Preprints.org

Article

Not peer-reviewed version

# Knowledge Evolution in the Mobile Industry via Embedding-Based Topic Growth and Typology Analysis

SungJin Jeon [†] , Woojun Jung [†] , Keuntae Cho [*]

Posted Date: 11 March 2026

doi: 10.20944/preprints202603.0320.v2

Keywords: mobile industry; regime; embedding; lineage; growth type

*Article*

# Knowledge Evolution in the Mobile Industry via Embedding-Based Topic Growth and Typology Analysis

**SungJin Jeon [1,†], Woojun Jung [1,†] and Keuntae Cho [2,*]**

[1] Graduate School of Management of Technology, Sungkyunkwan University, Suwon 16419, Republic of Korea
[2] Department of Systems Management Engineering, Sungkyunkwan University, Suwon 16419, Republic of Korea
* Correspondence: ktcho@skku.edu; Tel.: +82-031-290-7602
† These authors contributed equally to this work.

**Abstract**

The mobile industry has experienced long-run changes in its knowledge structure, including identifiable transition points observable through meaning-based analysis. Using abstracts from 86,674 mobile-industry publications published between 2005 and 2024, we embed documents with SPECTER2, build year-specific embedding distributions, and derive knowledge regimes by combining change-point detection with inter-year distribution distances. We then extract regime-specific topics via clustering and reconstruct topic lineages by aligning topic similarities to classify inheritance, differentiation, convergence, and disappearance. The analysis delineates three regimes spanning 2005 to 2012, 2013 to 2019, and 2020 to 2024, with pronounced transitions around 2012 to 2013 and 2019 to 2020. Regime 1 centers on foundational technologies such as wireless communication, power, sensors, and reliability. Regime 2 expands toward platforms, apps, and data analytics alongside cross-domain convergence. Regime 3 is characterized by strengthened 5G operations and data-driven services, together with the independent rise of policy, governance, and regulation topics. Transitions reflect recombination built on inherited knowledge rather than abrupt replacement, and post-transition topics display distinct growth typologies by network position and growth pattern. By integrating embedding-based change-point detection with topic-lineage reconstruction, we provide a reproducible account of regime transitions and quantitative evidence to inform the timing of corporate R&D, standard and platform strategies, and policy and regulatory design.

**Keywords:** mobile industry; regime; embedding; lineage; growth type

---

## 1. Introduction

The mobile industry has shifted from handset centered competition to platform competition in which operating systems, app stores, and developer ecosystems shape the direction of innovation.[1,2] Even as functional convergence has progressed, the smartphone industry has continued to exhibit persistent product differentiation across manufacturers, and a single dominant design has not fully stabilized.[3] Platform diffusion dynamics interact with ecosystem feedback mechanisms such as network effects, implying that innovative success can be determined not only by device performance but also by the interplay between complementary supply and user scale.[4] App stores have also evolved beyond distribution channels into governance devices that reshape market access and innovation incentives through rule setting and enforcement, becoming a central agenda in debates on digital governance and regulation.[5] As regulatory regimes such as the European Union Digital

 doi:10.20944/preprints202603.0320.v2



Markets Act, DMA, increasingly adopt ex ante constraints on gatekeeper platforms, the institutional environment surrounding app store governance is also changing.[6]

With the introduction of fifth generation networks, network computing integration paradigms such as mobile edge computing have expanded, reshaping service architectures and operational modes.[7] Relatedly, the rise of discussions on zero touch management oriented toward autonomous network operations indicates that operational automation and data driven decision making have become core competitive factors in communications infrastructure.[8] The mobile industry therefore represents a setting in which technological generation shifts in networks and devices, platform and ecosystem shifts in the app economy and revenue models, and institutional and governance shifts in rules and regulation overlap, creating a high likelihood of transition periods in which the knowledge system is reallocated and concentrated.[5,8]

In such a rapidly evolving environment, tracing change solely through market outcomes or product launch data tends to rely on indicators observable only after transitions have materialized, making it difficult to detect early signals. Scholarly knowledge, by contrast, constitutes an upstream layer in which new concepts, methods, and evaluative frames accumulate, allowing early signals of industrial change to surface first in text.[9] Prior work that analyzes mobile ecosystem knowledge flows through patent citation networks illuminates interfirm knowledge movement and structural change, yet additional approaches are needed to reconstruct how topical content is reorganized at the textual level in an integrated manner.[10] Moreover, in a digital economy where data driven value creation is increasingly salient, issues such as data access, platform dominance, and policy can become intertwined with the competitive structure of the mobile industry.[11]

Platform policy shifts such as tracking restrictions in iOS increasingly need to be interpreted as transitions that combine technological and institutional change.[12] Studies on the relationship between targeting efficiency and privacy in mobile advertising further suggest that constraints on data access can affect performance and competition, supporting the view that changes in data governance may intersect with industrial innovation pathways.[13] Nonetheless, much of the existing mobile research focuses on specific technological domains or limited time windows, leaving insufficient integrated quantitative evidence on when the industry knowledge system undergoes structural transitions over the long run and which topics move toward the central axis after such transitions.[9,10] Capturing transitions requires moving beyond keyword frequencies to examine distributional shifts in a document level semantic space, and scientific literature embedding models in the SPECTER family provide a useful foundation by offering document level meaning representations at scale.[14,15] Accordingly, there is a need for research that tracks mobile industry knowledge flows through year to year changes in semantic distributions, objectively identifies transition points, and systematically reconstructs how topical structures are reallocated across transitions, including centrality shifts and processes of convergence and differentiation.

Research on mobile industry trends often summarizes distributions through publication growth, keyword frequencies, or citation networks, which limits the ability to explicitly identify structural transition points in the knowledge system. Even when topic modeling is applied, studies that reconstruct how topics persist through time as lineages, how they split through differentiation, how they merge through convergence, and what disappears are relatively rare. In domains such as the mobile industry where technology, platforms, and institutions become rapidly entangled, presenting a list of topics is insufficient to explain the meaning of transition periods. It is necessary to represent the relational structure among topics and their movement pathways together.

Distinguishing growth typologies after transitions can clarify the nature of change and its strategic implications by indicating which topics move toward the core of the knowledge system, which persist in the periphery, and which surge during specific periods. This motivates an integrated approach that constructs year specific knowledge distributions from embedding based document representations, identifies transition points from distributional shifts, and links topic structures before and after transitions into lineages. Mobile industry evolution should not be viewed as the cumulative growth of a single technology. Rather, it progresses through the reallocation of

 doi:10.20944/preprints202603.0320.v2



knowledge exploration and exploitation, the emergence of new problem frames enabled by knowledge recombination, and interactions with shifting platform rules and institutional environments. Building on this perspective, we interpret regime transitions as structural changes in innovation phases and treat topic lineage events such as inheritance, convergence, and differentiation as observable indicators of knowledge recombination.

