classification: a New Era for the World'S Intellectual Property Offices. Technol Innov. 2013;15(2):125–7. https://doi.org/10.3727/194982413x1350843068915

[81] Mueller SC, Sandner PG, Welpe IM. Monitoring innovation in electrochemical energy storage technologies: A patent-based approach. Appl Energy. Elsevier Ltd; 2015;137:537–44. https://doi.org/10.1016/j.apenergy.2014.06.082

[82] Kim J, Jeong B, Kim D. Patent Analytics. Patent Analytics. 2021. https://doi.org/10.1007/978-981-16-2930-3

[83] Risch J, Garda S, Krestel R. Hierarchical document classification as a sequence generation task. Proc ACM/IEEE Jt Conf Digit Libr. 2020;(April):147–55. https://doi.org/10.1145/3383583.3398538

[84] Fallah MH, Fishman E, Reilly RR. Forward patent citations as predictive measures for diffusion of emerging technologies. PICMET Portl Int Cent Manag Eng Technol Proc. 2009;420–7. https://doi.org/10.1109/PICMET.2009.5262201

[85] Duguet E, MacGarvie M. How well do patent citations measure flows of technology?

Evidence from French innovation surveys. Econ Innov New Technol. 2005;14(5):375–93. https://doi.org/10.1080/1043859042000307347

[86] Érdi P, Makovi K, Somogyvári Z, Strandburg K, Tobochnik J, Volf P, et al. Prediction of emerging technologies based on analysis of the US patent citation network. Scientometrics. 2013;95(1):225–42. https://doi.org/10.1007/s11192-012-0796-4

[87] Su HN, Lee PC. Dynamic and quantitative exploration on technology evolution mechanism: The case of electrical conducting polymer nanocomposite. PICMET Portl Int Cent Manag Eng Technol Proc. 2009;2433–40. https://doi.org/10.1109/PICMET.2009.5261834

[88] Watanabe C, Tsuji YS, Griffy-Brown C. Patent statistics: Deciphering a “real” versus a “pseudo” proxy of innovation. Technovation. 2001;21(12):783–90. https://doi.org/10.1016/S0166-4972(01)00025-6

[89] Phelps C, Heidl R, Wadhwa A. Knowledge, Networks, and Knowledge Networks: A Review and Research Agenda. Journal of Management. 2012. 1115–1166 p. https://doi.org/10.1177/0149206311432640

[90] Shibata N, Kajikawa Y, Takeda Y, Matsushima K. Comparative study on methods of detecting research fronts using different types of citation. J Am Soc Inf Sci Technol. 2009 Mar;60(3):571–80. https://doi.org/10.1002/asi.20994

[91] Kayal A. Measuring the pace of technological progress: Implications for technological forecasting. Technol Forecast Soc Change. 1999;60(3):237–45. https://doi.org/10.1016/S0040-1625(98)00030-4

[92] Efimenk I V., Khoroshevsky VF, Noyons ECM. Anticipating Future Innovation Pathways Through Large Data Analysis [Internet]. 2016. 360 p. https://doi.org/10.1007/978-3-319-39056-7

[93] Jeong Y, Yoon B. Development of patent roadmap based on technology roadmap by analyzing patterns of patent development. Technovation. Elsevier; 2015;39–40(1):37–52. https://doi.org/10.1016/j.technovation.2014.03.001

[94] Kim J, Lee S. Forecasting and identifying multi-technology convergence based on patent data: the case of IT and BT industries in 2020. Scientometrics. Springer Netherlands; 2017;111(1):47–65. https://doi.org/10.1007/s11192-017-2275-4

[95] Tseng F-M, Liu Y-L, Wu H-H. Market penetration among competitive innovation products: The case of the Smartphone Operating System. J Eng Technol Manag. Elsevier B.V.; 2014 Apr;32:40–59. https://doi.org/10.1016/j.jengtecman.2013.10.002

[96] Danneels E. Disruptive Technology Reconsidered: A Critique and Research Agenda. J Prod Innov Manag. 2004 Jul;21(4):246–58. https://doi.org/10.1111/j.0737-6782.2004.00076.x

[97] National Academy Of Sciences. Persistent forecasting of disruptive technologies - report 2. Persistent Forecasting of Disruptive Technologies - Report 2. 2010. 1–341 p. https://doi.org/10.17226/12834

[98] Fenn J, Raskino M. Understanding Gartner ' s Hype Cycles , 2009. Analysis. 2009;(July).

[99] O'Leary DE. Gartner's hype cycle and information system research issues. Int J Account Inf Syst. Elsevier Inc.; 2008;9(4):240–52. https://doi.org/10.1016/j.accinf.2008.09.001

[100] Steinert M, Leifer L. Scrutinizing Gartner ' s hype cycle approach Scrutinizing Gartner ' s Hype Cycle Approach. 2022;(October 2014).

[101] Campbell RS. Patent trends as a technological forecasting tool. World Pat Inf. 1983;5(3):137–43. https://doi.org/10.1016/0172-2190(83)90134-5

[102] Yoon B, Park Y. A text-mining-based patent network: Analytical tool for high-technology trend. J High Technol Manag Res. 2004;15(1):37–50. https://doi.org/10.1016/j.hitech.2003.09.003

[103] Hussain M, Tapinos E, Knight L. Scenario-driven roadmapping for technology foresight. Technol Forecast Soc Change. Elsevier; 2017;124(April):160–77. https://doi.org/10.1016/j.techfore.2017.05.005

[104] Chen K, Jarboe K, Wolfe J. Long-range scenario construction for technology assessment. Technol Forecast Soc Change. 1981;20(1):27–40. https://doi.org/10.1016/0040-1625(81)90039-1

[105] Trippe AJ. Patinformatics: Tasks to tools. World Pat Inf. 2003;25(3):211–21. https://doi.org/10.1016/S0172-2190(03)00079-6

[106] Hirschey M, Richardson VJ. Valuation effects of patent quality: A comparison for

Japanese and U.S. firms. Pacific Basin Financ J. 2001;9(1):65–82. https://doi.org/10.1016/S0927-538X(00)00038-X

