



Ph.D. Defense

Exploring the Technological Phylogenetic Network: Construction, Historical Analysis, and Future Predictions of Technological Evolution

Youwei He

Ph.D. candidate, SNU

TEMPEP

heyuwei2022@gmail.com

2024. 11. 20.

© 2025 Youwei He. All rights reserved.

CONTENTS

Introduction

Conceptual background

Research question

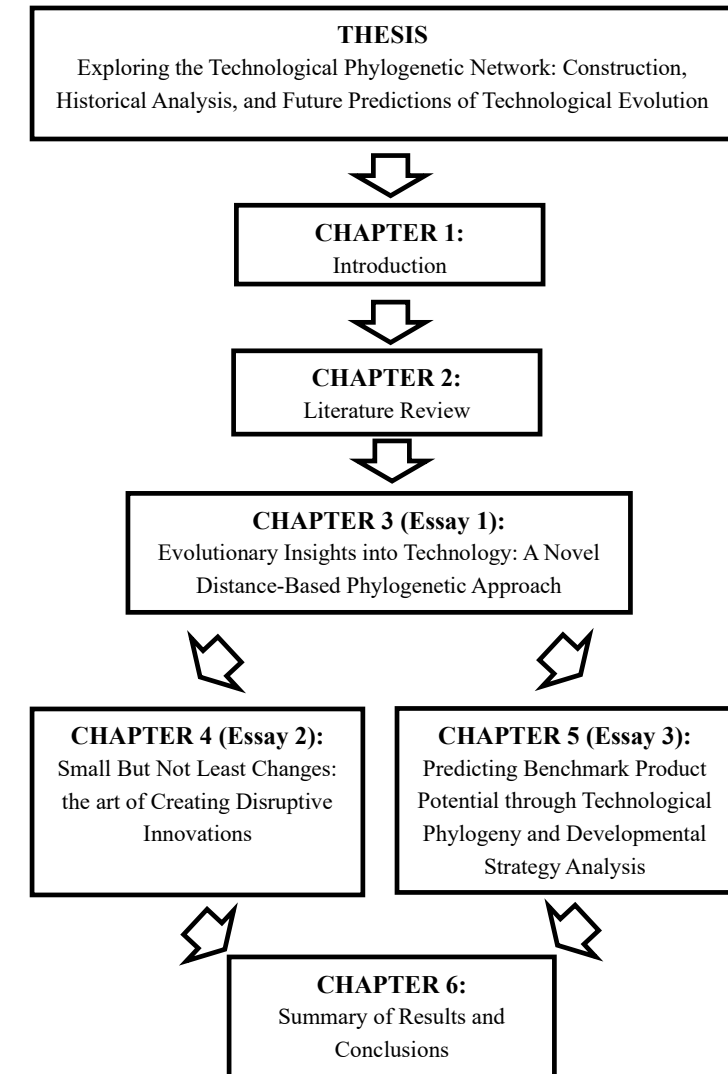
Essay I: Evolutionary Insights into Technology: A Novel Distance-Based Phylogenetic Approach

Essay II: Small But Not Least Changes: The Art of Creating Disruptive Innovations

Essay III: Predicting Benchmark Product Potential through Technological Phylogeny and Developmental Strategy Analysis

Conclusion

Reference



Introduction

Research Background

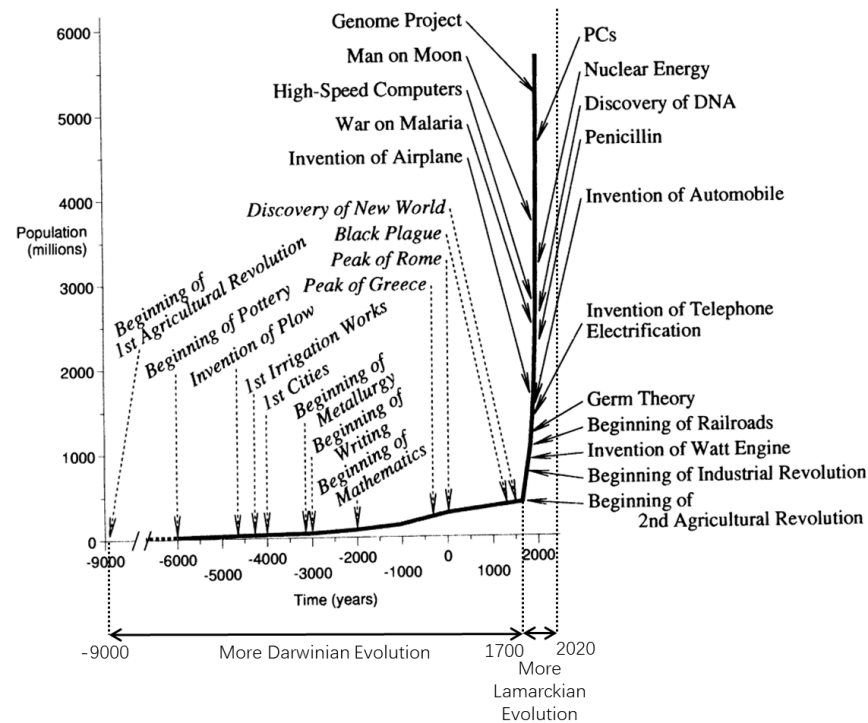
- Neo-Darwinism, refers to the modern synthesis of Darwin's theory of natural selection combined with genetics.
- Universal (Generalized) Darwinism applies Darwinian principles of variation, selection, and inheritance to non-biological systems like social, cultural, and technological evolution.
- Lamarckism suggests that organisms can acquire traits during their lifetime and pass them on to their offspring.

Theory	Mechanism of Evolution	Inheritance	Current Status	Phylogeny
Neo-Darwinism	Evolution through natural selection based on random genetic mutations (Mayr, 2001)	Genetic inheritance through DNA, primarily vertical gene transfer (VGT)	Widely accepted in biological evolution; limited applicability to technological evolution (Nelson & Winter, 1982).	Biological
Generalized Darwinism	Selection processes in evolving systems, applicable beyond biology (Hodgson & Knudsen, 2010)	Inheritance includes genetic, social, and technological elements, considering horizontal and vertical transmission	Increasingly used in social sciences, economics, and technology studies to explain adaptation and selection (Arthur, 2009; Mehmood et al., 2019)	Biological Non-biological
Lamarckism	Inheritance of acquired traits, involving adaptive change during an organism's lifetime (Bowler, 1989)	Transmission of acquired characteristics across generations, involving both direct modification and learning	Less accepted in biology but has influenced theories of cultural and technological evolution, where learning and adaptation are critical (Cimoli & Dosi, 1995)	Non-biological



Introduction

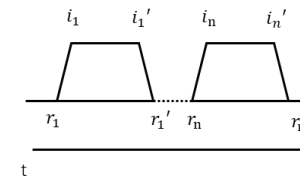
Comparison of Technology Evolution Theories (Generalized Darwinism and Lamarckism)



A. The growth of the world population and major technological events (adapted from Fogel, 1999)

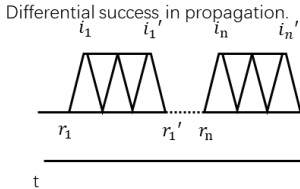
B. Darwinian Evolution

Replication of a cultural code (routine) with random component + recombination
Differential success in propagation.



C. Lamarckian Evolution

Environmental or idiosyncratic stimuli change the replicators' code (routine). Replication of altered code (routine).



r : replicating routine.

i : interactors (social identities) instructed by the replicating routine.

B.C. Selection models (adapted from Knudsen, 2002)

Technology change and evolution models (adapted from Fogel, 1999; Knudsen, 2002)

Conceptual background

Background of phylogeny

Phylogenetic phylogeny in biology

Neo-Darwinism and biological phylogeny dominant the biology. Darwin's tree of life is a classic example that represents relationships among all living organisms (Darwin, 1859). Literature on evolutionary biology discusses how phylogenetic methods are used to identify lineage-specific adaptations and reconstruct ancestral states (Felsenstein, 1985).

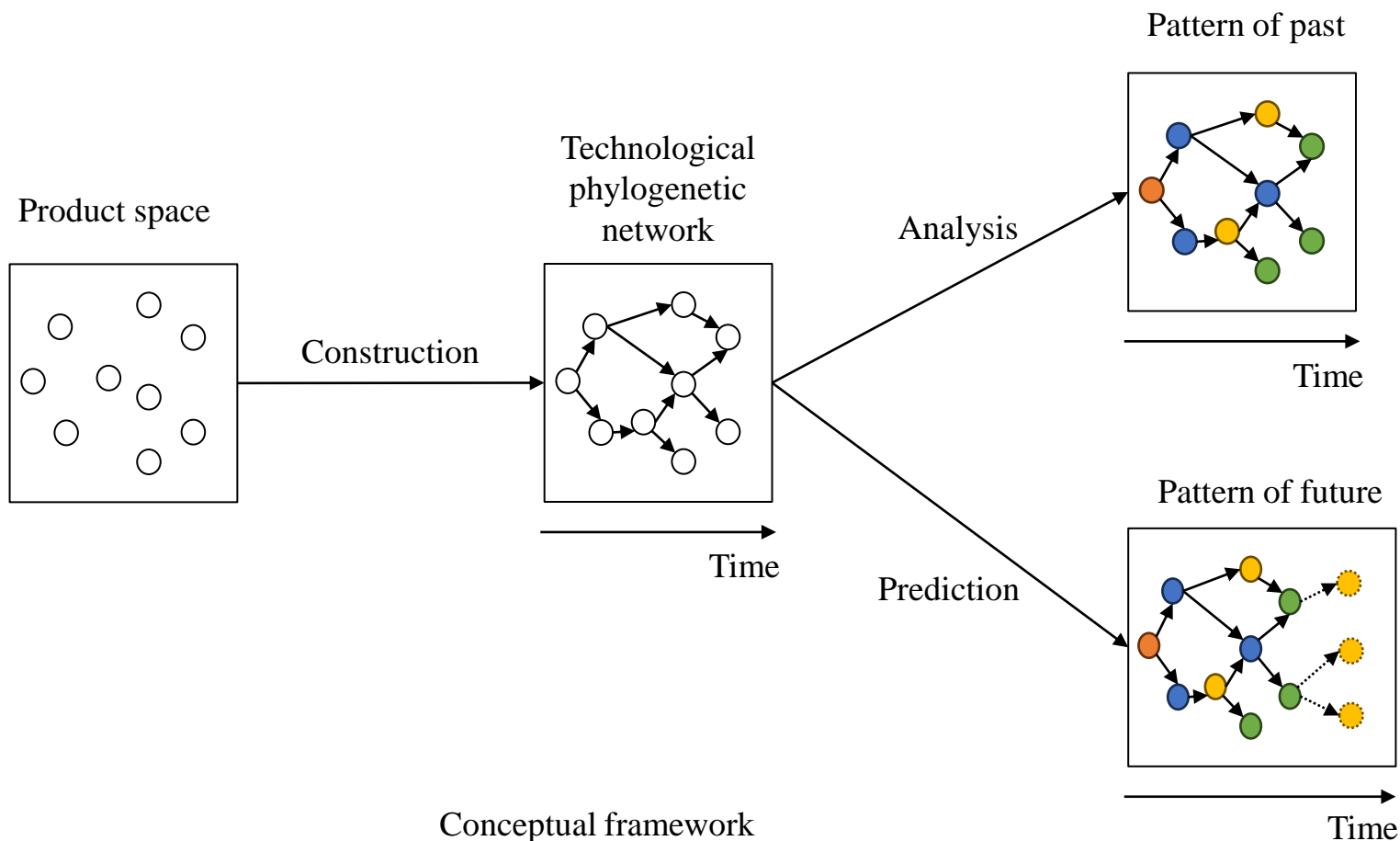
Phylogenetic phylogeny in non-biology

Universal Darwinism and Lamarckism have indeed found acceptance in non-biological fields and Non-biological phylogeny has been applied in these areas as well. In technological contexts, phylogenetic approaches have been employed to trace the evolution of innovations and to identify common technological "ancestors" (Arthur, 2009; Basalla, 1988; Lee et al., 2022). Cultural phylogenetics, for instance, applies phylogenetic techniques to analyze relationships between cultural artifacts and technologies (Straffon, 2016). Studies demonstrate that phylogenetic networks are particularly useful for capturing the complex, reticulate nature of technological evolution, characterized by frequent recombination and reuse of innovations (Fleming & Sorenson, 2001).