Academically, the study offers an integrated framework that explains long run knowledge change in the mobile industry from a structural perspective centered on distributional transitions and the reconfiguration of topic lineages rather than reducing it to a growth narrative of specific technologies. Methodologically, it combines embedding based transition point detection with topic alignment based lineage reconstruction to provide a reproducible procedure for identifying the existence and character of transition periods. Practically, by distinguishing topics that move toward the central axis after transitions from those that persist at the periphery, the findings provide evidence that can support corporate R and D prioritization, platform and standard strategies, and exploration of collaboration and investment opportunities. From a policy perspective, the results provide quantitative implications that can inform the timing of support and regulatory design and the prioritization of policy agendas, grounded in structural shifts such as the rise of data, governance, and regulation after the fifth-generation transition.

Against this backdrop, the present study addresses three interconnected research questions. RQ1: At what temporal boundaries does the mobile industry knowledge system undergo statistically verifiable distributional transitions, and what auxiliary structural indicators---centroid displacement and dispersion change---characterize the nature of each transition? RQ2: How do individual topics persist, merge, split, emerge, or disappear across regime boundaries, and what weighted-contribution criteria enable conservative, reproducible classification of these evolutionary events? RQ3: What two-dimensional typology, combining structural centrality in the topic transition network with temporal growth intensity, best differentiates topics whose growth reflects core architectural recon-figuration from those reflecting peripheral issue-driven surges, and what strategic and regulatory prescriptions follow from each type?

These research questions collectively address the three gaps identified in the literature: the absence of distributional, embedding-based transition detection for mobile industry knowledge (Gap 1); the lack of a conservative, size-weighted topic lineage reconstruction procedure covering inheritance, differentiation, convergence, and disappearance (Gap 2); and the unavailability of a multidimensional growth typology integrating structural network position with temporal growth pattern (Gap 3). The proposed framework is reproducible, scalable to corpora of 86,000-plus documents, and yields results that are robust across 432 hyperparameter combinations with 97.2% boundary stability, as demonstrated in the sensitivity analysis reported in Section 5.3.

The remainder of this paper is organized as follows. Section 2 reviews the theoretical background and prior studies on the mobile industry knowledge structure, platform ecosystems, transition point analysis, and topic evolution and lineage reconstruction. Section 3 describes the data collection and preprocessing, document embedding construction, transition point detection procedure, regime specific topic extraction, and the methods for topic alignment and growth typology classification. Section 4 presents the results on transition points and regime identification, regime specific topic structures, and analyses of topic lineages and growth typologies. Section 5 discusses the findings and derives implications from academic, industrial, and policy perspectives. Section 6 concludes by summarizing contributions, discussing limitations, and suggesting directions for future research.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 March 2026 doi:10.20944/preprints202603.0320.v2



## 2. Literature Review

### *2.1. Research on the Mobile Industry*

The mobile industry exhibits a multi sided platform structure that simultaneously matches users and complementors, and the theoretical foundation is two sided market theory, which emphasizes that cross side network effects shape market competition and the pace of innovation.[16–18] Platforms coordinate participation on both sides through pricing, access, and rules, and they design both value creation and value capture, providing a core analytical lens for understanding the mobile OS app store payment and advertising and developer ecosystem.[18] In such markets, competition can be reconfigured less by share levels per se than by the process through which platform boundaries expand and overlap, and platform envelopment strategies that absorb adjacent markets can rapidly reorder industrial boundaries and innovation trajectories.[19]

Platform research has extended into ecosystem theory, strengthening the view that innovation outcomes depend not only on focal firm capabilities but also on interdependence among components and coordination failures.[20,21] Ecosystems are analyzable as structures rather than as simple networks, and accumulated discussions show that competitive advantage varies with which components perform which roles and where bottlenecks arise.[21,22] Industrial platforms can induce external innovation and accelerate ecosystem level innovation through modularity, interface design, and governance, which has refined platform leadership and orchestration as central strategic themes.[23,24]

As platforms grow, their relationships with complementors involve not only collaboration but also competition, and tensions and shifts in innovation incentives can arise when platforms enter complementor domains.[25] Platforms must also choose strategic tradeoffs between openness and control, and governance, rules, and fee structures can structurally reallocate participant behavior within ecosystems.[26,27] Within this setting, antitrust frameworks and competition policy debates for multi sided platforms require regulatory logics distinct from those of traditional industries, motivating calls for competition policy frames tailored to the digital era.[28,29] Work that systematically integrates the platform competition literature synthesizes these core issues such as competition, dominance, governance, ecosystems, and complementor strategies and provides coordinates for research design.[30]

Finally, the mobile industry has faced an increased likelihood of transition periods in which technological, platform, and institutional elements change simultaneously, as the generational shift from 4G to 5G has redefined networks as foundational infrastructures for services and operations.[31,32] Edge computing in particular alters constraints related to latency, bandwidth, and service design through tighter coupling of networks and computing, and it is summarized as a driver that can trigger structural change in mobile service architectures and operating modes.[33–35] On the operational side, standardization and automation discussions such as zero touch service management indicate that data driven operational paradigms have emerged as core competitive factors in communications infrastructure.[36]

### *2.2. Bibliometric Studies and Knowledge Flow Research in the Mobile Industry*

In domains such as the mobile industry where technological and service generations turn over rapidly, bibliometrics and science mapping have been widely used to capture the accumulation and diffusion of knowledge quantitatively. From a network perspective, scholarly knowledge can be represented through citation relations among papers and patents, and general theories for analyzing such multiplex networks provide a basis for quantitatively summarizing science and technology knowledge flows.[37] Network analytic methodologies have also been refined to explain how knowledge is organized and connected using structural indicators such as centrality, betweenness, and community structure.[38]

Two representative linkage rules in bibliometric networks are bibliographic coupling and co citation. Bibliographic coupling assumes that two documents are more similar when they share the

 doi:10.20944/preprints202603.0320.v2



same references, and there is classical work proposing linkage rules for coupling across documents.[39] Co citation defines similarity by how frequently two documents are cited together by third documents, and it has become a core tool in the analytic tradition that distinguishes research fronts from intellectual bases to characterize knowledge structures.[40] These linkage rules provide static similarity, but reconstructing how knowledge flows unfold along temporal paths requires explicit path extraction.