[107] Narin F. Patent bibliometrics. Scientometrics. 1994;30(1):147–55. https://doi.org/10.1007/BF02017219

[108] Griliches Z. Patent Statistics as Economic Indicators: A Survey. J Econ Lit. 1990;28(4):1661–707. https://doi.org/10.1016/S0169-7218(10)02009-5

[109] Daim T, Lai KK, Yalcin H, Alsoubie F, Kumar V. Forecasting technological positioning through technology knowledge redundancy: Patent citation analysis of IoT, cybersecurity, and Blockchain. Technol Forecast Soc Change. Elsevier; 2020;161(October 2019):120329. https://doi.org/10.1016/j.techfore.2020.120329

[110] Shibata N, Kajikawa Y, Matsushima K. Topological analysis of citation networks to discover the future core articles. J Am Soc Inf Sci Technol. 2007 Apr;58(6):872–82. https://doi.org/10.1002/asi.20529

[111] Roepke S, Moehrle MG. Sequencing the evolution of technologies in a system-oriented way: The concept of technology-DNA. J Eng Technol Manag - JET-M. 2014;32:110–28. https://doi.org/10.1016/j.jengtecman.2013.08.005

[112] Ko N, Yoon J, Seo W. Analyzing interdisciplinarity of technology fusion using knowledge flows of patents. Expert Syst Appl. Elsevier Ltd; 2014;41(4 PART 2):1955–63. https://doi.org/10.1016/j.eswa.2013.08.091

[113] Fontana R, Nuvolari A, Verspagen B. Mapping technological trajectories as patent citation networks. An application to data communication standards. Econ Innov New Technol. 2009;18(4):311–36. https://doi.org/10.1080/10438590801969073

[114] Huang Y, Zhu F, Guo Y, Porter A, Zhu D. Identifying Technology Evolution Pathways Based on Patent Citation Network and Tech M Ining --Illustrated for Dye - Sensitized Solar Cells. Futur Technol Anal. 2014;Brussels(November):27–8. https://doi.org/10.13140/RG.2.1.3168.1125

[115] You H, Li M, Hipel KW, Jiang J, Ge B, Duan H. Development trend forecasting for coherent light generator technology based on patent citation network analysis. Scientometrics. Springer Netherlands; 2017;111(1):297–315. https://doi.org/10.1007/s11192-017-2252-y

[116] Kim E, Cho Y, Kim W. Dynamic patterns of technological convergence in printed electronics technologies: Patent citation network. Scientometrics. 2014;98(2):975–98. https://doi.org/10.1007/s11192-013-1104-7

[117] Lee PC, Su HN, Wu FS. Quantitative mapping of patented technology - The case of electrical conducting polymer nanocomposite. Technol Forecast Soc Change. Elsevier Inc.; 2010;77(3):466–78. https://doi.org/10.1016/j.techfore.2009.08.006

[118] Chen R, Park J, Smith A. Measuring Growth Potential of Communities in the Patent Citation Network. 2014;1–11.

[119] Takano Y, Mejia C, Kajikawa Y. Unconnected component inclusion technique for patent network analysis: Case study of Internet of Things-related technologies. J Informetr. Elsevier Ltd; 2016;10(4):967–80. https://doi.org/10.1016/j.joi.2016.05.004

[120] Ha SH, Liu W, Cho H, Kim SH. Technological advances in the fuel cell vehicle: Patent portfolio management. Technol Forecast Soc Change. Elsevier Inc.; 2015;100:277–89. https://doi.org/10.1016/j.techfore.2015.07.016

[121] Chang SBSM, Lai KK, Chang SBSM. Exploring technology diffusion and classification of business methods: Using the patent citation network. Technol Forecast Soc Change. Elsevier Inc.; 2009;76(1):107–17. https://doi.org/10.1016/j.techfore.2008.03.014

[122] Hung SW, Wang AP. Examining the small world phenomenon in the patent citation network: A case study of the radio frequency identification (RFID) network. Scientometrics. 2010;82(1):121–34. https://doi.org/10.1007/s11192-009-0032-z

[123] Marco AC. The dynamics of patent citations. Econ Lett. 2007;94(2):290–6. https://doi.org/10.1016/j.econlet.2006.08.014

[124] Choi J, Hwang YS. Patent keyword network analysis for improving technology development efficiency. Technol Forecast Soc Change. Elsevier Inc.; 2014;83(1):170–82. https://doi.org/10.1016/j.techfore.2013.07.004

[125] Kim H, Hong S, Kwon O, Lee C. Concentric diversification based on technological capabilities: Link analysis of products and technologies. Technol Forecast Soc Change. Elsevier Inc.; 2017;118:246–57. https://doi.org/10.1016/j.techfore.2017.02.025

[126] Lee WS, Han EJ, Sohn SY. Predicting the pattern of technology convergence using big-data technology on large-scale triadic patents. Technol Forecast Soc Change. Elsevier

Inc.; 2015;100:317–29. https://doi.org/10.1016/j.techfore.2015.07.022

[127] Feng X, Fuhai L. Patent text mining and informetric-based patent technology morphological analysis: An empirical study. Technol Anal Strateg Manag. 2012;24(5):467–79. https://doi.org/10.1080/09537325.2012.674669

[128] Suzuki S. Introduction to Patent Map Analysis. Japan Pat Off. 2011;

[129] Trappey C V., Wu HY, Taghaboni-Dutta F, Trappey AJC. Using patent data for technology forecasting: China RFID patent analysis. Adv Eng Informatics. Elsevier Ltd; 2011;25(1):53–64. https://doi.org/10.1016/j.aei.2010.05.007

[130] Lee S, Yoon B, Park Y. An approach to discovering new technology opportunities: Keyword-based patent map approach. Technovation. 2009;29(6–7):481–97. https://doi.org/10.1016/j.technovation.2008.10.006

[131] Yoon J, Park H, Kim K. Identifying technological competition trends for R&D planning using dynamic patent maps: SAO-based content analysis. Scientometrics. 2013;94(1):313–31. https://doi.org/10.1007/s11192-012-0830-6