Conceptual background

Conceptual framework of the thesis



Research Question

Research Question

“How can technological phylogenetic networks be reliably constructed and utilized to analyze technological evolution in the past and predict trends in the future?”

Research gap

While phylogenetic networks have been extensively used in biological studies, their application in technological evolution is relatively novel. There is a lack of comprehensive methodologies and mathematical definition for constructing these networks for products and using them to analyze past technological trends and predict future developments.

Sub research questions

Q1. How can a reliable technological phylogeny be constructed using a general distance-based method? (Define and validate the method)

Q2. How can technological phylogeny be utilized to analyze the historical evolution of products, specifically to identify the design principles of disruptive products? (Analyze the past)

Q3. How can technological phylogeny be employed to predict benchmark product potential from the perspective of technological evolution? (Predict the future)

Essay I

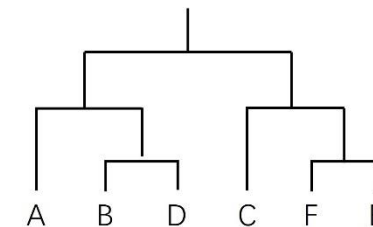
Research Question

How can a reliable technological phylogeny be constructed using a general method based on the distance of technological innovations?
(Define and validate the method)

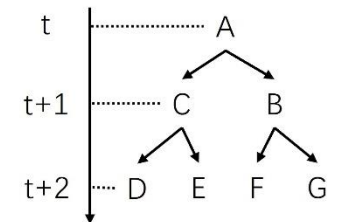
Research gap

1. Biological phylogeny

The assumptions of biological phylogeny based on the Neo-Darwinism, include a constant rate of genetic variation and fixed nucleotides. Additionally, the timing of species emergence is uncertain and can only be inferred. Evolution must be based on species rather than individuals.



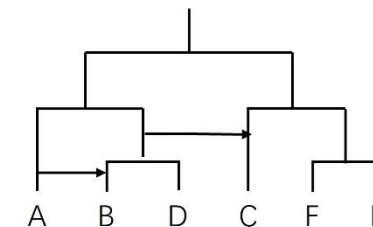
a. Biological phylogenetic tree



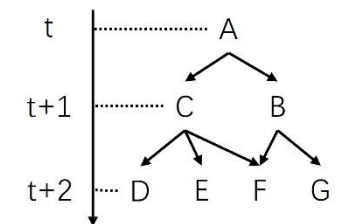
c. Technological phylogenetic tree

2. Technological phylogeny

The previous research has applied phylogenetic methods to technological evolution, introducing the concept of product phylogenetic tree to technological evolution for the first time (Lee et al., 2022). However, there are certain limitations. First, it is limited to the genetic representation of technology, but the genetic representation is not perfect for technology (Ziman, 2003). Second, there are time window constraints. Third, the tree form is less effective than the network form in representing horizontal technology transfer. Fourth, there is no strict mathematical definition. Fifth, it is still based on taxon, but Lamarckism make the individual evolution possible.



b. Biological phylogenetic network



d. Technological phylogenetic network

Essay I

Research Question

How can a reliable technological phylogeny be constructed using a general method based on the distance of technological innovations?
(Define and validate the method)

Research gap

Previous studies on the phylogenetic approach, whether in the biological or technical fields, have not validated this method. Instead, they have derived it directly from its definition. But how can we be certain that this phylogenetic approach reliably reflects the evolution of products?

Operational definition	Technological characteristics	Technological content	Tree	Network	Approach type	Validation	Literature
Yes	Yes	No	Yes	No	biological	No	(Chavalarias & Cointet, 2013 ; Khanafiah & Situngkir, 2006 ; McCarthy et al., 2000 ; O'Brien et al., 2001 ; Tëmkin & Eldredge, 2007 ; Youngblood et al., 2021)
No	Yes	No	No	Yes	technological	No	(Carignani et al., 2019)
Yes	Yes	No	Yes	No	technological	No	(Jeong & Lee, 2024 ; Jeong et al., 2023 ; Lee et al., 2023 ; Lee et al., 2022 ; Park et al., 2024)
Yes	Yes	Yes	Yes	Yes	technological	Yes	This study

Essay I

Research Question

How can a reliable technological phylogeny be constructed using a general method based on the distance of technological innovations?

The detailed mathematical definitions of the Phylogenetic Network are in the Annexes.

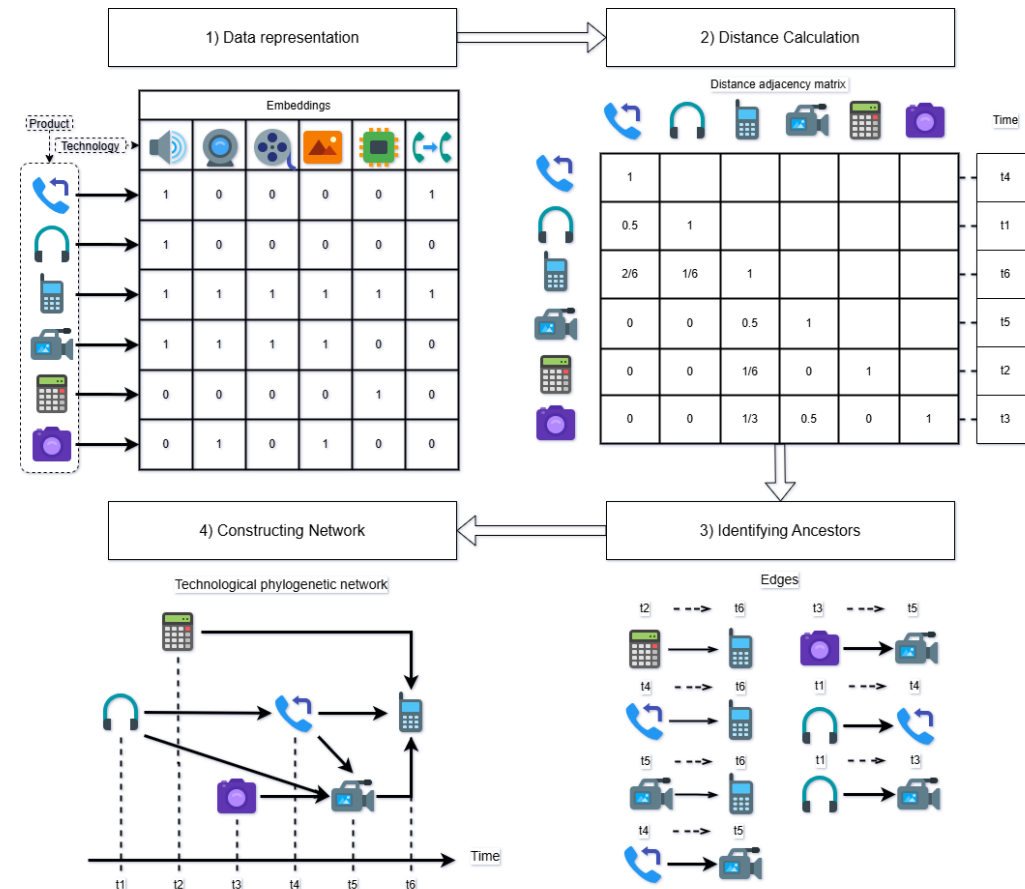
Steps for Construction of Phylogenetic Network

Step 1 Data Representation

Step 2 Distance Calculation

Step 3 Identifying Ancestors

Step 4 Constructing Network



Essay I

Validation of the Method

Criterion 1: Path Dependence versus Technology Transfer

Criterion 2: Uneven Distribution of Technological Influence

Criterion 3: Cluster Effect and Innovation Ecosystem

Criterion 4: Gradual and Radical Technological Evolution

Criterion 5: Temporal Proximity in Innovation

Technological Evolution (Wagner & Rosen, 2014)	Criterion				
	1	2	3	4	5
Extinction and replacement		√			
Descent with modification	√			√	√
Horizontal information transfer	√				
Combinatorial innovation	√				
Exaptation	√				
Ecosystem engineering			√		
Episodic change				√	√
Multiples and singletons		√			



Data

Product: ZOL Mobile¹

1815 mobile phone models with 9946 technical features from 2001 to 2020 (Gene based)

Paper: OpenAlex²

We analyzed 27,700 AI-related papers across 41 academic fields over 5-year intervals from 1900 to 2020. Using Sentence-BERT (Reimers & Gurevych, 2019), we embedded the abstracts into 384-dimensional vectors and calculated the mean embeddings to represent each academic field. (Content based)