Main path analysis was proposed to identify the central paths along which knowledge actually flows in citation networks, and approaches that reconstruct core streams based on connectivity structures have continued to develop.[41] Efficient algorithms and implementations that enable main path analysis on large scale networks have subsequently diffused across diverse science and technology fields.[42] Studies that reconstruct technological trajectories using patent citation networks have highlighted both the usefulness of main path analysis and limitations such as missing important nodes and producing overly complex paths, thereby proposing directions for improvement.[43] Related approaches that interpret patent citations in staged ways to quantify inventive progress can also be understood as efforts to jointly explain paths and stages of technological evolution.[44] More recently, studies have proposed improvements to patent based main path extraction or generalized procedures, further expanding the empirical toolbox for analyzing technological evolution.[45] In the 5G domain as well, research combining patent citation networks with main path analysis has emerged to trace technological development streams, accumulating analytic grounds that connect generational transitions with knowledge flows in the mobile industry.[46]

Knowledge structure change can appear not only as paths but also as abrupt surges, and a representative quantitative device for capturing such surges is burst detection. Burst detection algorithms that model the short term intensification of specific terms or topics in document streams have been used as a methodological basis for detecting trend transitions across many fields.[47] Visual analytic tools that map temporal variation in research fronts and intellectual bases and detect clues of transitions through measures such as centrality based pivotal points have also become representative approaches supporting the interpretation of knowledge structure change.[48]

From an empirical standpoint, these methodologies have been applied to quantify research fronts and hot topics in specific themes such as 5G security and 5G applications.[49] However, in domains like the mobile industry where industrial, standards, patent, and scholarly knowledge accumulate simultaneously, it is difficult to capture the direction and speed of innovation using a single data source, increasing the need to jointly design patent citation based network analysis with bibliometric and text analytic approaches. Research monitoring brokerage roles in patent citation networks from an open innovation perspective quantitatively shows that knowledge movement across firms and technologies can be linked to innovation strategy.[50] The conceptual framework of open innovation further emphasizes that innovation activities have shifted from closed internal R and D toward actively combining external knowledge and pathways, and it has been used as a background theory for industry ecosystem analysis.[51]

Finally, recent bibliometric analysis has improved substantially in reproducibility and scalability alongside advances in tool ecosystems. VOSviewer has been widely used as a representative tool for visualizing and mapping large scale bibliometric networks.[52] Theoretical formalizations that seek to integrate mapping and clustering under a unified principle have also been proposed, providing a basis for improving the consistency of tool based results.[53] Bibliometrix provides an R based open source workflow for conducting science mapping, offering an integrated procedure that spans data processing, analysis, and visualization.[54] In sum, quantitative prior research on mobile industry knowledge flows has expanded toward deriving knowledge structures based on bibliographic coupling and co citation, reconstructing trajectories via citation and patent networks, detecting transition clues through bursts and visual analytics, and building reproducible workflows through tool based pipelines.

### 2.3. Methodological Review

#### 2.3.1. Change Point Detection

To argue for regime shifts in knowledge structures within long run corpora, it is necessary to detect changes in distributions themselves rather than changes in publication volume or keyword frequency. In multivariate settings, nonparametric methods that estimate multiple change points can directly infer when regimes change from data, providing a core foundation for transition analysis.[55] Change point detection has also advanced in computational efficiency, and approaches that detect optimal change points with linear cost have been proposed, expanding feasibility for large scale time series applications.[56] Reviews synthesizing change point detection methods emphasize that method choice should depend on data characteristics and objectives, such as offline versus online settings and univariate versus multivariate cases.[57]

#### 2.3.2. Measuring Distributional Distance

Quantifying the strength of transitions requires statistics that measure distributional differences, and distance based statistics such as energy statistics provide a framework that expresses distributional differences as distances.[58] Results showing theoretical connections between distance based tests and reproducing kernel Hilbert space based tests provide justification for method selection in distribution testing.[59] Maximum Mean Discrepancy, formulated as a kernel based two sample test, can capture distributional differences broadly and can therefore be used to support the magnitude and direction of candidate transition periods.[60]

#### 2.3.3. Embedding Based Topic Modeling

Topic modeling has become a representative technique for extracting themes from science and technology literature.[61] Dynamic topic models that explicitly incorporate time have also been proposed, and efforts to model long run trends have continued.[62,63] More recently, advances in contextual embeddings have reshaped topic construction methods. Pretrained BERT family models substantially improved language representations,[64] and pretrained models specialized for scientific text have also been proposed, strengthening the foundation for analyzing science and technology documents.[65]

Embedding based topic construction relies on dimensionality reduction and clustering as core procedures, and UMAP has been widely used to transform high dimensional semantic space into representations that are suitable for visualization and clustering.[66] HDBSCAN is well suited to text corpora because it can detect topics of varying sizes and densities without pre specifying the number of clusters, and it provides practical advantages by separating noise while forming stable clusters.[67,68] BERTopic combines embedding, dimensionality reduction, clustering, and class based TF IDF to construct interpretable topics in large scale text, and it has become a representative implementation of embedding based topic modeling.[69]

#### 2.3.4. Topic Evolution and Lineage Reconstruction

Explaining topic evolution requires the perspective that topics change not only by increasing or decreasing but also through events such as emergence, disappearance, differentiation, and convergence. Approaches have been proposed to visually represent topic flows to make complex changes interpretable,[70] and work that structurally connects transitions among hierarchical topics provides a logical basis for topic lineage reconstruction.[71] Reviews of the development of topic modeling organize model families and evaluation issues and emphasize the need for procedure design aligned with research purposes such as exploration, tracking, and explanation.[72]

 doi:10.20944/preprints202603.0320.v2



*2.4. Limitations of Prior Studies and Research Gaps*

The review of prior work suggests three research gaps. First, studies on knowledge flows in the mobile industry have relied largely on patent citation networks or keyword frequency analysis, and approaches that directly detect distributional shifts in document level semantic space remain insufficient. This limits the ability to capture early signals of transitions and to identify transition periods objectively.[9,10] Second, topic modeling studies often present topic structures at specific points in time, leaving a lack of research that reconstructs topic lineages across time by integrating inheritance, differentiation, convergence, and disappearance. In particular, conservative decision procedures that use weighted contribution to avoid over identifying merge and split events have not yet been systematically established. Third, topic growth is often evaluated using a single indicator such as increases in document counts, and multidimensional typology systems that combine structural position in transition networks with temporal growth patterns remain absent. As a result, interpretive frames are limited in distinguishing reconfigurational surges within the core from issue driven surges in the periphery.