[132] Jeong S, Yoon B. A systemic approach to exploring an essential patent linking standard and patent maps: Application of generative topographic mapping (GTM). EMJ - Eng Manag J. 2013;25(1):48–57. https://doi.org/10.1080/10429247.2013.11431965

[133] Son C, Suh Y, Jeon J, Park Y. Development of a GTM-based patent map for identifying patent vacuums. Expert Syst Appl. Elsevier Ltd; 2012;39(3):2489–500. https://doi.org/10.1016/j.eswa.2011.08.101

[134] Joung J, Kim K. Monitoring emerging technologies for technology planning using technical keyword based analysis from patent data. Technol Forecast Soc Change. Elsevier Inc.; 2017;114:281–92. https://doi.org/10.1016/j.techfore.2016.08.020

[135] Suh JH, Park SC. Service-oriented Technology Roadmap (SoTRM) using patent map for R&D strategy of service industry. Expert Syst Appl. Elsevier Ltd; 2009;36(3 PART 2):6754–72. https://doi.org/10.1016/j.eswa.2008.08.062

[136] Chen SH, Huang MH, Chen DZ. Identifying and visualizing technology evolution: A case study of smart grid technology. Technol Forecast Soc Change. Elsevier Inc.; 2012;79(6):1099–110. https://doi.org/10.1016/j.techfore.2011.12.011

[137] Kim YG, Suh JH, Park SC. Visualization of patent analysis for emerging technology. Expert Syst Appl. 2008;34(3):1804–12. https://doi.org/10.1016/j.eswa.2007.01.033

[138] Yoon BU, Yoon CB, Park YT. On the development and application of a self-organizing feature map-based patent map. R D Manag. 2002;32(4):291–300. https://doi.org/10.1111/1467-9310.00261

[139] Huang Y, Zhu D, Qian Y, Zhang Y, Porter AL, Liu Y, et al. A hybrid method to trace technology evolution pathways: a case study of 3D printing. Scientometrics. Springer Netherlands; 2017;111(1):185–204. https://doi.org/10.1007/s11192-017-2271-8

[140] Zhang L, Li L, Li T, Zhang Q. PatentLine: Analyzing technology evolution on multi-view patent graphs. SIGIR 2014 - Proc 37th Int ACM SIGIR Conf Res Dev Inf Retr. 2014;1095–8. https://doi.org/10.1145/2600428.2609518

[141] Wu JL, Chang PC, Tsao CC, Fan CY. A patent quality analysis and classification system using self-organizing maps with support vector machine. Appl Soft Comput J. Elsevier B.V.; 2016;41:305–16. https://doi.org/10.1016/j.asoc.2016.01.020

[142] Kim M, Park Y, Yoon J. Generating patent development maps for technology monitoring using semantic patent-topic analysis. Comput Ind Eng. Elsevier Ltd; 2016;98:289–99. https://doi.org/10.1016/j.cie.2016.06.006

[143] Yang SY, Soo VW. Comparing the conceptual graphs extracted from patent claims. Proc - IEEE Int Conf Sens Networks, Ubiquitous, Trust Comput. 2008;(1):394–9. https://doi.org/10.1109/SUTC.2008.87

[144] Yang SY, Soo VW. Extract conceptual graphs from plain texts in patent claims. Eng Appl Artif Intell. Elsevier; 2012;25(4):874–87. https://doi.org/10.1016/j.engappai.2011.11.006

[145] Tseng YH, Lin CJ, Lin YI. Text mining techniques for patent analysis. Inf Process Manag. 2007;43(5):1216–47. https://doi.org/10.1016/j.ipm.2006.11.011

[146] Daim TU, Rueda G, Martin H, Gerdsri P. Forecasting emerging technologies: Use of bibliometrics and patent analysis. Technol Forecast Soc Change. 2006;73(8):981–1012. https://doi.org/10.1016/j.techfore.2006.04.004

[147] Yoon B, Lee S. Applicability of patent information in technological forecasting: A sector-specific approach. J Intellect Prop Rights. 2012;17(1):37–45.

[148] Albert T. Measuring Technology Maturity [Internet]. Measuring Technology Maturity. Wiesbaden: Springer Fachmedien Wiesbaden; 2016. 1689–1699 p. https://doi.org/10.1007/978-3-658-12132-7

[149] Fleming L, Sorenson O. Technology as a complex adaptive system: Evidence from patent data. Res Policy. 2001;30(7):1019–39. https://doi.org/10.1016/S0048-7333(00)00135-9

[150] Temel S, Dabić M, Murat Ar I, Howells J, Ali Mert, Yesilay RB. Exploring the relationship between university innovation intermediaries and patenting performance. Technol Soc. 2021 Aug;66:101665. https://doi.org/10.1016/j.techsoc.2021.101665

[151] Ernst H. The Use of Patent Data for Technological Forecasting: The Diffusion of CNC-Technology in the Machine Tool Industry. Small Bus Econ. 1997;9(4):361–81. https://doi.org/10.1023/A:1007921808138

[152] Liu CY, Wang JC. Forecasting the development of the biped robot walking technique in Japan through S-curve model analysis. Scientometrics. 2010;82(1):21–36. https://doi.org/10.1007/s11192-009-0055-5

[153] Smojver V, Štorga M, Potočki E. An extended methodology for the assessment of technical invention evolution. In: Marjanović, Dorian ; Štorga, Mario ; Pavković, Neven ; Bojčetić, Nenad ; Škec S, editor. Proceedings of International Design Conference, DESIGN. Design Society; 2016. p. 1135–44.

[154] Shields LBE, Gertz TA, Wilson KC, Figg GL, Hester ST, Honaker JT. The S-curve discontinuity theory predicts the path towards a "well" society and increased longevity. Med Hypotheses. Elsevier; 2018;121(June):99–102. https://doi.org/10.1016/j.mehy.2018.09.006

[155] Floyd C. Managing Technology for Corporate Success. Gower; 1997. 261 p.