¹ mobile.zol.com.cn

² <https://openalex.org/>

Criterion, hypothesis, and theory

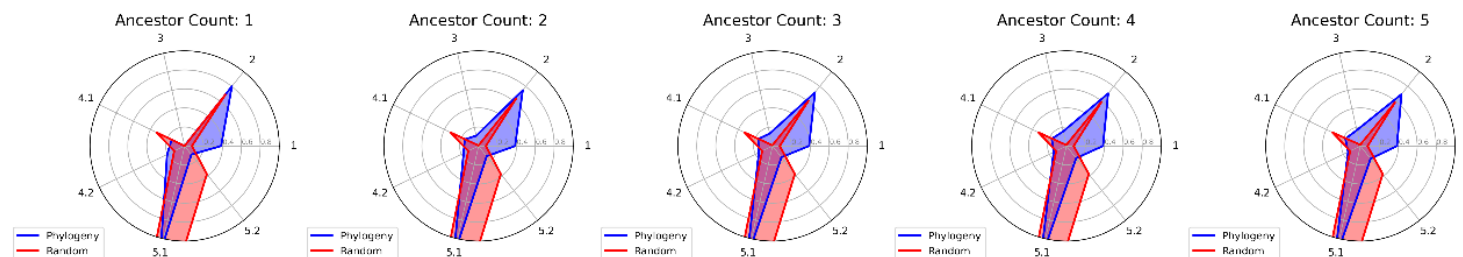
Criterion	Hypothesis (Product)	Hypothesis (Paper)	Theory
1	H0: The proportion of edges connecting nodes of the same brand in phylogeny should be greater than that in a random network.	H0: The proportion of edges connecting nodes of the same field in phylogeny should be greater than that in a random network.	Path dependency and lock-in (Cecere et al., 2014; David, 1985; Foxon, 2014), Cumulative Nature of Innovation (Basalla, 1988; Nelson, 1985), Dynamic Capabilities (Teece, 2007; Teece et al., 1997)
2	H1: In phylogeny, the Gini index of the node's outdegrees should be greater than that in a random network.	H1: In phylogeny, the Gini index of the node's outdegrees should be less than that in a random network.	Dominant effect (Cecere et al., 2015; Srinivasan et al., 2006), Pareto Principle (Pareto, 1964), Power Law Distribution (Zipf, 2016), Network Effect (Katz & Shapiro, 1985), Economies of Scale (Stigler, 1958), Winner-Takes-All (Frank & Cook, 1995; Schulz & Mayerhoffer, 2021), Market Concentration (Bain, 1951), Social centralization (Linzhuo et al., 2020)
3	H3: In phylogeny, the clustering coefficient should be greater than a random network.	H3: In phylogeny, the clustering coefficient should be greater than a random network.	Innovation ecosystem (Frenkel & Maital, 2014; Wagner & Rosen, 2014), co-evolution (Murray, 2002), niche (Kemp et al., 1998)
4	H4.1: In phylogeny, the average distance between the focal product and its ancestor should be smaller than that in the random network.	H4.1: In phylogeny, the average distance between the focal product and its ancestor should be smaller than that in the random network.	Punctuated equilibrium (Loch & Huberman, 1999; Mokyr, 1990), incremental and radical innovation (Norman & Verganti, 2014), Creative Destruction (Schumpeter, 2013), S-Curve (Christensen, 1992), technological discontinuity (Anderson & Tushman, 2018; Tushman & Anderson, 2018), Episodic change (Wagner & Rosen, 2014)
	H4.2: The Gini coefficient of this distance should be larger than that in the random network.	H4.2: The Gini coefficient of this distance should be larger than that in the random network.	
5	H5.1: The difference in years between the edge-connected ancestor and the focal node should be less in the phylogeny network than in the random network.	H5.1: The difference in years between the edge-connected ancestor and the focal node should be less in the phylogeny network than in the random network.	The trend of citing the latest research (Foster et al., 2015), the importance of the latest research (Fortunato et al., 2018), the importance of the latest knowledge combination (Youn et al., 2015)
	H5.2: The Gini of phylogeny should be less .	H5.2: The Gini of phylogeny should be greater .	

Testing approach

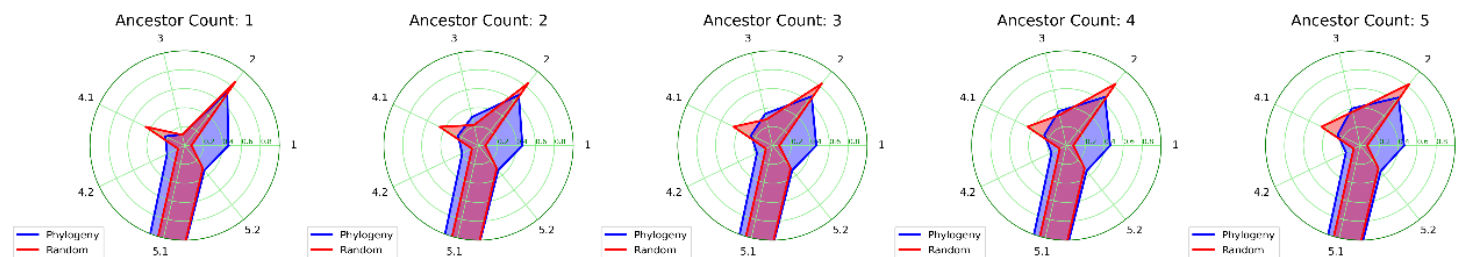
1. Phylogenetic Network Construction
2. Random Rewiring Network Generation
3. Metric Calculation
4. Hypothesis Testing (Permutation)

Testing results (permutation)

a. Product data



b. Paper data



Testing results (permutation)

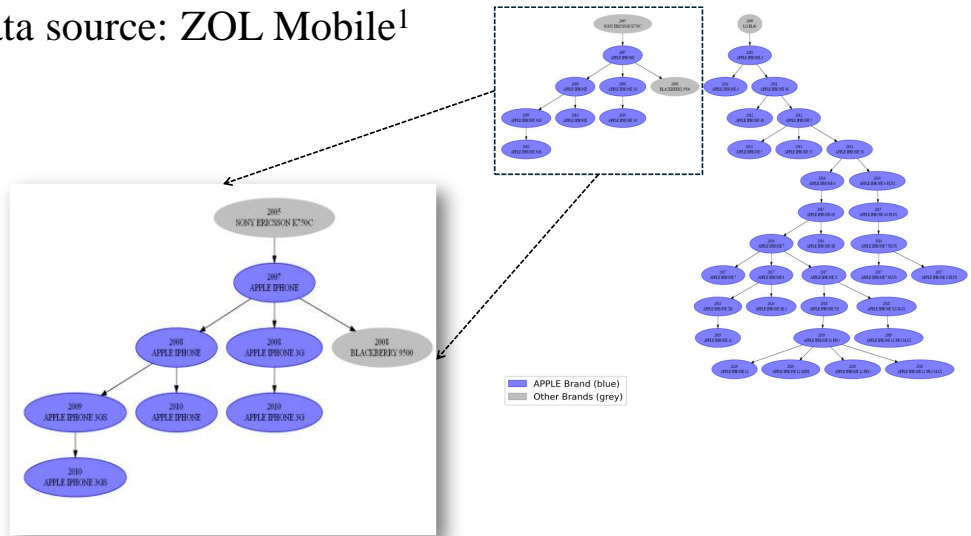
Criterion	Ancestors Hypothesis	AI related Papers					Mobile phone products				
		1	2	3	4	5	1	2	3	4	5
1	H1	0.3911 *	0.39 **	0.3895 *	0.3894 **	0.3894 **	0.3079 **	0.308 **	0.3085 *	0.3089 **	0.3093 **
2	H2	-0.141 **	-0.1632 **	-0.1732 **	-0.1785 **	-0.1812 **	0.11 *	0.1148 **	0.1107 **	0.1129 ***	0.1106 **
3	H3	0.0018	0.0875 **	0.0605 **	0.0404 **	0.0369 **	0.0019 **	0.1056 **	0.1267 **	0.1472 *	0.1583 **
4	H4.1	-0.2293 **	-0.2151 **	-0.205 **	-0.1971 *	-0.1907 **	-0.1762 **	-0.1662 **	-0.1594 **	-0.1546 *	-0.1508 **
4	H4.2	0.128 **	0.1141 **	0.1044 **	0.0961 *	0.0909 *	0.0834 **	0.0546 **	0.0392 **	0.0298 **	0.0236 **
5	H5.1	-19.6469 **	-19.0853 **	-18.7896 **	-18.5171 **	-18.0646 ***	-0.1508 **	-0.1508 **	-0.1508 **	-0.1508 **	-0.1508 **
5	H5.2	0.0264 **	0.0244 **	0.0256 **	0.0352 **	0.0385 *	-0.263 **	-0.2392 **	-0.2265 **	-0.2136 **	-0.2053 **

Note: Significance Indicators: '***' indicates highly significant ($p < 0.001$), '**' indicates significant ($p < 0.01$), '*' indicates moderately significant ($p < 0.05$).