These three gaps collectively point to a deeper epistemological limitation in prior mobile industry knowledge studies: the inability to distinguish between gradual cumula-tive growth within a stable paradigm and punctuated structural reconfiguration that re-orients the knowledge system toward a new problem center. This distinction has direct implications for technology management: firms that cannot differentiate between incre-mental and architectural knowledge change are prone to under-invest during bursty re-configuration phases and over-invest in mature peripheral domains. The present study's integrated framework---combining distributional regime detection, conservative lineage reconstruction, and multidimensional growth typology---is designed to provide precisely this discriminatory capacity at the level of a large-scale longitudinal corpus.

A fourth gap is also identified. No prior bibliometric study has explicitly connected embedding-based topic centrality to Teece's (1986) co-specialized assets framework, linked topic spike patterns to real options theory (McGrath, 1997), or mapped peripheral bursty trajectories onto Christensen's (1997) disruptive innovation model. Bridging this theoretical gap is a secondary contribution of the present study, pursued through the three formal propositions derived in Section 5.1 and the typology-specific strategic prescriptions in Section 5.2. Together, these contributions position the study at the intersection of computational bibliometrics and strategic management of technology, advancing both the methodological toolkit and the theoretical vocabulary available for analyzing knowledge evolution in high-velocity industries.

To address these gaps, this study proposes a methodological framework that integrates E Divisive and MMD based regime detection, topic alignment that combines similarity with weighted contribution, and a two by two typology that jointly considers structural position and growth patterns.

## 3. Materials and Methods

The purpose of this study is to empirically examine how scholarly knowledge in the mobile industry accumulates over time, when it is structurally reconfigured, and which thematic strands subsequently converge toward the core or persist at the periphery. To this end, we construct a large scale longitudinal text corpus based on Web of Science abstracts published between 2005 and 2024 and implement an integrated analytical pipeline consisting of four steps, as summarized in Figure 1. First, we quantify year specific knowledge distributions by generating document embeddings. Second, we detect transition periods and segment the observation window into regimes based on distributional shifts. Third, we construct regime specific topic structures. Fourth, we align topics across regimes to derive topic lineages and classify topic growth typologies by jointly considering structural position and temporal growth patterns.

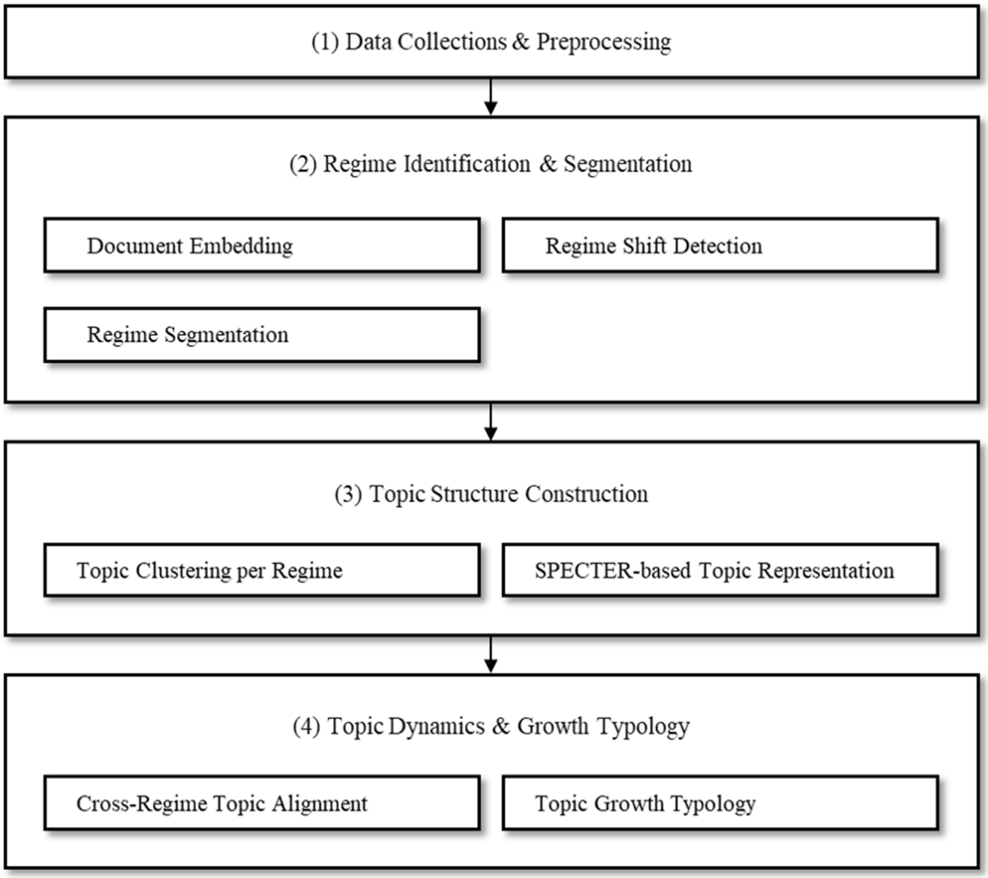


**Figure 1.** Summary of research procedure.

### *3.1. Data Collection and Preprocessing*

#### 3.1.1. Data Collection

This step constructs the literature corpus required for the analysis. Building on earlier prior studies, we first organized the mobile industry literature into analytical categories and then developed a structured block based search query accordingly. As shown in Table 1, the target set is restricted to scholarly articles related to the mobile industry and the smartphone industry, identified through this query. To ensure reliability and reproducibility, we use the Web of Science Core Collection as the data source. The analysis period spans 2005 to 2024 to capture long run dynamics in which technological generational shifts and industrial restructuring accumulate over time. We restrict document type to Article and extract bibliographic information required for subsequent analyses, including authors, titles, abstracts, keywords, journals, affiliations, and citation metadata. This process yields a final dataset of 86,674 records.

#### 3.1.2. Preprocessing

The preprocessing stage refines abstract text into an analyzable format. We first organized documents by publication year to construct a longitudinal corpus and then removed function words with low semantic contribution, such as articles, prepositions, and conjunctions, as well as generic terms that reduce discriminative power.