[156] Murata Y, Nakajima R, Okamoto R, Tamura R. Localized knowledge spillovers and patent citations: A distance-based approach. Rev Econ Stat. 2014 Dec;96(5):967–85. https://doi.org/10.1162/REST_a_00422

[157] Mas-Tur A, Roig-Tierno N, Sarin S, Haon C, Sego T, Belkhouja M, et al. Co-citation, bibliographic coupling and leading authors, institutions and countries in the 50 years of Technological Forecasting and Social Change. Technol Forecast Soc Change. 2021

Apr;165:120487. https://doi.org/10.1016/j.techfore.2020.120487

[158] Small H. Tracking and predicting growth areas in science. Proc ISSI 2005 10th Int Conf Int Soc Sci Inf. 2005;1(3):13–23. https://doi.org/10.1007/s11192-006-0132-y

[159] Braam RR, Moed HF, van Raan AFJ. Mapping of science by combined co-citation and word analysis. I. Structural aspects. J Am Soc Inf Sci. 1991;42(4):233–51. https://doi.org/10.1002/(SICI)1097-4571(199105)42:4<233::AID-ASI1>3.0.CO;2-I

[160] Adams J. Early citation counts correlate with accumulated impact. Scientometrics. 2005;63(3):567–81. https://doi.org/10.1007/s11192-005-0228-9

[161] Hummon NP, Dereian P. Connectivity in a citation network: The development of DNA theory. Soc Networks. 1989;11(1):39–63. https://doi.org/10.1016/0378-8733(89)90017-8

[162] Ryo M, Jeschke JM, Rillig MC, Heger T. Machine learning with the hierarchy-of-hypotheses (HoH) approach discovers novel pattern in studies on biological invasions. Res Synth Methods. 2020;11(1):66–73. https://doi.org/10.1002/jrsm.1363

[163] Benson CL, Magee CL. A hybrid keyword and patent class methodology for selecting relevant sets of patents for a technological field. Scientometrics. 2013;96(1):69–82. https://doi.org/10.1007/s11192-012-0930-3

[164] Xie Z, Miyazaki K. Evaluating the effectiveness of keyword search strategy for patent identification. World Pat Inf. Elsevier Ltd; 2013;35(1):20–30. https://doi.org/10.1016/j.wpi.2012.10.005

[165] lens.org [Internet].

[166] Google Patents [Internet].

[167] Smojver V, Štorga M, Potočki E. Determining the life cycle phase of a technology based on patent data. Teh Vjesn. 2019; https://doi.org/10.17559/TV-20171123182106

[168] Haupt R, Kloyer M, Lange M. Patent indicators for the technology life cycle development. Res Policy. 2007;36(3):387–98. https://doi.org/10.1016/j.respol.2006.12.004

[169] Štorga M, Mostashari A, Stanković T. Visualisation of the organisation knowledge structure evolution. J Knowl Manag. 2013;17(5):724–40. https://doi.org/10.1108/JKM-

02-2013-0058

[170] Cash P, Stanković T, Štorga M. Using visual information analysis to explore complex patterns in the activity of designers. Des Stud. 2014;35(1):1–28. https://doi.org/10.1016/j.destud.2013.06.001

[171] Linyuan LL, Zhou T. Link prediction in complex networks: A survey. Phys A Stat Mech its Appl. Elsevier B.V.; 2011;390(6):1150–70. https://doi.org/10.1016/j.physa.2010.11.027

[172] Wei Y, Spencer G. Measuring the value of accurate link prediction for network seeding. Computational Social Networks. Springer International Publishing; 2017. 1–35 p. https://doi.org/10.1186/s40649-017-0037-3

[173] Chiu C, Zhan J. Deep learning for link prediction in dynamic networks using weak estimators. IEEE Access. IEEE; 2018;6:35937–45. https://doi.org/10.1109/ACCESS.2018.2845876

[174] Gao F, Musial K, Cooper C, Tsoka S. Link prediction methods and their accuracy for different social networks and network metrics. Sci Program. 2015;2015(i). https://doi.org/10.1155/2015/172879

[175] Liben-Nowell D, Kleinberg J. The link prediction problem for social networks. Int Conf Inf Knowl Manag Proc. 2003;(November 2003):556–9. https://doi.org/10.1145/956958.956972

[176] Barrat A, Barthélemy M, Vespignani A. Dynamical processes on complex networks. Dynamical Processes on Complex Networks. 2008. 1–347 p. https://doi.org/10.1017/CBO9780511791383

[177] Zhou T, Lü L, Zhang YC. Predicting missing links via local information. Eur Phys J B. 2009;71(4):623–30. https://doi.org/10.1140/epjb/e2009-00335-8

[178] Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982;143(1):29–36. https://doi.org/10.1148/radiology.143.1.7063747

[179] Herlocker JL, Konstan JA, Terveen LG, Riedl JT. Evaluating collaborative filtering recommender systems. ACM Trans Inf Syst. 2004;22(1):5–53. https://doi.org/10.1145/963770.963772

[180] Geisser S. Predictive inference: An introduction. Predictive Inference: An Introduction. 2017. 1–264 p. https://doi.org/10.1201/9780203742310

[181] Smojver V, Štorga M, Zovak G. Exploring knowledge flow within a technology domain by conducting a dynamic analysis of a patent co-citation network. J Knowl Manag. 2020;ahead-of-p(ahead-of-print). https://doi.org/10.1108/jkm-01-2020-0079

[182] Adamic LA, Adar E. Friends and neighbors on the Web. Soc Networks. 2003;25(3):211–30. https://doi.org/10.1016/S0378-8733(03)00009-1

[183] Barabási AL, Jeong H, Néda Z, Ravasz E, Schubert A, Vicsek T. Evolution of the social network of scientific collaborations. Phys A Stat Mech its Appl. 2002;311(3–4):590–614. https://doi.org/10.1016/S0378-4371(02)00736-7

[184] Moore DW. Headlamp History and Harmonization. Umtri. 1998;(June):26.