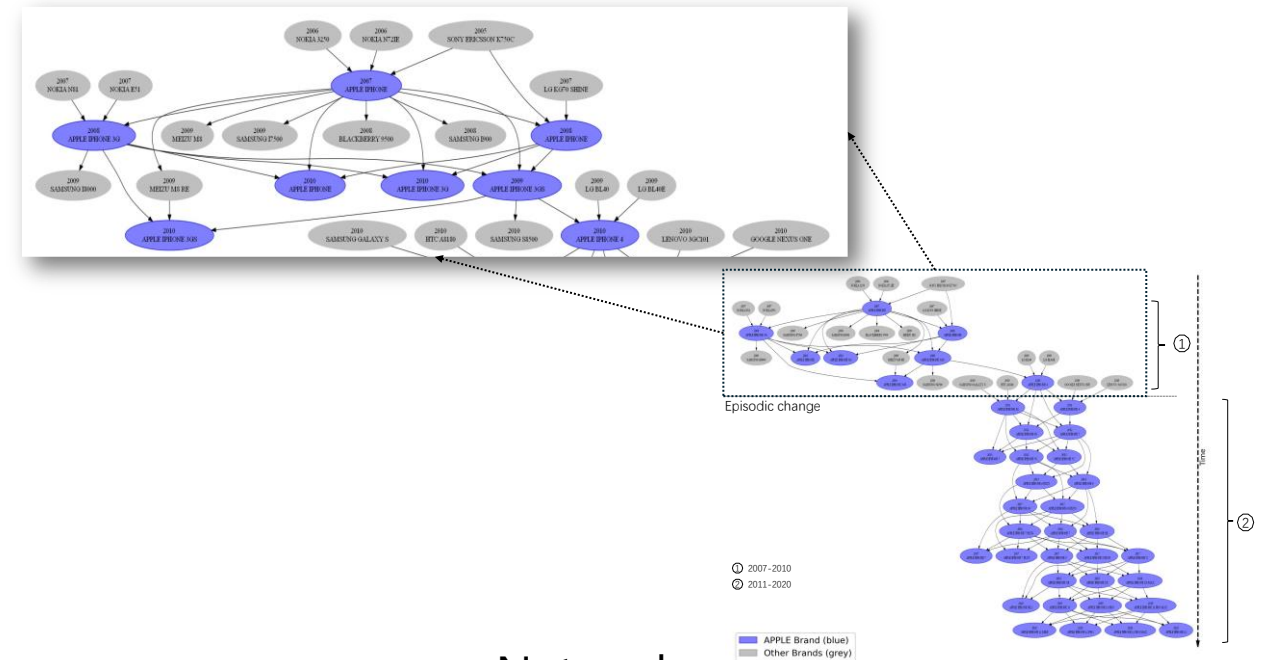
Empirical application

Phylogenetic network of Apple mobile phones

Data source: ZOL Mobile¹



Tree



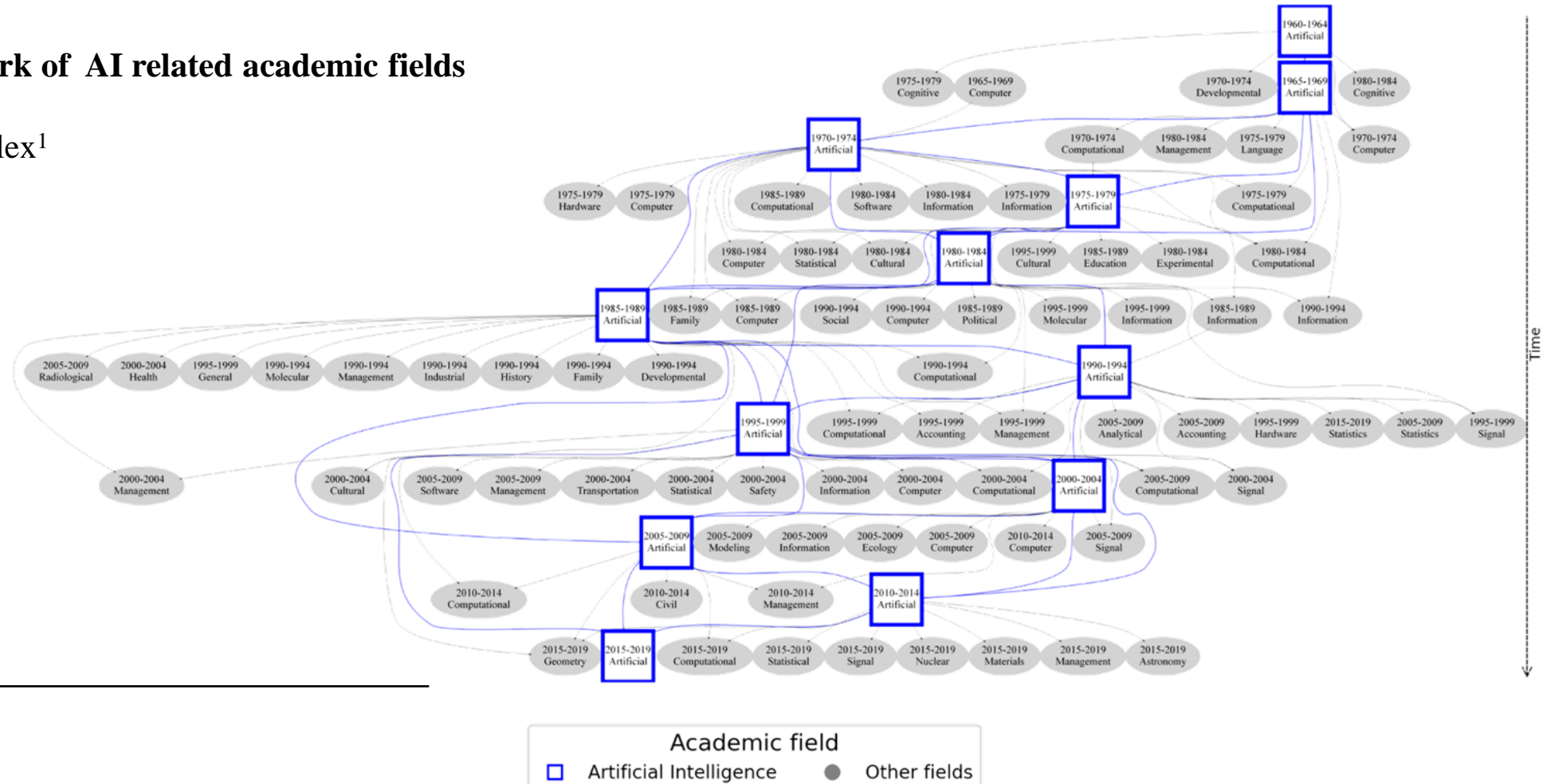
Network



Empirical application

Phylogenetic network of AI related academic fields

Data source: OpenAlex¹



¹ <https://openalex.org/>



Essay I

Implication

We propose a general method for constructing technological phylogenies and provide its mathematical definition. This method can be constructed based not only on the technological genes of products but also on the embeddings of technological contents. Additionally, we have constructed and analyzed technological phylogenetic networks using data from scientific papers and cell phones. Furthermore, we validated the method using cell phone data, demonstrating its reliability.

Essay II

Research Question

How can a novel method leveraging product information networks and the D index be developed to identify and measure disruptive products? What is the principle of disruptive products?

Research gap

The main research gap is the lack of a comprehensive method to measure and analyze disruptive product using product information. And there is no principle of how to make disruptive products.

Technology level disruption	Market level disruption	Product	Phylogenetic approach	Literature
No	Yes	No	No	(Christensen & Raynor, 2013 ; Christensen, 2015 ; Govindarajan & Kopalle, 2006a ; Linton, 2002 ; Schmidt & van der Sijde, 2022 ; Tellis, 2006)
No	Yes	Yes	No	(Linton, 2002)
Yes	No	No	No	(Funk & Owen-Smith, 2017 ; Park et al., 2023 ; Wang et al., 2024 ; Wang et al., 2022 ; Wu et al., 2019)
No	No	Yes	Yes	(Lee et al., 2023 ; Lee et al., 2022 ; Park et al., 2024)
Yes	Yes	Yes	Yes	This study



Essay II

Research Question

How can a novel method leveraging product information networks and the D index be developed to identify and measure disruptive products? What is the principle of disruptive products?

Data

Car data from Edmunds¹. From 2012 to 2023, there are 4496 products with 38127 technological features. For the binary features, we applied one-hot encoding. For the continuous features, we normalized them to the range [0,1].

¹ <https://www.edmunds.com/>

Essay II

Research Question

How can a novel method leveraging product information networks and the D index be developed to identify and measure disruptive products? What is the principle of disruptive products?

Methods

Based on the technological product network of products and the Disruption (D) index (Wu et al., 2019), we define the product disruption index (PDI).

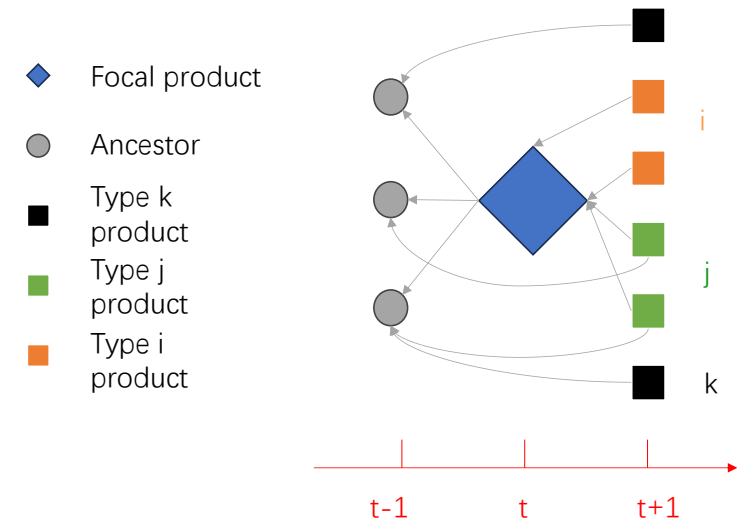
$$PDI = \frac{n_i - n_j}{n_i + n_j + n_k}$$

Where:

n_i : the number of i type nodes that cite the focal paper but not the ancestors of the focal node.

n_j : the number of j type nodes that cite both the focal node and any of its ancestors.

n_k : the number of k type nodes that cite an ancestor of the focal node but not the focal node.



Part of product network.

This network contains the focal product in the center. At time $t+1$, type i , j , and k products are present.

Essay II

Regressions (Pooled OLS)

Dependent variable: PDI

Inherit but change a small number of important genes is the principle for making disruptive product.

	(1)	(2)	(3)	(4)	(5)	(6)
	Frequency weight			TF-IDF weight		
Unweighted inherited gene rate	0.268*** (0.031)			0.360*** (0.035)		
Weighted inherited gene rate	-0.472*** (0.075)			-0.259*** (0.031)		
Ancestor PDI	0.363*** (0.011)	0.363*** (0.011)	0.363*** (0.011)	0.315*** (0.012)	0.310*** (0.011)	0.315*** (0.012)
Car rating	0.013* (0.006)	0.012+ (0.006)	0.013* (0.006)	0.014* (0.006)	0.014* (0.006)	0.014* (0.006)
Price(log)	-0.009** (0.003)	-0.009** (0.003)	-0.009** (0.003)	-0.009** (0.003)	-0.009** (0.003)	-0.009** (0.003)
Brand new gene rate	0.415*** (0.030)	0.380*** (0.029)	0.415*** (0.030)	0.269*** (0.031)	0.146*** (0.035)	0.269*** (0.031)
Weighted lost gene rate		0.151+ (0.083)			0.416*** (0.035)	
Unweighted lost gene rate		-0.168*** (0.036)			-0.516*** (0.039)	
Weighted new gene rate			0.472*** (0.075)			0.259*** (0.031)
Unweighted new gene rate			-0.268*** (0.031)			-0.360*** (0.035)
N	3694	3694	3694	3694	3694	3694
R2	0.345	0.338	0.345	0.350	0.362	0.350

Notes: This table evaluates the relationship between different measures of the use of product genetic information, survey data from website, price and 1-year PDI. Estimates are from ordinary-least-squares regressions. Each coefficient is tested against the null hypothesis of being equal to 0 using a two-sided t-test. We do not adjust for multiple hypothesis testing. Robust standard errors are shown in parentheses. + p<0.1; * p<0.05; ** p<0.01; *** p<0.001.

Essay II

Regressions (Fixed effects)

Dependent variable: PDI

After accounting for fixed effects of brand and year, the brand-new gene rate is negatively correlated with the PDI. This indicates that the blind introduction of new technologies can reduce the disruptiveness of a product.