**Table 1.** Web of Science search formula.

| Category | Formula |
|---|---|
| Devices, OS, industry, ecosystem | TS_BLOCK1 = ((smartphone OR 'smart phone' OR 'mobile phone' OR cellphone OR 'cell phone' OR 'cellular phone' OR handset OR 'mobile device' OR 'feature phone' OR tablet OR 'mobile terminal' OR iPhone OR Android OR iOS OR Symbian OR 'Windows Phone' OR BlackBerry) AND ('mobile industr' OR 'smartphone industr' OR 'handset industr' OR ecosystem OR 'value chain' OR 'supply chain' OR manufacturing OR production OR platform OR 'app store' OR 'business model' OR market OR vendor OR OEM OR brand OR 'market share' OR competition OR strateg OR pricing OR 'intellectual property' OR patent OR standard OR '3GPP' OR 'LTE' OR '5G')) |
| Telecom operators, performance metrics, policy | TS_BLOCK2 = (((mobile OR cellular OR wireless) NEAR/3 (operator OR carrier OR telecom OR 'service provider' OR MNO OR MVNO)) AND (ARPU OR churn OR tariff OR spectrum OR licensing OR regulation OR 'network sharing' OR roaming OR 'base station' OR RAN OR 'core network' OR 'VoLTE' OR '5G NR' OR 'non-standalone' OR 'standalone' OR vendor*)) |
| Components, semiconductors, display, supply chain, production | TS_BLOCK3 = (((smartphone OR handset OR 'mobile device') NEAR/3 (chip OR SoC OR baseband OR modem OR 'RF front end' OR display OR OLED OR AMOLED OR LTPO OR 'camera module' OR 'image sensor' OR CIS OR battery OR charger OR PMIC OR memory OR DRAM OR NAND OR 'touch panel' OR 'cover glass' OR 'Gorilla Glass' OR PCB OR packaging OR SiP OR foundry OR fab OR TSMC OR 'Samsung Foundry' OR Qualcomm OR MediaTek OR Sony)) AND (market OR vendor OR 'supply chain' OR manufacturing OR production OR capacity OR 'lead time' OR shortage)) |
| Mobile apps, services, ecosystem | TS_BLOCK4 = ((mobile NEAR/3 (app OR 'app store' OR 'mobile service' OR 'mobile payment' OR 'm-payment' OR 'mobile banking' OR fintech OR 'ride-hailing' OR 'social media' OR 'messaging app')) AND (market OR monetization OR platform OR ecosystem OR competition OR pricing OR adoption OR diffusion)) |
| Standards, SEP, patents, licensing, royalties | TS_BLOCK5 = (('3GPP' OR 'Release 15' OR 'Release 16' OR 'Release 17' OR 'LTE' OR 'LTE-Advanced' OR '5G' OR '5G NR' OR 'New Radio' OR '6G' OR ETSI OR 'IMT-2020' OR 'IMT-2030' OR 'standard-essential patent' OR SEP OR FRAND) AND (industry OR market OR standard OR patent OR litigation OR licensing OR royalty)) |

To mitigate concept fragmentation caused by spelling and wording variations, we applied standardization for key terms and handled synonyms by unifying equivalent expressions, such as platform and ecosystem and standard and standardization. These steps minimize noise in subsequent embedding and clustering procedures and improve the interpretability of the resulting topics.

### *3.2. Regime Identification & Segmentation*

#### 3.2.1. Document Embedding

The purpose of this section is to convert longitudinally accumulated paper abstracts into embedding based numerical vectors so that semantic distributions can be compared across years. Because raw text is not directly suitable for inter year comparison or for measuring distributional shifts, each document is mapped to a vector representation that preserves semantics, enabling similarity and distance based analyses in a shared coordinate space.

The time ordered abstract corpus is encoded using the SPECTER2 model. SPECTER2 is a pretrained model that learns scholarly document representations using citation contexts and is well suited to capturing semantic proximity in science and technology texts.[14,15] Each abstract is transformed into a fixed length vector of 768 dimensions, such that semantically similar documents are located closer to one another in the embedding space.

To reduce the influence of vector magnitude and to improve the stability of distributional comparisons across years, L2 normalization is applied. This enables consistent estimation of distribution based statistics required for transition analysis, such as shifts in the centroid and changes in dispersion. All analyses are implemented in Python using Google Colab.

#### 3.2.2. Regime Shift Detection

The purpose of this section is to quantitatively identify regime shifts in which the mobile industry knowledge structure changes beyond gradual variation and exhibits a distinct distributional character. In this study, a regime shift is defined as a point at which continuity between consecutive year level embedding distributions weakens substantially and the knowledge system is reorganized into a different structure. Specifically, for a given year t, a regime shift is assumed when the embedding distribution before t and the embedding distribution after t differ in a statistically meaningful way.

Regime shift detection is conducted in two stages. First, to generate candidate boundaries, a multiple change point detection procedure is applied to the embedding time series. Document embeddings are arranged by year and an E Divisive based nonparametric method is used to detect multiple boundary indices where distributions change. This stage produces a candidate set of years that may mark distributional reconfiguration and serves as an exploratory step for prioritizing strong transitions.

Second, to confirm key boundaries among the candidates, a time series of distributional distances based on Maximum Mean Discrepancy is constructed. For each adjacent year pair t and t plus 1, the distributional difference is computed as MMD between the embedding sets of the two years, forming a year pair change series. Boundaries are then prioritized around peak segments where the change magnitude increases sharply. Years for which E Divisive candidates and MMD peaks are jointly observed are treated as high likelihood transition points. To avoid excessive sensitivity to a single year, regime shifts are finalized by cross validating the stage one candidates against the stage two peak segments.

Finally, two auxiliary indicators are used to assess the structural plausibility of the detected shifts. Centroid shift is computed to measure how far the mean location of the distribution moves across the boundary and to evaluate whether an MMD increase is driven by a translation of the overall topical center. Dispersion change is also computed to examine whether the shift is associated with expansion or contraction of distributional spread, supporting interpretation in terms of knowledge diversification or convergence. This design goes beyond statistical detection by enabling characterization of the structural nature of regime shifts. All analyses are implemented in Python using Google Colab.

#### 3.2.3. Regime Segmentation

The purpose of this section is to segment the full observation period into multiple regimes using the finalized transition years as boundaries, thereby establishing comparable period specific structures for subsequent analysis. Using the detected boundary years, the 2005 to 2024 corpus is partitioned into contiguous intervals and each interval is treated as a relatively homogeneous knowledge system. Each regime then becomes the unit of analysis for constructing topic structures and analyzing topic dynamics. After segmentation, regime specific document sets, embedding distributions, and topic clusters can be constructed independently, and structural differences and inheritance relationships before and after transitions can be systematically compared in later steps. All analyses are implemented in Python using Google Colab.