[185] Daniel N, Donahoe PE. Thermal aspects of LED automotive headlights. 5th IEEE Veh Power Propuls Conf VPPC ’09. 2009;1193–9. https://doi.org/10.1109/VPPC.2009.5289714

[186] OECD. Patents and Innovation: Trends and Policy Challenges. Organ Econ Co-operation Dev. 2004;(12):32.

[187] Li Z, Liu C, Wang Y, Yan B, Yang C, Yang J, et al. An overview on memristor crossabr based neuromorphic circuit and architecture. IEEE/IFIP Int Conf VLSI Syst VLSI-SoC. 2015;2015-Octob:52–6. https://doi.org/10.1109/VLSI-SoC.2015.7314391

[188] Misra J, Saha I. Artificial neural networks in hardware: A survey of two decades of progress. Neurocomputing. Elsevier; 2010;74(1–3):239–55. https://doi.org/10.1016/j.neucom.2010.03.021

[189] Zhao C, Liu L, Yi Y. Design and analysis of real time spiking neural network decoder for neuromorphic chips. ACM Int Conf Proceeding Ser. 2019; https://doi.org/10.1145/3354265.3354280

[190] Goteti US, Zaluzhnyy IA, Ramanathan S, Dynes RC, Frano A. Low-temperature emergent neuromorphic networks with correlated oxide devices. Proc Natl Acad Sci. 2021 Aug 31;118(35):e2103934118. https://doi.org/10.1073/pnas.2103934118

[191] Sanchez ASBPC. NEUROMORPHIC COMPUTING AND SENSING 2021. 2021.

[192] Strumsky D, Lobo J, van der Leeuw S. Using patent technology codes to study technological change. Econ Innov New Technol. 2012;21(3):267–86. https://doi.org/10.1080/10438599.2011.578709

[193] Cascini G, Rotini F, Russo D. Networks of trends: Systematic definition of evolutionary scenarios. Procedia Eng. 2011;9:355–67. https://doi.org/10.1016/j.proeng.2011.03.125

[194] Youn H, Strumsky D, Bettencourt LMA, Lobo J. Invention as a combinatorial process: Evidence from US patents. J R Soc Interface. 2015;12(106). https://doi.org/10.1098/rsif.2015.0272

[195] Cascini G, Fantechi A, Spinicci E. Natural language processing of patents and technical documentation. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics). 2004;3163:508–20. https://doi.org/10.1007/978-3-540-28640-0_48

[196] Obradović T, Vlačić B, Dabić M. Open innovation in the manufacturing industry: A review and research agenda. Technovation. 2021 Apr;102:102221. https://doi.org/10.1016/j.technovation.2021.102221

11. APPENDIX A:

Table A-1 Link prediction results- PA algorithm applied to subset T1a

Percentage of predicted links being verified	Similarity measure value	Number of predicted links	Number of correctly predicted links	Precision
Top 10 %	339231	21	16	0.76
Top 20 %	301539	80	58	0.72
Top 30 %	263846	390	266	0.68
Top 40 %	226154	1936	1263	0.65

Table A-2 Link prediction results- PA algorithm applied to subset T2a

Percentage of predicted links being verified	Similarity measure value	Number of predicted links	Number of correctly predicted links	Precision
Top 10 %	339231	21	16	0.76
Top 20 %	301539	80	58	0.72
Top 30 %	263846	390	266	0.68
Top 40 %	226154	1936	1263	0.65

Table A-3Link prediction results- PA algorithm applied to subset T3a

Percentage of predicted links being verified	Similarity measure value	Number of predicted links	Number of correctly predicted links	Precision
Top 10 %	339231	21	17	0.80
Top 20 %	301539	80	60	0.75
Top 30 %	263846	389	269	0.69
Top 40 %	226154	1926	1264	0.65

Table A-4 Link prediction results- PA algorithm applied to subset T4a

Percentage of predicted links being verified	Similarity measure value	Number of predicted links	Number of correctly predicted links	Precision
Top 10 %	339231	20	17	0.85
Top 20 %	301539	75	56	0.74
Top 30 %	263846	373	261	0.69
Top 40 %	226154	1556	962	0.61

Table A-5Link prediction results- PA algorithm applied to subset T5a

Percentage of predicted links being verified	Similarity measure value	Number of predicted links	Number of correctly predicted links	Precision
Top 10 %	325442	21	16	0.76
Top 20 %	289282	89	63	0.70
Top 30 %	253122	385	266	0.69
Top 40 %	216961	2282	1249	0.54

Table A-6 Link prediction results- PA algorithm applied to subset T6a

Percentage of predicted links being verified	Similarity measure value	Number of predicted links	Number of correctly predicted links	Precision
Top 10 %	296465	21	19	0.90
Top 20 %	263524	75	65	0.86
Top 30 %	230584	711	607	0.85
Top 40 %	197643	1722	1165	0.67

Table A-7 Link prediction results- Adamic/Adar Index algorithm applied to subset T1b

Percentage of predicted links being verified	Similarity measure value	Number of predicted links	Number of correctly predicted links	Precision
Top 10 %	29,1	542	376	0.69
Top 20 %	25,9	830	376	0.45
Top 30 %	22,4	830	376	0.45
Top 40 %	19,4	830	376	0.45

Table A-8 Link prediction results- Adamic/Adar Index algorithm applied to subset T2b

Percentage of predicted links being verified	Similarity measure value	Number of predicted links	Number of correctly predicted links	Precision
Top 10 %	54,59	6	2	0.33
Top 20 %	48,52	8	2	0.25
Top 30 %	42,45	83	26	0.31
Top 40	36,39	1016	533	0.52

Table A-9 Link prediction results- Adamic/Adar Index algorithm applied to subset T3b

Percentage of predicted links being verified	Similarity measure value	Number of predicted links	Number of correctly predicted links	Precision
Top 10 %	64,56	101	20	0.2
Top 20 %	57,39	107	20	0.19
Top 30 %	50,21	130	20	0.15
Top 40 %	43,04	193	21	0.1

Table A-10 Link prediction results- Adamic/Adar Index algorithm applied to subset T4b