	(1)	(2)	(3)	(4)	(5)	(6)
	Frequency weight			TF-IDF weight		
Unweighted inherited gene rate	0.247*** (0.046)			0.543*** (0.047)		
Weighted inherited gene rate	-0.209* (0.098)			-0.383*** (0.038)		
Ancestor PDI	0.581*** (0.024)	0.581*** (0.024)	0.581*** (0.024)	0.521*** (0.025)	0.524*** (0.024)	0.521*** (0.025)
Car rating	0.009+ (0.005)	0.009+ (0.005)	0.009+ (0.005)	0.010+ (0.005)	0.008 (0.005)	0.010+ (0.005)
Price(log)	-0.013* (0.006)	-0.013* (0.005)	-0.013* (0.006)	-0.013* (0.005)	-0.012* (0.005)	-0.013* (0.005)
Brand new gene rate	-0.215*** (0.046)	-0.235*** (0.044)	-0.215*** (0.046)	-0.432*** (0.047)	-0.499*** (0.055)	-0.432*** (0.047)
Weighted lost gene rate		-0.020 (0.126)			0.456*** (0.046)	
Unweighted lost gene rate		-0.178*** (0.052)			-0.622*** (0.055)	
Weighted new gene rate			0.209* (0.098)			0.383*** (0.038)
Unweighted new gene rate			-0.247*** (0.046)			-0.543*** (0.047)
Year	✓	✓	✓	✓	✓	✓
Brand	✓	✓	✓	✓	✓	✓
N	3694	3694	3694	3694	3694	3694
R2	0.529	0.528	0.529	0.553	0.556	0.553

Notes: This table evaluates the relationship between different measures of the use of product genetic information, survey data from website, price and 1-year PDI. Estimates are from multiply fixed effects regressions. Each coefficient is tested against the null hypothesis of being equal to 0 using a two-sided t-test. We do not adjust for multiple hypothesis testing. Robust standard errors are shown in parentheses. + p<0.1; * p<0.05; ** p<0.01; *** p<0.001.

Essay II

Robustness (Logit fixed effects regression)

Dependent variable: PDI dummy (If PDI>0, dummy=1,else dummy=0)

	(1)	(2)	(3)	(4)	(5)	(6)
	Frequency weight			TF-IDF weight		
Unweighted inherited gene rate	4.977*** (0.934)			7.076*** (1.010)		
Weighted inherited gene rate	-4.440* (2.053)			-3.513*** (0.777)		
Ancestor PDI	7.462*** (0.371)	7.431*** (0.374)	7.462*** (0.371)	7.046*** (0.384)	7.258*** (0.385)	7.046*** (0.384)
Car rating	0.024 (0.205)	0.030 (0.204)	0.024 (0.205)	0.045 (0.205)	0.022 (0.204)	0.045 (0.205)
Price(log)	-0.177 (0.114)	-0.170 (0.113)	-0.177 (0.114)	-0.194+ (0.114)	-0.183 (0.114)	-0.194+ (0.114)
Brand new gene rate	-3.206*** (0.785)	-3.669*** (0.776)	-3.206*** (0.785)	-5.184*** (0.819)	-4.620*** (0.923)	-5.184*** (0.819)
Weighted lost gene rate		-1.284 (2.361)			1.764+ (0.904)	
Unweighted lost gene rate		-3.036** (0.988)			-5.155*** (1.068)	
Weighted new gene rate			4.440* (2.053)			3.513*** (0.777)
Unweighted new gene rate			-4.977*** (0.934)			-7.076*** (1.010)
N	3733	3733	3733	3733	3733	3733
R2	0.324	0.321	0.324	0.331	0.323	0.331

Standard errors in parentheses = "+ p<0.1, * p<0.05, ** p<0.01, *** p<0.001"

Implication

This study provides a novel approach to measuring product disruptiveness through the construction of product similarity network. By integrating product technological evolutionary data (from phylogenetic network) and PDI, it offers a quantitative framework for understanding how products achieve disruption status. The findings highlight the importance of new technologies in driving product disruptiveness and suggest that inheriting traits from disruptive ancestors can enhance a product's disruptiveness. This approach not only fills a gap in existing literature but also offers practical insights for product design and innovation management.

Essay III

Research Question

How can we define and identify the benchmark product? What makes a product to be a benchmark? Can we predict the potential for a product to become a benchmark product?

Research gap

Despite extensive research on new product development (NPD) and innovation, a significant gap exists in studies specifically focusing on identifying and predicting benchmark products. Existing literature often lacks a systematic approach to understanding how products achieve benchmark status (Ulrich & Eppinger, 2016).

Data	Method	Results and implications	Time available	Literature
Patent	Supervised models	Emerging technologies	Early stage	(Kyebambe et al., 2017 ; Lee et al., 2018 ; Zhou et al., 2020)
Patent	Supervised models	Technology convergence	Middle stage	(Cho et al., 2021 ; C. Lee et al., 2021 ; San Kim & Sohn, 2020)
Product	Supervised models with phylogenetic tree	Benchmark product	Late stage	This study

Essay III

Research Question

How can we define and identify the benchmark product? What makes a product to be a benchmark? Can we predict the potential for a product to become a benchmark product?

Methods

Data source: ZOL Mobile¹

Cell phones in the Chinese market from 2001 to 2020. Out of 3,430 cell phone models, across 63 different brands and 11,006 technological features.

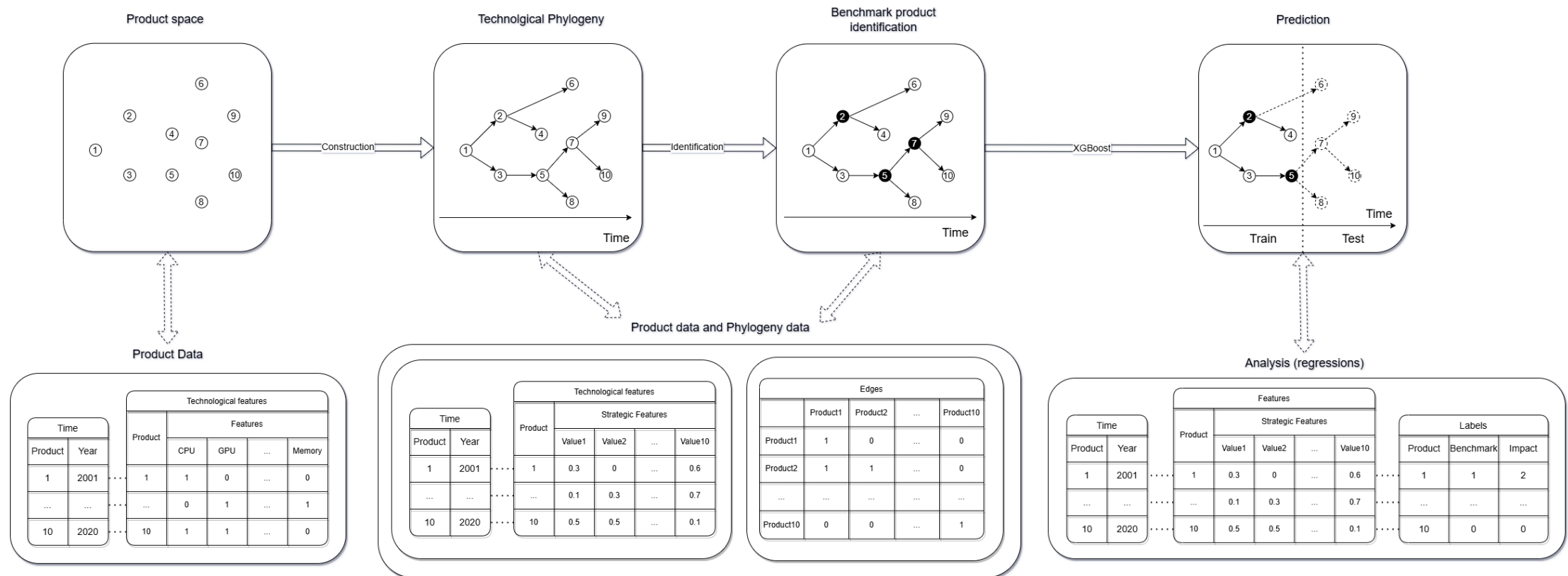
We use a technological phylogenetic tree to identify the benchmark products by years. We define the benchmark products as the top $n\%$ most impactful products within 5 years by the phylogenetic tree. The impacts can be measured by the descendant counts (Shi & Evans, 2023). We also use the XGBoost (Chen et al., 2016) to predict the benchmark products by years.

¹ mobile.zol.com.cn



Essay III

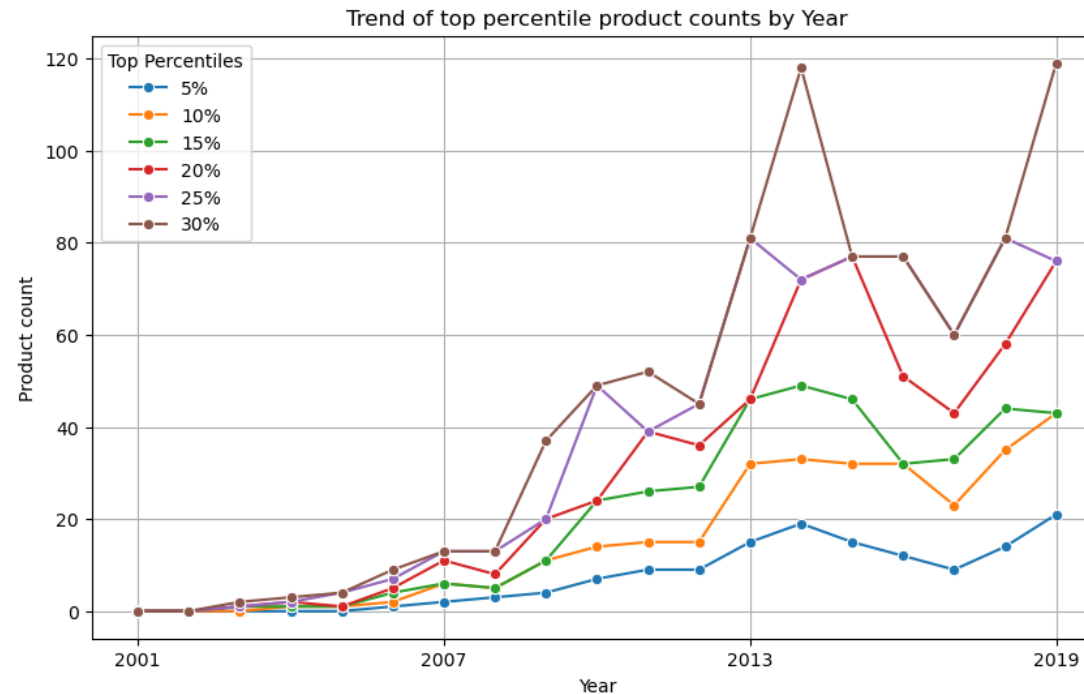
Conceptual framework





Essay III

Benchmark product count trend (by top percentiles)



Product impact refers to the number of descendants a product has within 5 years.

A **benchmark product** is defined as a product that falls within the top percentile of product impact in the same year.