### *3.3. Topic Structure Construction*

#### 3.3.1. Topic Clustering per Regime

The purpose of this section is to derive topics and construct regime specific topic structures by identifying density patterns in document embeddings within each segmented regime. Regime level topic clustering reveals how studies produced within a given period are organized into subtopic sets and provides the basic units for cross regime topic alignment and topic dynamics analysis.

Document embeddings obtained in Section 3.2 are split using the regime boundaries to form an embedding set for each regime. HDBSCAN, a density based clustering method, is then applied to cluster each regime embedding distribution into topics. Because HDBSCAN forms clusters based on density variation, the number of topics is not fixed in advance and can be determined flexibly by the data structure within each regime. HDBSCAN also separates noise points, allowing documents that do not stably belong to any topic to be treated as outliers, which improves topic homogeneity.

To enhance interpretability of the derived topic structures, regime specific topics are projected into two dimensional space using UMAP for visualization. Because UMAP aims to preserve neighborhood relations from the high dimensional embedding space, it is used to inspect relative distances among topics, the degree of topic separation, and regime level structural differences in an intuitive way.

Finally, representative keywords are computed for each topic by aggregating the abstract texts of documents assigned to the topic. Class based TF IDF is applied to extract highly weighted terms that characterize each topic, providing a basis for topic labeling and semantic interpretation.

#### 3.3.2. SPECTER based Topic Representation

The purpose of this section is to represent regime specific topics derived by HDBSCAN in a consistent manner in the SPECTER embedding space, so that topic meanings can be summarized quantitatively and used in downstream steps, including cross regime topic alignment and topic dynamics analysis. Topic level representations are constructed by summarizing multiple document embeddings into a single representative vector while also producing interpretable descriptors such as representative documents and keywords.

For each topic, a representative vector is defined to capture the central tendency of document embeddings within the topic. Specifically, the representative vector of topic k is computed as the centroid, the mean of the document embeddings assigned to the topic. This represents the topic as a single point in the SPECTER semantic space and serves as a key input for topic similarity computation, cross regime matching, and lineage reconstruction. Because centroid based representations reflect shared meaning within a topic, they provide comparability even when the number and distribution of topics differ across regimes.

Because a representative vector alone is not directly interpretable, representative documents and representative keywords are additionally identified for each topic. The representative document is defined as the document whose embedding is closest to the topic centroid, serving as an exemplar that best captures the topic content. Representative keywords are obtained by aggregating abstracts within each topic and extracting the top terms using class based TF IDF weighting. Each topic is thus described by a three component representation consisting of the centroid vector, a representative document, and a representative keyword set, enabling both quantitative comparison and qualitative interpretation.

To present topic structures intuitively, a topic map is constructed by projecting topic centroid vectors into two dimensional space using UMAP. This visualization illustrates relative distances and cluster structures among topics and supports inspection of regime specific topic configurations and topic positions such as central versus peripheral locations. However, UMAP is used only for visualization, while quantitative comparison and alignment are performed in the original high dimensional embedding space based on similarities among topic centroids.

In summary, the SPECTER based topic representation provides a standardized topic level expression by defining topics quantitatively through centroid vectors, ensuring interpretability through representative documents and class based TF IDF keywords, and visualizing structures through a

 doi:10.20944/preprints202603.0320.v2



UMAP based topic map. This representation supports subsequent cross regime topic alignment and growth typology analysis.

### *3.4. Topic Dynamics and Growth Typology*

#### 3.4.1. Cross Regime Topic Alignment

The purpose of this section is to temporally connect topics that are independently derived within each regime, reconstruct topic inheritance relationships, and systematically identify evolutionary events such as *birth*, *death*, *merge*, *split*, and *recombination*. Because regime specific topics are constructed from period specific data distributions, they do not share a common topic index system. An explicit alignment procedure is therefore required to generate linkage edges by matching topic representative vectors across adjacent regimes.

Semantic continuity between topic $k$ in regime $T$ and topic $l$ in regime $T$ plus 1 is assessed using similarity between topic representative vectors. Each topic is summarized by two attributes, a representative vector defined as the centroid of document embeddings within the topic and a topic size defined as the number of documents assigned to the topic. For all topic pairs $k$ and $l$ across adjacent regimes, a cosine similarity matrix is computed to form candidate links.

To prevent excessive link creation, three criteria are applied sequentially to confirm one to one inheritance relationships. First, a similarity threshold $\tau$ is introduced so that only pairs satisfying $sim(k,l) \geq \tau$ are retained as candidates. To determine $\tau$ in a data driven manner, a permutation based null distribution is constructed and $\tau$ is set automatically to exceed similarity levels expected under random matching, for example 0.7. Second, a margin criterion is used to ensure that the best match is sufficiently dominant. Specifically, for each topic $k$, the margin is computed as the difference between the highest and the second highest similarity scores, $margin(k) = best(k) - second(k)$, and one to one inheritance is confirmed only when $margin(k) \geq \delta$. The value of $\delta$ is set empirically based on the margin distribution, for example 0.03. Third, to reduce misalignment caused by one sided matching, a mutual top $N$ condition is applied. A link is accepted only when the best candidate $l$ for topic $k$ also includes $k$ among its top $N$ candidates. As a result, a one to one inheritance link is defined as a pair that jointly satisfies $sim \geq \tau$, $margin \geq \delta$, and the mutual top $N$ condition.

After confirming one to one inheritance links, the remaining connections are further interpreted to classify evolutionary events. Topic *birth* is defined for a topic $l$ in regime $T$ plus 1 when no valid incoming link from the previous regime exceeds the threshold, which corresponds to $In(l) = 0$. Conversely, topic *death* is defined for a topic $k$ in regime $T$ when no valid outgoing link to the next regime exceeds the threshold, which corresponds to $Out(k) = 0$. However, classifying many to one or one to many patterns as *merge* or *split* solely because the number of links is at least two risks over identification. To address this, the study introduces weights that reflect not only similarity but also topic size and confirms events based on the extent to which a source topic explains a target topic.