Percentage of predicted links being verified	Similarity measure value	Number of predicted links	Number of correctly predicted links	Precision
Top 10 %	67,68	95	14	0.15
Top 20 %	60,16	97	14	0.14
Top 30 %	52,64	124	14	0.11
Top 40 %	45,12	159	14	0.09

Table A-11 Link prediction results- Adamic/Adar Index algorithm applied to subset T5b

Percentage of predicted links being verified	Similarity measure value	Number of predicted links	Number of correctly predicted links	Precision
Top 10 %	67,5	95	14	0.15
Top 20 %	60	97	14	0.14
Top 30 %	52,5	124	14	0.11
Top 40 %	45	159	14	0.09

Table A-12 Link prediction results- Adamic/Adar Index algorithm applied to subset T6b

Percentage of predicted links being verified	Similarity measure value	Number of predicted links	Number of correctly predicted links	Precision
Top 10 %	67,32	525	0	0
Top 20 %	59,84	635	0	0
Top 30 %	52,36	652	0	0
Top 40 %	44,88	697	0	0

Table A-13 Link prediction results- Adamic/Adar Index algorithm applied to subset T7b

Percentage of predicted links being verified	Similarity measure value	Number of predicted links	Number of correctly predicted links	Precision
Top 10 %	67,32	525	0	0
Top 20 %	59,84	635	0	0
Top 30 %	52,36	652	0	0
Top 40 %	44,88	697	0	0

Table A-14 Link prediction results- Adamic/Adar Index algorithm applied to subset T8b

Percentage of predicted links being verified	Similarity measure value	Number of predicted links	Number of correctly predicted links	Precision
Top 10 %	67,32	525	0	0
Top 20 %	59,84	635	0	0
Top 30 %	52,36	652	0	0
Top 40 %	44,88	697	0	0

Table A-15 Link prediction results- Adamic/Adar Index algorithm applied to subset T9b

Percentage of predicted links being verified	Similarity measure value	Number of predicted links	Number of correctly predicted links	Precision
Top 10 %	67,32	525	0	0
Top 20 %	59,84	633	0	0
Top 30 %	52,36	652	0	0
Top 40 %	44,88	801	0	0

Table A-16 Link prediction results- Jaccard coefficient algorithm applied to subset T1b

Percentage of predicted links being verified	Similarity measure value	Number of predicted links	Number of correctly predicted links	Precision
Top 10 %	0,86	174	10	0,057
Top 20 %	0,77	853	392	0,46
Top 30 %	0,67	877	393	0,45
Top 40 %	0,57	961	457	0,48

Table A-17 Link prediction results- Jaccard coefficient algorithm applied to subset T2b

Percentage of predicted links being verified	Similarity measure value	Number of predicted links	Number of correctly predicted links	Precision
Top 10 %	0,88	425	136	0,32
Top 20 %	0,78	978	311	0,32
Top 30 %	0,68	1149	330	0,29
Top 40 %	0,58	1407	466	0,33

Table A- 18 Link prediction results- Jaccard coefficient algorithm applied to subset T3b

Percentage of predicted links being verified	Similarity measure value	Number of predicted links	Number of correctly predicted links	Precision
Top 10 %	0,9	108	13	0,12
Top 20 %	0,8	695	164	0,24
Top 30 %	0,7	927	164	0,18
Top 40 %	0,6	1060	164	0,15

Table A-19 Link prediction results- Jaccard coefficient algorithm applied to subset T4b

Percentage of predicted links being verified	Similarity measure value	Number of predicted links	Number of correctly predicted links	Precision
Top 10 %	0,9	89	10	0,11
Top 20 %	0,8	845	78	0,09
Top 30 %	0,7	1151	78	0,07
Top 40 %	0,6	1264	78	0,06

Table A-20 Link prediction results- Jaccard coefficient algorithm applied to subset T5b

Percentage of predicted links being verified	Similarity measure value	Number of predicted links	Number of correctly predicted links	Precision
Top 10 %	0,9	89	10	0,11
Top 20 %	0,8	860	14	0,02
Top 30 %	0,7	1158	17	0,01
Top 40 %	0,6	1286	17	0,01

Table A-21 Link prediction results- Jaccard coefficient algorithm applied to subset T6b

Percentage of predicted links being verified	Similarity measure value	Number of predicted links	Number of correctly predicted links	Precision
Top 10 %	0,9	46	2	0,04
Top 20 %	0,8	1602	2	0,001
Top 30 %	0,7	2030	2	0,001
Top 40 %	0,6	2222	2	0,001

Table A-22 Link prediction results- Jaccard coefficient algorithm applied to subset T7b

Percentage of predicted links being verified	Similarity measure value	Number of predicted links	Number of correctly predicted links	Precision
Top 10 %	0,9	107	0	0
Top 20 %	0,8	1713	8	0,005
Top 30 %	0,7	2030	8	0,004
Top 40 %	0,6	2338	8	0,003

Table A-23 Link prediction results- Jaccard coefficient algorithm applied to subset T8b

Percentage of predicted links being verified	Similarity measure value	Number of predicted links	Number of correctly predicted links	Precision
Top 10 %	0,9	189	0	0
Top 20 %	0,8	1800	8	0,004
Top 30 %	0,7	2232	8	0,004
Top 40 %	0,6	2536	8	0,003

Table A-24 Link prediction results- Jaccard coefficient algorithm applied to subset T9b

Percentage of predicted links being verified	Similarity measure value	Number of predicted links	Number of correctly predicted links	Precision
Top 10 %	0,9	195	0	0
Top 20 %	0,8	1913	0	0
Top 30 %	0,7	3093	185	0,06
Top 40 %	0,6	4879	185	0,04

Table A-25 Link prediction results- RAI algorithm applied to subset T1b

Percentage of predicted links being verified	Similarity measure value	Number of predicted links	Number of correctly predicted links	Precision
Top 10 %	0,77	167	23	0,14
Top 20 %	0,69	893	399	0,45
Top 30 %	0,60	901	401	0,45
Top 40 %	0,52	909	402	0,44