Essay III

Variable definition

Variable Name	Description	Related theories
Count of development	The count of new technologies in the product. I	Knowledge-Based View (Grant, 1996), Creative destruction (Schumpeter, 2013), First-Mover Advantages (Lieberman & Montgomery, 1988; Kerin et al., 1992)
Value of development	The weighted value of the new technologies in the product	Knowledge-Based View (Grant, 1996), Creative destruction (Schumpeter, 1942), First-Mover Advantages (Kerin et al., 1992; Lieberman & Montgomery, 1988)
Count of inheritance	The count of the existing technologies in the product in the same brand	Technological trajectories (Dosi, 1982), Path dependency (Arthur, 1989)
Value of inheritance	The weighted value of the existing technologies in the product in the same brand	Technological trajectories (Dosi, 1982), Path dependency (Arthur, 1989)
Count of transfer	The count of the existing technologies in the product in the other brands	Technology transfer (Alkhazaleh et al., 2022), Technology diffusion (Geroski, 2000; Loch & Huberman, 1999; Rogers et al., 2014)
Value of transfer	The weighted value of the existing technologies in the product in the other brands	Technology transfer (Alkhazaleh et al., 2022), Technology diffusion (Geroski, 2000; Loch & Huberman, 1999; Rogers et al., 2014)
Count of technologies	The count of the total technologies in the product	Technology Sourcing and Combination (Arora et al., 2004; Fleming, 2001)
Value of technologies	The weighted value of the total technologies in the product	Technology Sourcing and Combination (Arora et al., 2004; Fleming, 2001)
Brand difference	The brand difference between the product and its ancestor	Open Innovation (Chesbrough, 2003)
Year difference	The year difference between the product and its ancestor	Technological discontinuities (Tushman & Anderson, 2018), Late mover advantage (Shankar et al., 1998; Zhou, 2002)
Product distance	The technological distance between the product and its ancestor	Technological discontinuities (Tushman & Anderson, 2018), Late mover advantage (Shankar et al., 1998; Zhou, 2002)

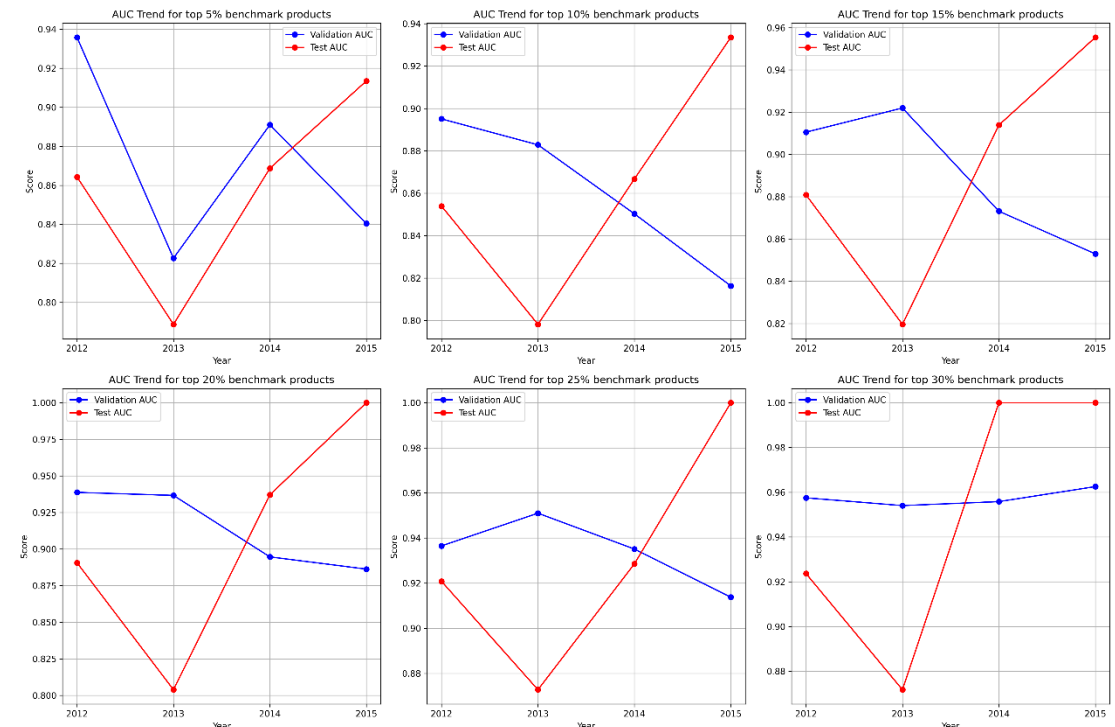
The weight of the technology is calculated using frequency weight.

Essay III

Prediction results

Prediction: Training on year (from 2001 to T), Testing on year T+1, Features: around 20 product variables (strategic variables), Label: Benchmark product dummy

Benchmark products (Top percentile)	AUC of Validation	AUC of Test
5%	0.8732	0.8703
10%	0.8652	0.8803
15%	0.8907	0.8939
20%	0.9140	0.9044
25%	0.9365	0.9352
30%	0.9616	0.9512



AUC stands for **Area Under the Curve**, specifically referring to the area under the **Receiver Operating Characteristic (ROC)** curve. It is a crucial metric for evaluating the performance of binary classification models.

Essay III

Regression results

Logit regressions of Top 5% benchmark product from 2005 to 2016

	(1)	(2)	(3)	(4)	(5)	(6)
	Benchmark product (Top 5%)					
Value of Development	96.096*** (20.8438)	95.114*** (21.0292)	44.8869• (25.876)	96.7693*** (27.768)	54.2482• (29.5873)	54.8671• (29.9711)
Year difference	1.2971*** (0.3065)	1.2563*** (0.3108)	1.7954*** (0.1643)	2.0335*** (0.1885)	0.5677 (0.3631)	0.5652 (0.365)
Value of Transfer		2.7241* (1.3754)	19.1527*** (2.884)	30.6739*** (3.7582)	29.3947*** (4.2776)	29.7932*** (4.3081)
Value of Inheritance			16.8134*** (2.5829)	27.9594*** (3.7376)	26.6958*** (4.1725)	26.9801*** (4.1997)
Technology count				-0.0545*** (0.0154)	-0.0754*** (0.0168)	-0.0775*** (0.017)
Product distance					23.5113*** (3.8035)	24.3384*** (3.8943)
Brand Difference						-0.1819 (0.3177)
Year	✓	✓	✓	✓	✓	✓
Brand	✓	✓	✓	✓	✓	✓
Obs.	1620	1620	1620	1620	1620	1620
Adj. Pseudo R2	0.0784	0.0816	0.1509	0.1824	0.2462	0.2437

Note: Significance Indicators: '***' indicates highly significant ($p < 0.001$), '**' indicates significant ($p < 0.01$), and '*' indicates moderately significant ($p < 0.05$).

Essay III

Regression results

Fixed Effect regressions of bottom percentile from 2005 to 2016

	(1)	(2)	(3)	(4)	(5)	(6)
	Benchmark (bottom percentile)					
Value of development	4.2947*** (0.9299)	4.2652*** (0.9231)	1.6713** (0.6566)	2.1195*** (0.6853)	-0.5390 (0.4598)	-0.5253 (0.4576)
Year difference	0.3058*** (0.0360)	0.3057*** (0.0361)	0.3078*** (0.0317)	0.3077*** (0.0316)	0.0595*** (0.0187)	0.0597*** (0.0185)
Value of transfer		0.0191 (0.0456)	0.5598*** (0.1400)	0.7114*** (0.1647)	0.5351*** (0.0607)	0.5248*** (0.0612)
Value of inheritance			0.5829*** (0.1351)	0.7281*** (0.1598)	0.6116*** (0.0570)	0.6068*** (0.0571)
Technology count				-0.0009*** (0.0003)	-0.0019*** (0.0002)	-0.0019*** (0.0002)
Product distance					2.6139*** (0.1286)	2.5804*** (0.1305)
Brand difference						0.0090 (0.0112)
Year	✓	✓	✓	✓	✓	✓
Brand	✓	✓	✓	✓	✓	✓
Obs.	1771	1771	1771	1771	1771	1771
R ²	0.661	0.661	0.687	0.689	0.870	0.870

Note: Significance Indicators: '***' indicates highly significant ($p < 0.001$), '**' indicates significant ($p < 0.01$), and '*' indicates moderately significant ($p < 0.05$).