As shown in Equation 1, the weighted contribution from topic $k$ in regime $T$ to topic $l$ in regime $T$ plus 1 is defined as $w(k,l) = S(k,l) \times n(k)$, where $S$ denotes cosine similarity and $n(k)$ denotes topic size. The incoming share of topic $k$ to topic $l$ is then computed, as shown in Equation 2, as $share\_in(k \rightarrow l) = w(k,l) / \Sigma\_\{k' \in In(l)\}\, w(k',l)$. Because this share is a relative contribution based on similarity multiplied by size rather than similarity alone, it enables a more conservative determination of whether a target topic is genuinely formed through *convergence*.

$$w(k,l) = S(k,l) \times n(k) \tag{1}$$

$$share\_in(k \rightarrow l) = \frac{w(k,l)}{\Sigma\_\{k' \in In(l)\}\, w(k',l)} \tag{2}$$

A *merge* is defined as a case in which a topic $l$ in regime $T$ plus 1 is formed through substantive contributions from multiple topics in regime $T$. To confirm a *merge*, two conditions are imposed. First, at least the top two *share_in* values must each be greater than or equal to $\beta$. Second, the cumulative

 doi:10.20944/preprints202603.0320.v2



sum of the top contributions must be greater than or equal to $\gamma$. For example, $\beta$ can be set to 0.20 and $\gamma$ to 0.70. In other words, a topic is classified as a *merge* only when at least two prior topics each contribute at a meaningful level and jointly account for most of the target topic.

A *split* is defined as the inverse of a *merge*. Specifically, a *split* occurs when a topic $k$ in regime $T$ branches into multiple topics in regime $T$ plus 1 through substantive outgoing contributions. To quantify branching, the outgoing share is computed, as shown in Equation 3, as *share_out(k→l) = w(k,l) / Σ_{l′ ∈ Out(k)} w(k,l′).* The confirmation criteria for *split* are set symmetrically to those for *merge*. A topic $k$ is classified as a *split* when at least the top two *share_out* values are each greater than or equal to $\beta$ and the cumulative sum of the top contributions is greater than or equal to $\gamma$.

$$share\_out(k \rightarrow l) \; = \frac{w(k,l)}{\Sigma\_\{l' \in Out(k)\}\, w(k,l')} \tag{3}$$

In summary, this section first confirms similarity based one to one inheritance links in a conservative manner. It then introduces share measures that incorporate both similarity and topic size to interpret the remaining multiple links while avoiding over identification of *merge* and *split* events. Finally, it classifies topic evolution using consistent rules that also cover *birth* and *death*. This design enables the growth typology analysis in the subsequent section to be conducted not at the level of isolated topics but at the level of temporally connected topic lineages.

### 3.4.2. Topic Growth Typology

This section presents a method for classifying growth typologies by combining the structural roles and temporal growth patterns of topic lineages constructed from cross regime topic alignment. Even when lineages exhibit similar growth trajectories, their roles within knowledge flows may differ. Conversely, lineages that occupy central positions may still display distinct growth dynamics. Accordingly, this study adopts a two by two typology that integrates network based structural indicators with time series based growth indicators rather than reducing topic growth to a single metric.

The unit of analysis is not an isolated topic within a single regime, but a topic lineage defined as a chain of topics connected across adjacent regimes through the alignment procedure described in Section 3.4.1. Each lineage may undergo continuation after birth, experience transformations such as merge, split, or recombination, or terminate through death. Growth typology classification is conducted by comparing where each lineage is positioned in the knowledge flow and how its scale changes over time. Lineages that disappear during the knowledge flow or those that are newly born are excluded from the growth typology analysis.

The typology is defined along two axes. Structural position refers to the centrality of a lineage within the topic transition graph. A lineage level structural score, struct_score, is computed based on centrality measures such as PageRank, degree, and k core. To remove scale differences and enable relative comparison across lineages, the structural score is converted into a percentile rank and normalized to the 0 to 1 range, which is used as the X axis value. Higher values indicate a more central, core position with stronger linkage and brokerage roles, whereas lower values indicate a more peripheral position and a higher likelihood of being locally bounded.

Growth pattern captures the temporal expansion dynamics of a lineage. A growth indicator is computed from the lineage level size time series, such as regime level document counts or topic size changes. In particular, a spike based measure is used to distinguish bursty growth from gradual accumulation by capturing whether rapid increases occur at specific points in time. The growth indicator is also converted into a percentile rank and normalized to the 0 to 1 range, which is used as the Y axis value. Higher values indicate stronger bursty growth characterized by short term surges, whereas lower values indicate persistent accumulation or stable maintenance without abrupt fluctuations.

Because both the X and Y axes are rank based measures in the 0 to 1 range, quadrant boundaries are defined using the median threshold of 0.5. Lineages are thus classified into four types: peripheral

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 March 2026 doi:10.20944/preprints202603.0320.v2



persistent for X below 0.5 and Y below 0.5, peripheral bursty for X below 0.5 and Y at least 0.5, core persistent for X at least 0.5 and Y below 0.5, and core bursty for X at least 0.5 and Y at least 0.5. Peripheral persistent lineages represent subtopics that accumulate stably within a limited scope. Peripheral bursty lineages are structurally peripheral but exhibit sharp attention spikes at specific times. Core persistent lineages correspond to foundational themes that accumulate over long periods in the center of the knowledge flow. Core bursty lineages represent themes that grow rapidly in the core and emerge as central domains after transition periods.

Finally, the classification results are visualized using an X Y scatter plot of structural position rank versus growth pattern rank to show where each lineage is located among the four types and to compare distributions across types. By jointly considering temporal change and network roles, this framework provides a systematic basis for explaining the formation, diffusion, and reconfiguration mechanisms of core themes in mobile industry knowledge flows.

## 4. Results

### *4.1. Regime Identification and Segmentation Results*

#### 4.1.1. Regime Identification Results

Table 2 reports the regime intervals identified through E Divisive based change point detection. The analysis segments the observation period into three regimes, Regime 1 spanning 2005 to 2012, Regime 2 spanning 2013 to 2019, and Regime 3 spanning 2020 to 2024. These results indicate that the embedding time series contains multiple boundaries at which the distribution changes, suggesting that the knowledge structure is likely reorganized into distinct configurations, particularly around the 2012 to 2013 and 2019 to 2020 transitions.

**Table 2.** E Divisive analysis results.

| Segment | Start | End |
|---|---|---|
| 1 | 2005 | 2012 |
| 2 | 2013 | 2019 |
| 3 | 2020 | 2024 |

Figure 2 presents a time series that summarizes embedding distribution differences between adjacent years using Maximum Mean Discrepancy. For each year boundary, the MMD value provides a single measure of how much the overall distribution changes from year t to year t plus 1, with larger values indicating greater structural change in the knowledge system. The largest peak is observed for the 2019 to 2020 transition, followed by a comparatively large increase for 2012 to 2013. Among the candidate boundaries suggested by E Divisive, these results indicate that 2019 to 2020 constitutes the strongest regime shift, while 2012 to 2013 represents the next most salient transition.