Table A-26 Link prediction results- RAI algorithm applied to subset T2b

Percentage of predicted links being verified	Similarity measure value	Number of predicted links	Number of correctly predicted links	Precision
Top 10 %	0,99	4	1	0,25
Top 20 %	0,88	7	2	0,29
Top 30 %	0,77	589	152	0,26
Top 40 %	0,66	1375	526	0,38

Table A-27 Link prediction results- RAI algorithm applied to subset T3b

Percentage of predicted links being verified	Similarity measure value	Number of predicted links	Number of correctly predicted links	Precision
Top 10 %	0,95	90	20	0,22
Top 20 %	0,85	110	20	0,18
Top 30 %	0,74	307	20	0,07
Top 40 %	0,64	584	20	0,03

Table A-28 Link prediction results- RAI algorithm applied to subset T4b

Percentage of predicted links being verified	Similarity measure value	Number of predicted links	Number of correctly predicted links	Precision
Top 10 %	1,1	1	0	0
Top 20 %	0,98	5	0	0
Top 30 %	0,85	103	14	0,14
Top 40 %	0,73	300	14	0,05

Table A-29 Link prediction results- RAI algorithm applied to subset T5b

Percentage of predicted links being verified	Similarity measure value	Number of predicted links	Number of correctly predicted links	Precision
Top 10 %	1,09	1	0	0
Top 20 %	0,97	5	0	0
Top 30 %	0,85	103	14	0,14
Top 40 %	0,73	353	17	0,05

Table A-30 Link prediction results- RAI algorithm applied to subset T6b

Percentage of predicted links being verified	Similarity measure value	Number of predicted links	Number of correctly predicted links	Precision
Top 10 %	1,071	1	0	0
Top 20 %	0,952	1	0	0
Top 30 %	0,833	458	0	0
Top 40 %	0,714	1021	2	0,002

Table A-31 Link prediction results- RAI algorithm applied to subset T7b

Percentage of predicted links being verified	Similarity measure value	Number of predicted links	Number of correctly predicted links	Precision
Top 10 %	1,071	1	0	0
Top 20 %	0,952	1	0	0
Top 30 %	0,833	571	8	0,01
Top 40 %	0,714	1104	8	0,007

Table A-32 Link prediction results- RAI algorithm applied to subset T8b

Percentage of predicted links being verified	Similarity measure value	Number of predicted links	Number of correctly predicted links	Precision
Top 10 %	1,071	1	0	0
Top 20 %	0,952	3	0	0
Top 30 %	0,833	658	8	0,012
Top 40 %	0,714	1191	8	0,007

Table A-33 Link prediction results- RAI algorithm applied to subset T9b

Percentage of predicted links being verified	Similarity measure value	Number of predicted links	Number of correctly predicted links	Precision
Top 10 %	1,071	1	0	0
Top 20 %	0,952	3	0	0
Top 30 %	0,833	599	8	0,013
Top 40 %	0,714	1132	0	0

Table A-34 Link prediction results- PA algorithm applied to subset T1b

Percentage of predicted links being verified	Similarity measure value	Number of predicted links	Number of correctly predicted links	Precision
Top 10 %	59427	103	1	0,01
Top 20 %	52824	168	1	0,006
Top 30 %	46221	18090	3	0,0002
Top 40 %	39618	37333	385	0,01

Table A-35 Link prediction results- PA algorithm applied to subset T2b

Percentage of predicted links being verified	Similarity measure value	Number of predicted links	Number of correctly predicted links	Precision
Top 10 %	221977	2	0	0
Top 20 %	197313	10	0	0
Top 30 %	172649	73	0	0
Top 40 %	147985	1170	22	0,02

Table A-36 Link prediction results- PA algorithm applied to subset T3b

Percentage of predicted links being verified	Similarity measure value	Number of predicted links	Number of correctly predicted links	Precision
Top 10 %	453033	2	0	0
Top 20 %	402696	16	0	0
Top 30 %	352359	1451	0	0
Top 40 %	302022	12352	0	0

Table A-37 Link prediction results- PA algorithm applied to subset T4b

Percentage of predicted links being verified	Similarity measure value	Number of predicted links	Number of correctly predicted links	Precision
Top 10 %	498560	1	0	0
Top 20 %	443164	3	0	0
Top 30 %	387769	294	0	0
Top 40 %	332373	2723	0	0

Table A-38 Link prediction results- PA algorithm applied to subset T5b

Percentage of predicted links being verified	Similarity measure value	Number of predicted links	Number of correctly predicted links	Precision
Top 10 %	514587	2	0	0
Top 20 %	457411	22	0	0
Top 30 %	400234	1291	0	0
Top 40 %	343058	7843	0	0

Table A-39 Link prediction results- PA algorithm applied to subset T6b

Percentage of predicted links being verified	Similarity measure value	Number of predicted links	Number of correctly predicted links	Precision
Top 10 %	556102	4	0	0
Top 20 %	494313	34	0	0
Top 30 %	432524	1919	0	0
Top 40 %	370735	13812	0	0

Table A-40 Link prediction results- PA algorithm applied to subset T7b

Percentage of predicted links being verified	Similarity measure value	Number of predicted links	Number of correctly predicted links	Precision
Top 10 %	556102	5	0	0
Top 20 %	494313	36	0	0
Top 30 %	432524	1923	0	0
Top 40 %	370735	14237	0	0

Table A-41 Link prediction results- PA algorithm applied to subset T8b

Percentage of predicted links being verified	Similarity measure value	Number of predicted links	Number of correctly predicted links	Precision
Top 10 %	556102	6	0	0
Top 20 %	494313	37	0	0
Top 30 %	432524	1923	0	0
Top 40 %	370735	14295	0	0

Table A-42 Link prediction results- PA algorithm applied to subset T8b

Percentage of predicted links being verified	Similarity measure value	Number of predicted links	Number of correctly predicted links	Precision
Top 10 %	558756	7	0	0
Top 20 %	496672	49	0	0
Top 30 %	434588	1763	0	0
Top 40 %	372504	14727	0	0