Essay III

Robustness

Fixed Effect

Regression results of
impact (in 5 years) of
products by year from
2006 to 2015

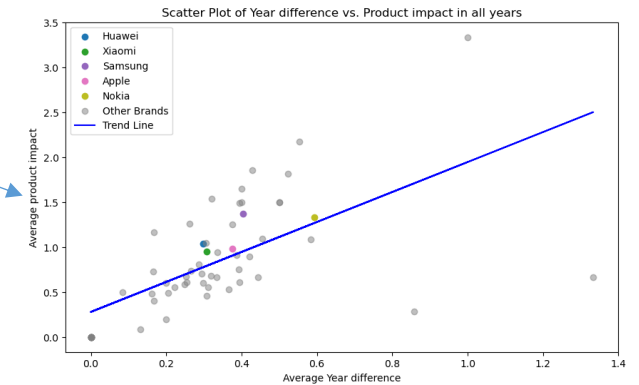
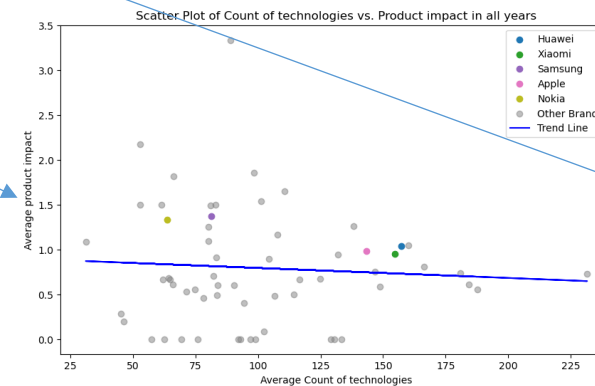
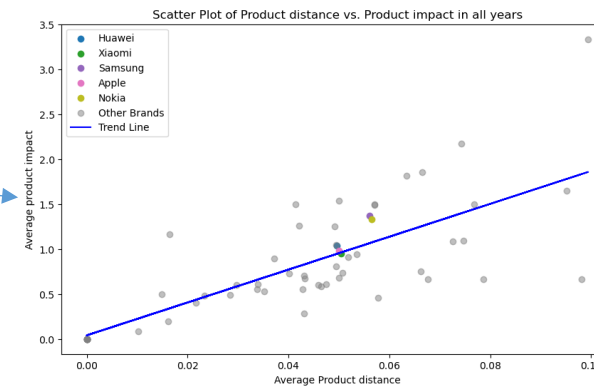
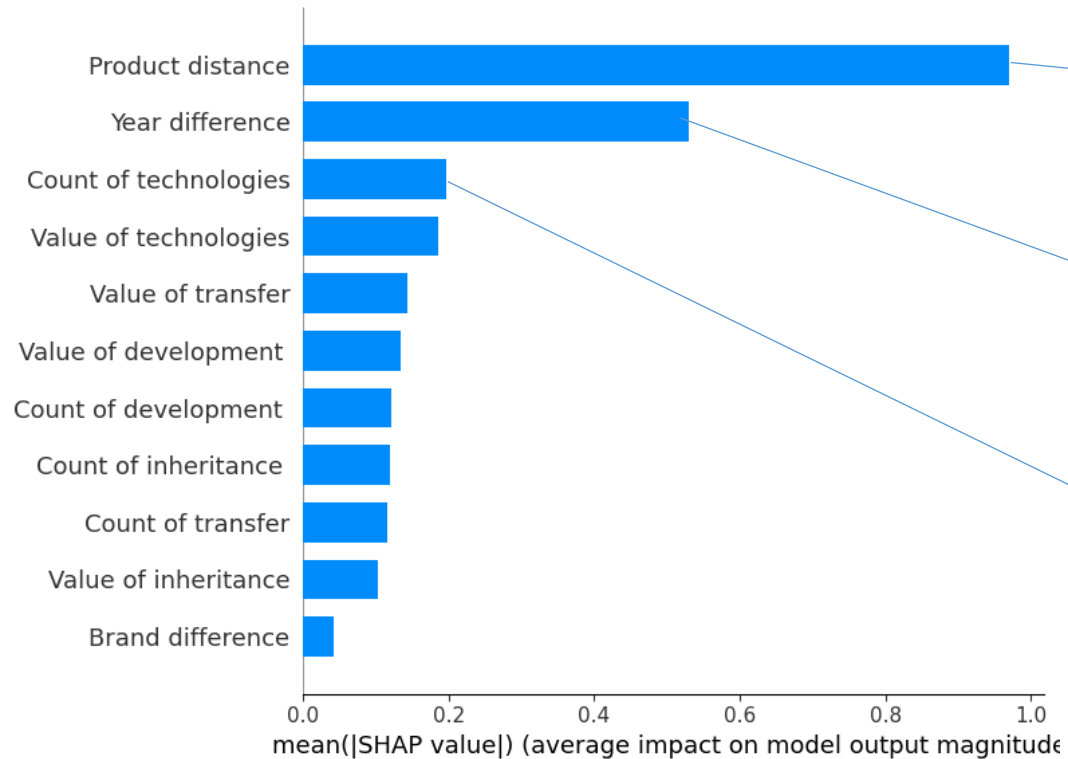
	(1)	(2)	(3)	(4)	(5)	(6)
	Product impacts					
Value of development	58.1585*** (14.4951)	54.5920*** (14.7029)	35.0140** (14.5171)	43.6371*** (15.0404)	24.9516* (13.4659)	24.7457* (13.3499)
Year difference	1.8685*** (0.2501)	1.8475*** (0.2484)	1.8640*** (0.2185)	1.8615*** (0.2156)	0.1168 (0.1313)	0.1138 (0.1325)
Value of transfer		2.3138* (1.3698)	6.3945*** (1.6188)	9.3110*** (1.9926)	8.0721*** (1.6791)	8.2277*** (1.7336)
Value of inheritance			4.3992*** (0.9976)	7.1943*** (1.3206)	6.3749*** (0.8979)	6.4471*** (0.9386)
Technology count				-0.0171*** (0.0045)	-0.0243*** (0.0043)	-0.0249*** (0.0047)
Product distance					18.3722*** (1.4367)	18.8768*** (2.0457)
Brand difference						-0.1361 (0.3272)
Year	√	√	√	√	√	√
Brand	√	√	√	√	√	√
Observations	1771	1771	1771	1771	1771	1771
R2	0.253	0.258	0.270	0.275	0.347	0.347

Note: Significance Indicators: '***' indicates highly significant ($p < 0.001$), '**' indicates significant ($p < 0.01$), and '*' indicates moderately significant ($p < 0.05$).



Robustness

SHAP Values





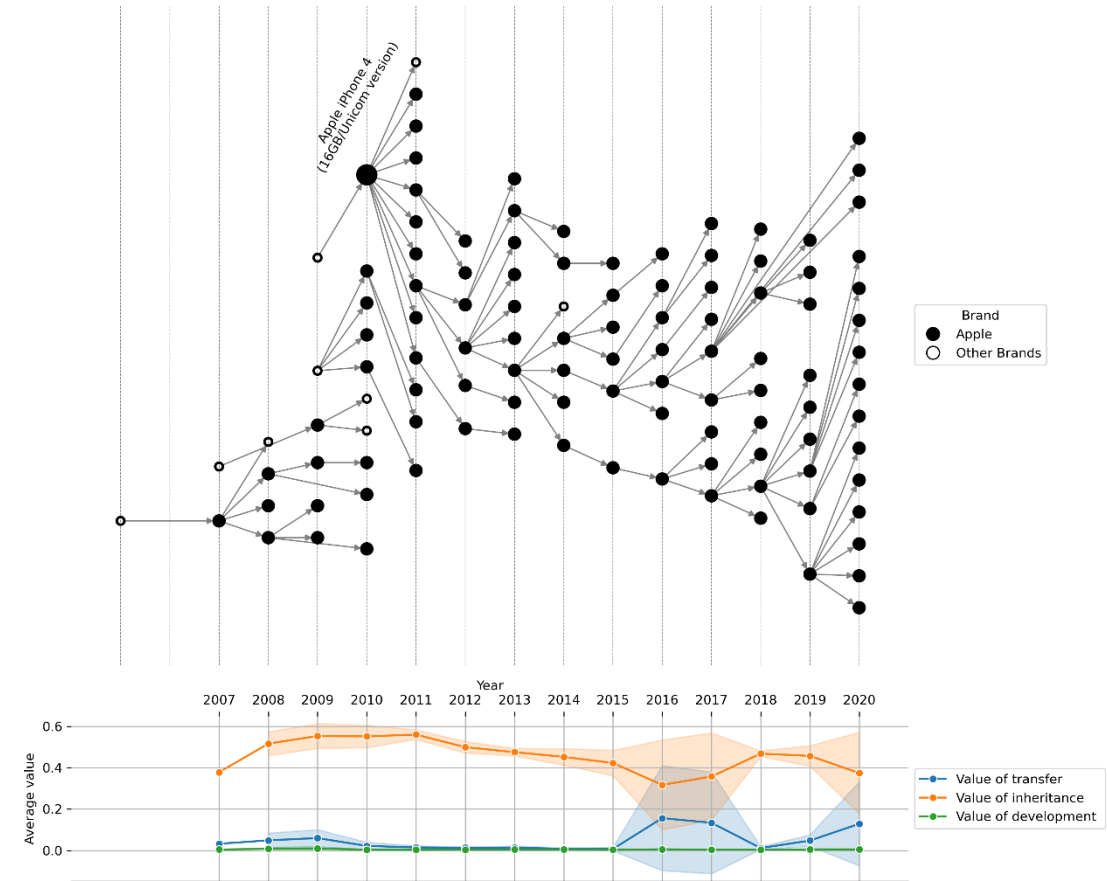
Essay III

Empirical application

Phylogenetic tree of Apple products

Table Average strategic values of Apple products

Year	Value of transfer	Value of inheritance	Value of development
2007	0.0328	0.3784	0.0049
2008	0.0493	0.5164	0.0089
2009	0.0602	0.5536	0.0103
2010	0.0224	0.5517	0.0041
2011	0.015	0.5603	0.0037
2012	0.0141	0.4995	0.0046
2013	0.0141	0.4754	0.0049
2014	0.0082	0.4526	0.0052
2015	0.0079	0.4231	0.0037
2016	0.1566	0.3171	0.006
2017	0.1341	0.3577	0.0044
2018	0.012	0.4682	0.0048
2019	0.0481	0.4573	0.0056
2020	0.1291	0.3747	0.0055



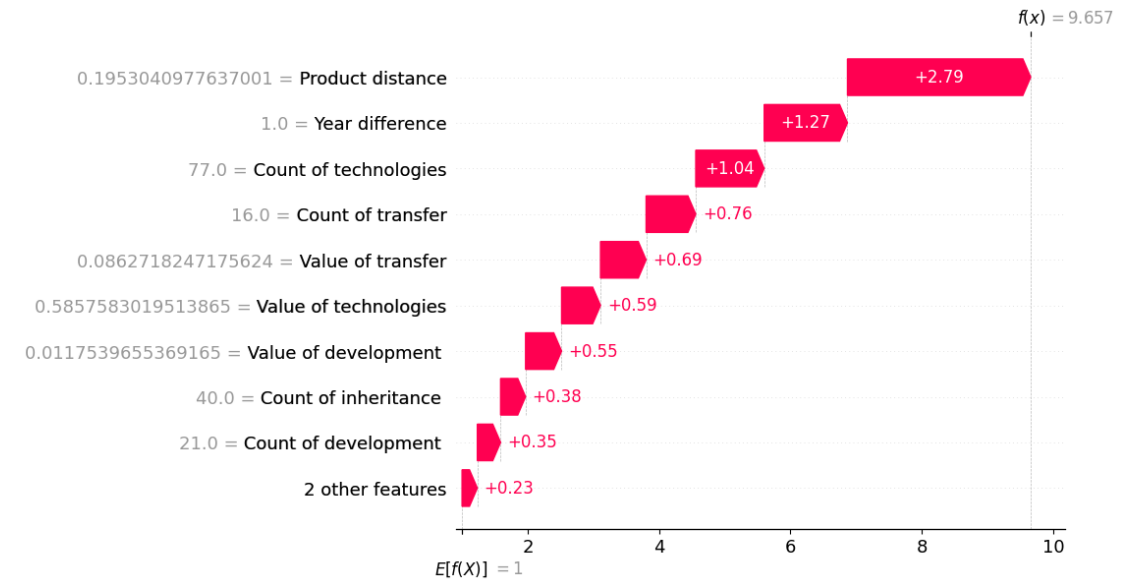


Essay III

Empirical application

Apple iPhone4

Product	Apple iPhone 4 (16GB/Unicom version)
Brand	Apple
Year	2010
Count of development	21
Value of development	0.0118
Count of Inheritance	40
Value of Inheritance	0.4877
Count of transfer	16
Value of transfer	0.0863
Technology count	77
Technology value	0.5858
Top 5%	Yes
Top 10%	Yes
Top 15%	Yes
Top 20%	Yes
Top 25%	Yes
Top 30%	Yes



Implications

This study introduces a method for identifying benchmark products by constructing technological phylogenetic trees. It highlights the product distance relative to ancestors as a key factor influencing its benchmark status from the perspective of technological evolution. The study develops a predictive model for the probability of a product becoming a benchmark, based on various technological and strategic variables. Experimental results demonstrate that this model effectively predicts benchmark products, providing valuable insights for practical product design.

Conclusion

This study makes significant contributions to technology management and forecasting by providing an evolutionary approach to understanding and predicting technological innovation. The distance-based phylogenetic framework enhances both theoretical understanding and practical application, helping organizations anticipate benchmark products and make informed strategic decisions. Additionally, the PDI aids in identifying disruptive products, while the ‘Small but not Least’ (SBNL) design principle offers a technical method for enhancing product disruptiveness. This approach addresses gaps in existing theories and aligns product innovation with scientific advancements.