12. APPENDIX B

Appendinx B provides an overview of the 4 link prediction algorithms used in this research. The symbols u,v denote nodes, $\Gamma(u)$ and $\Gamma(v)$ denote the neighbour sets of these nodes, $Sim(u,v)$ denotes the similarity value of these nodes, k denotes the average degree and N denotes the number of nodes in the network. The following link prediction algorithms are used:

a) Resource Allocation Index [177] – assumes that node *u* can send some resource to node *y*, with their common neighbors playing the roles of transmitters. Each transmitter has a unit of resource, and will averagely distribute it to all neighbors. The similarity between *u* and *v* can be defined as the amount of resource *v* received from *u*. The Resource Allocation index can be stated as:

$$Sim(u,v) = \sum_{w \in \Gamma(u) \cap \Gamma(v)} \frac{1}{|\Gamma(w)|} \tag{4}$$

b) Jaccard Coefficient – measures the probability that both *u* and *v* have a common neighbour, for a randomly selected neighbour that either *u* or *v* have. It is a static measure used for comparing the similarity of sample sets. The complexity of this algorithm is $O(Nk^2)$ The Jaccard Coefficient can be stated as:

$$Sim(u,v) = \frac{|\Gamma(u) \cap \Gamma(v)|}{|\Gamma(u) \cup \Gamma(v)|}$$

c) Adamic-Adar index [182] - refines the simple counting of common features by weighting rarer features more heavily formalizing the intuitive notion that a low-degree neighbour is more likely to indicate a future connection then a high-degree one. Therefore, a common neighbour of a node pair with a smaller number of neighbours contributes more to the similarity value then a pair with a larger number of neighbours. The complexity of this algorithm is $O(Nk^2)$. The Adamic-Adar index can be stated as (note *z* is the low-degree neighbor):

$$Sim(u,v) = \sum_{z \in \Gamma(u) \cap \Gamma(v)} \frac{1}{\log|\Gamma(z)|}$$

d) Preferential Attachment [183] is an algorithm based on the rich-get-richer (aka power - law) intuition that nodes with many connections tend to create more new connection. How rich two nodes are is calculated by multiplying the number of neighbours they have. The complexity of this algorithm is $O(N^2k^2)$. The Preferential Attachment algorithm can be stated as:

$$Sim(u,v) = |\Gamma(u)| \cdot |\Gamma(v)|$$

BIOGRAPHY

Vladimir Smojver was born in Zagreb, Croatia, in 1990. He graduated high school, MIOC Zagreb, in 2009 and enrolled into the Faculty of Mechanical Engineering and Naval Architecture (FSB) at the University of Zagreb. He gained his bachelor’s degree in 2013 and gained his master’s degree in 2014. During his studies he was a teacher’s assistant in several courses of the Department of design and product development. He was awarded the Rector's Award for research work. At the end of his graduate program, he was awarded the title Magna Cum Laude. He also received a scholarship from the Stipo Lozić Baškarad foundation based on academic excellence.

After gaining his master’s degree, he started working in the Center for Vehicles of Croatia (CVH) as an expert associate and started his Ph.D. research at Faculty of Mechanical Engineering and Naval Architecture at the University of Zagreb (FSB). He has co-authored 2 journal papers and 2 conference papers.

From 2015 to 2018 he was part of a Croatian Science Foundation (CSF) project “Models and Methods of Innovation Management in Complex Engineering Systems Development – MInMED”.

As part of his doctoral research, he attended two PhD workshops held on the Faculty of Mechanical Engineering and Naval Architecture. In addition, he enrolled in two summer schools organized by the Technical University of Denmark.

ŽIVOTOPIS

Vladimir Smojver rođen je u Zagrebu, Hrvatska, 1990. godine. Završio je srednju školu MIOC u Zagrebu 2009. godine i upisao Fakultet strojarstva i brodogradnje (FSB) Sveučilišta u Zagrebu. Diplomu prvostupnika inženjera strojarstva stekao je je 2013. godine, a zvanje magistra inženjera strojarstva 2014. godine. Tijekom studija bio je demonstrator na više kolegija Katedre za konstruiranje i razvoj proizvoda. Nagrađen je Rektorovom nagradom za istraživački rad. Na kraju diplomskog studija dobio je titulu Magna Cum Laude. Također je dobio stipendiju zaklade Stipo Lozić Baškarad na temelju akademske izvrsnosti.

Nakon stjecanja titule magistra, počinje raditi u Centru za vozila Hrvatske (CVH) kao stručni suradnik te upisuje doktorski studij na Fakultetu strojarstva i brodogradnje (FSB). Koautor je 2 rada u časopisu i 2 rada na međunarodnim konferencijama.

Od 2015. do 2018. sudjeluje kao vanjski suradnik na projektu Hrvatske zaklade za znanost „Modeli i metode upravljanja inovacijama u razvoju kompleksnih inženjerskih sustava – MINMED“.

U sklopu doktorskog rada pohađao je dvije doktorske radionice na Fakultetu strojarstva i brodogradnje. Osim toga, sudjelovao je na dvije međunarodne ljetne škole u organizaciji Danskog tehničkog sveučilišta.

BIBLIOGRAPY

Journal papers:

Smojver V, Štorga M, Zovak G. Exploring knowledge flow within a technology domain by conducting a dynamic analysis of a patent co-citation network. Journal of Knowledge Management. 2020 Jun 17. https://doi.org/10.1108/JKM-01-2020-0079

Smojver V, Štorga M, Potočki E. Determining the life cycle phase of a technology based on patent data. Tehnički vjesnik. 2019 Feb 16;26(1):222-9. https://doi.org/10.17559/TV-20171123182106

Conference papers:

Smojver V, Štorga M, Potočki E. An extended methodology for the assessment of technical invention evolution. InDS 84: Proceedings of the DESIGN 2016 14th International Design Conference 2016 (pp. 1135-1144).

Smojver V, Potočki E, Štorga M. A visual analysis of technical knowledge evolution based on patent data. InDS 87-6 Proceedings of the 21st International Conference on Engineering Design (ICED 17) Vol 6: Design Information and Knowledge, Vancouver, Canada, 21-25.08. 2017 2017 (pp. 307-316).