End of Presentation

Thank you😊

E: heyounwei2022@gmail.com

G: <https://heyounwei2021.github.io/>

Bibliography

- Anderson, P., & Tushman, M. L. (2018). Technological discontinuities and dominant designs: A cyclical model of technological change. In *Organizational Innovation* (pp. 373-402). Routledge.
- Arthur, W. B. (2009). *The nature of technology: What it is and how it evolves*. Penguin UK.
- Bain, J. S. (1951). Relation of profit rate to industry concentration: American manufacturing, 1936–1940. *The Quarterly Journal of Economics*, 65(3), 293-324.
- Basalla, G. (1988). *The evolution of technology*. Cambridge University Press.
- Bowler, P. J. (1989). *Evolution: the history of an idea*. Univ of California Press.
- Cecere, G., Corrocher, N., & Battaglia, R. D. (2015). Innovation and competition in the smartphone industry: Is there a dominant design? *Telecommunications Policy*, 39(3-4), 162-175.
- Cecere, G., Corrocher, N., Gossart, C., & Ozman, M. (2014). Lock-in and path dependence: an evolutionary approach to eco-innovations. *Journal of Evolutionary Economics*, 24, 1037-1065.
- Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*,
- Christensen, C. M. (1992). Exploring the limits of the technology S-curve. Part I: component technologies. *Production and operations management*, 1(4), 334-357.
- Cimoli, M., & Dosi, G. (1995). Technological paradigms, patterns of learning and development: an introductory roadmap. *Journal of Evolutionary Economics*, 5, 243-268.
- Darwin, C. (1859). *On the origin of species by means of natural selection* (Vol. 167). John Murray, London.

Bibliography

- David, P. A. (1985). Clio and the Economics of QWERTY. *The American economic review*, 75(2), 332-337.
- Felsenstein, J. (1985). Phylogenies and the comparative method. *The American Naturalist*, 125(1), 1-15.
- Fleming, L., & Sorenson, O. (2001). Technology as a complex adaptive system: evidence from patent data. *Research policy*, 30(7), 1019-1039.
- Fogel, R. W. (1999). Catching up with the economy. *American Economic Review*, 89(1), 1-21.
- Fortunato, S., Bergstrom, C. T., Börner, K., Evans, J. A., Helbing, D., Milojević, S., Petersen, A. M., Radicchi, F., Sinatra, R., & Uzzi, B. (2018). Science of science. *Science*, 359(6379), eaao0185.
- Foster, J. G., Rzhetsky, A., & Evans, J. A. (2015). Tradition and innovation in scientists' research strategies. *American sociological review*, 80(5), 875-908.
- Foxon, T. J. (2014). Technological lock-in and the role of innovation. In *Handbook of sustainable development* (pp. 304-316). Edward Elgar Publishing.
- Frank, R., & Cook, P. (1995). It's a Winner-Takes-All Market—Top Money Goes to Top Performing People or Products. *Washington Monthly*.
- Frenkel, A., & Maital, S. (2014). Mapping national innovation ecosystems: Foundations for policy consensus. Edward Elgar Publishing.
- Hodgson, G. M., & Knudsen, T. (2010). Darwin's conjecture: The search for general principles of social and economic evolution. University of Chicago Press.
- Katz, M. L., & Shapiro, C. (1985). Network externalities, competition, and compatibility. *The American economic review*, 75(3), 424-440.
- Kemp, R., Schot, J., & Hoogma, R. (1998). Regime shifts to sustainability through processes of niche formation: the approach of strategic niche management. *Technology analysis & strategic management*, 10(2), 175-198.
- Knudsen, T. (2002). Economic selection theory. *Journal of Evolutionary Economics*, 12, 443-470.

Bibliography

- Lee, J.-D., Jeong, D., Jung, E.-Y., Kim, Y., Kim, J., He, Y., & Choi, S. (2022). Mapping the evolutionary pattern of mobile products: a phylogenetic approach. *IEEE Transactions on Engineering Management*, 71, 4776-4790.
- Linzhuo, L., Lingfei, W., & James, E. (2020). Social centralization and semantic collapse: Hyperbolic embeddings of networks and text. *Poetics*, 78, 101428.
- Loch, C. H., & Huberman, B. A. (1999). A punctuated-equilibrium model of technology diffusion. *Management science*, 45(2), 160-177.
- Mayr, E. (2001). *What evolution is* Basic Books. New York, 349, 349.
- Mehmood, T., Alzoubi, H. M., & Ahmed, G. (2019). Schumpeterian entrepreneurship theory: Evolution and relevance. *Academy of Entrepreneurship Journal*, 25(4).
- Mokyr, J. (1990). Punctuated equilibria and technological progress. *The American economic review*, 80(2), 350-354.
- Murray, F. (2002). Innovation as co-evolution of scientific and technological networks: exploring tissue engineering. *Research policy*, 31(8-9), 1389-1403.
- Nelson, R. R. (1985). *An evolutionary theory of economic change*. harvard university press.
- Nelson, R. R., & Winter, S. G. (1982). The Schumpeterian tradeoff revisited. *The American economic review*, 72(1), 114-132.
- Norman, D. A., & Verganti, R. (2014). Incremental and radical innovation: Design research vs. technology and meaning change. *Design issues*, 30(1), 78-96.
- Pareto, V. (1964). *Cours d'économie politique* (Vol. 1). Librairie Droz.
- Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. *arXiv preprint arXiv:1908.10084*.
- Schulz, J., & Mayerhoffer, D. M. (2021). Equal chances, unequal outcomes? Network-based evolutionary learning and the industrial dynamics of superstar firms. *Journal of Business Economics*, 91(9), 1357-1385.
- Schumpeter, J. A. (2013). *Capitalism, socialism and democracy*. routledge.

Bibliography

- Shi, F., & Evans, J. (2023). Surprising combinations of research contents and contexts are related to impact and emerge with scientific outsiders from distant disciplines. *Nature Communications*, 14(1), 1641.
- Srinivasan, R., Lilien, G. L., & Rangaswamy, A. (2006). The emergence of dominant designs. *Journal of marketing*, 70(2), 1-17.
- Stigler, G. J. (1958). The economies of scale. *The Journal of Law and Economics*, 1, 54-71.
- Straffon, L. M. (2016). The applications and challenges of cultural phylogenetics in archaeology: An introduction. *Cultural phylogenetics: Concepts and applications in archaeology*, 1-15.
- Teece, D. J. (2007). Explicating dynamic capabilities: the nature and microfoundations of (sustainable) enterprise performance. *Strategic management journal*, 28(13), 1319-1350.
- Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic management journal*, 18(7), 509-533.
- Tushman, M. L., & Anderson, P. (2018). Technological discontinuities and organizational environments. In *Organizational innovation* (pp. 345-372). Routledge.
- Ulrich, K. T., & Eppinger, S. D. (2016). *Product design and development*. McGraw-hill.
- Wagner, A., & Rosen, W. (2014). Spaces of the possible: universal Darwinism and the wall between technological and biological innovation. *Journal of the Royal Society Interface*, 11(97), 20131190.
- Wu, L., Wang, D., & Evans, J. A. (2019). Large teams develop and small teams disrupt science and technology. *nature*, 566(7744), 378-382.
- Youn, H., Strumsky, D., Bettencourt, L. M., & Lobo, J. (2015). Invention as a combinatorial process: evidence from US patents. *Journal of the Royal Society Interface*, 12(106), 20150272.



Bibliography

Ziman, J. M. (2003). Technological innovation as an evolutionary process. Cambridge University Press.

Zipf, G. K. (2016). Human behavior and the principle of least effort: An introduction to human ecology. Ravenio books.

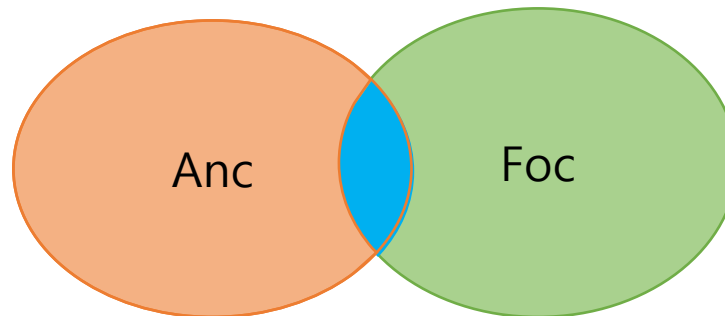



Annexes

Inherited gene rate: $\frac{Anc \cap Foc}{Foc}$

New gene rate: $\frac{Anc - Foc}{Foc}$

Lost gene rate: $\frac{Anc - Foc}{Anc}$



$Anc - Foc$: 

$Anc \cap Foc$: 



The definition of technological phylogenetic tree and network

- 1) Tree: $T = G(V, E)$ and Network: $N = G(V, E)$, while T is a directed acyclic graph (DAG) with the limitation that one node can only has one ancestor, and N is a directed acyclic graph without any limitation.
- 2) Distance-based method for inferring the optimum technological phylogenetic tree and network (Occam's razor):

Definition: The parsimony score of a phylogenetic tree is the sum of the substitution costs along all the edges of the tree

$$PS(T, X) = \sum_{e_{i,j} \in E} d(e_{i,j}), \text{ subject to } t_j < t_i, \text{ and } e_{i,j} \in E, \text{ and } i, j \in V,$$

Where:

$PS(T, X)$ is the parsimony length of the phylogeny.

X is the vector of the embedding or the chromosome of genes

e is an edge in the phylogenetic tree T .

$d(e_{i,j})$ is the distance of edge $e_{i,j}$.

i is the focal node's index

j is any node's index which appeared before the focal node i



The objective function of technological phylogenetic tree and network

The objective function of optimum technological phylogenetic tree refers to:

$$T^* = \operatorname{argmin}_T [PS(T, X)] = \operatorname{argmin}_T \left[\sum_{i=1}^m d(e_{i,j}) \right], \text{ subject to } t_j < t_i, \text{ and } e_{i,j} \in E, \text{ and } i, j \in V,$$

The objective function of the optimum technological phylogenetic network refers to:

$$N^* = \operatorname{argmin}_N [PS(N, X)] = \operatorname{argmin}_N \left[\sum_{i=1}^m \sum_{k=1}^n d(e_{i,j_k}) \right], \text{ subject to } t_{j_k} < t_i \forall k, \text{ and } e_{i,j_k} \in E, \text{ and } i, j_k \in V,$$

Where:

N^* represents the optimized network structure. m is the node count in V .

j_1, j_2, \dots, j_n are the indexes of the nodes corresponding to the n smallest distances for each focal node i .

The additional sum \sum_k^n represents the selection of the n edges with the smallest distances for each focal node i .

$d(e_{i,j_k})$ is the distance associated with edge e_{i,j_k} .