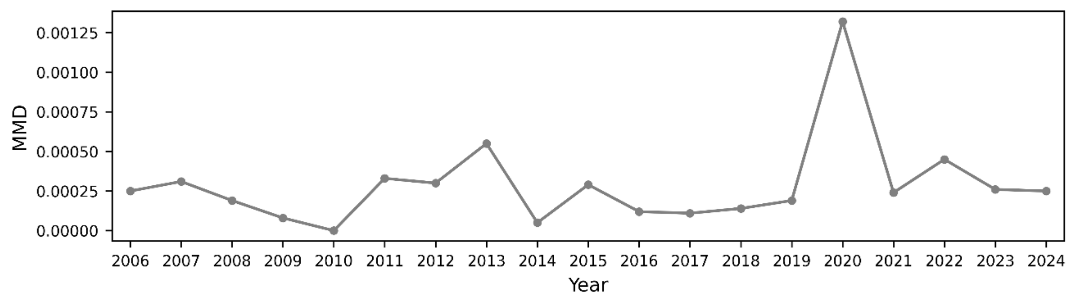


**Figure 2.** MMD analysis results.

 doi:10.20944/preprints202603.0320.v2



Figure 3 shows year level publication volumes for mobile industry related articles from 2005 to 2024 together with the regime intervals determined by the confirmed transition years of 2012 to 2013 and 2019 to 2020. The regimes displayed in the figure, Regime 1 spanning 2005 to 2012, Regime 2 spanning 2013 to 2019, and Regime 3 spanning 2020 to 2024, are defined by boundaries at which discontinuous changes in the document embedding distribution are detected. The publication trend is provided as supplementary evidence to illustrate how the segmentation corresponds to changes in the scale of research production.

In Regime 1 spanning 2005 to 2012, annual publication volume increases gradually from 980 to 2,137, indicating steady accumulation of research output. In Regime 2 spanning 2013 to 2019, publication volume expands from 2,564 to 5,796 and the growth slope becomes notably steeper, suggesting a transition to an expansion phase in research production. In Regime 3 spanning 2020 to 2024, publication volume jumps sharply to 7,921 in 2020 and remains high thereafter, reaching 9,170 in 2021, 9,692 in 2022, 9,804 in 2023, and 10,004 in 2024. Notably, the 2019 to 2020 boundary corresponds to the largest MMD peak and is also associated with a stepwise upward shift in publication volume around the same period.

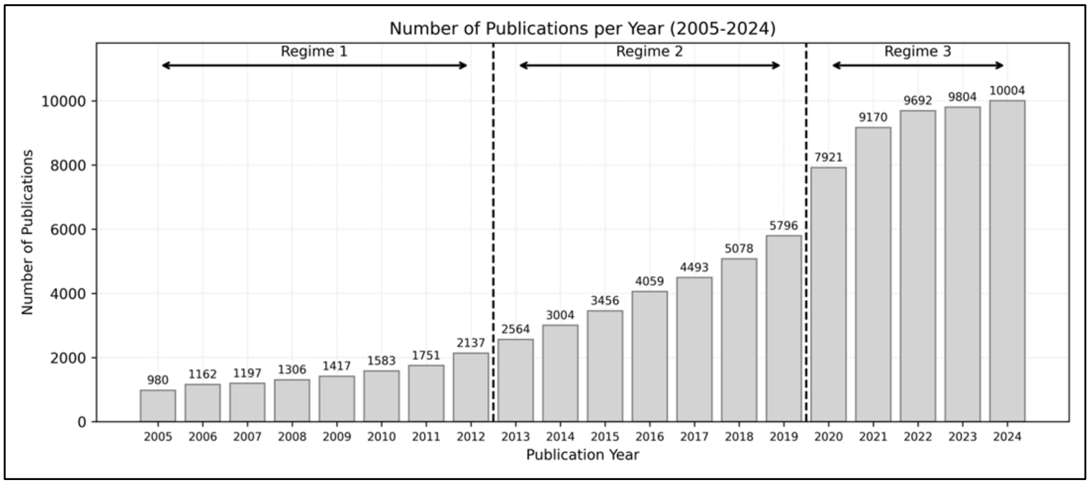


**Figure 3.** Yearly publication trends and confirmed regime segmentation.

#### 4.1.2. Regime Validation Results

To further support the validity of the regime segmentation, additional indicators are used to examine whether the observed increase in distributional distance reflects structural reconfiguration rather than sampling fluctuation. Figure 4 reports centroid shift, which measures the year to year displacement of the mean location of the embedding distribution, adjusted by distributional spread for comparability. A large centroid shift indicates that the topical center of documents moves as a whole in a specific direction at the boundary, suggesting that translation of the distributional center is a key driver of the transition. The results show a particularly large centroid shift for the 2019 to 2020 boundary, with a comparatively large movement also observed for 2011 to 2012. This supports the interpretation that the 2019 to 2020 transition is not merely a diversification effect but a structural change characterized by a clear relocation of the knowledge distribution center.

 doi:10.20944/preprints202603.0320.v2



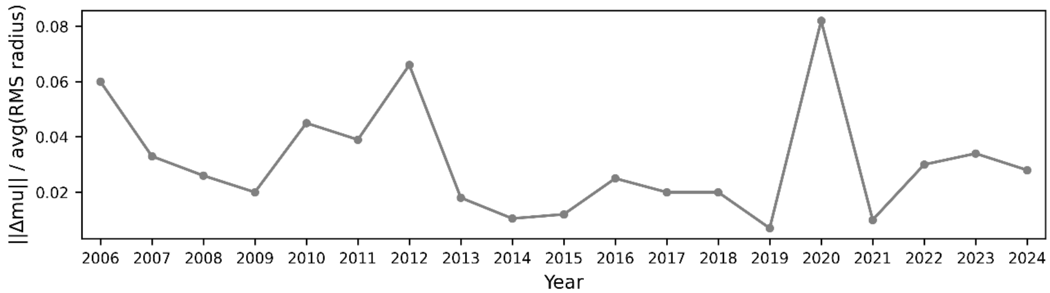


**Figure 4.** Centroid shift results.

Figure 5 reports dispersion change and shows whether the spread of the embedding distribution, measured by RMS radius, increases or decreases relative to the previous year. Positive values indicate expansion or diversification, whereas negative values indicate contraction or convergence. For the 2019 to 2020 boundary, dispersion change is negative, indicating a post transition convergence pattern in which the distribution becomes more concentrated in a particular direction. Taken together, the 2019 to 2020 transition is a boundary at which a sharp increase in distributional difference captured by the MMD peak, a substantial centroid translation indicated by centroid shift, and a decrease in dispersion indicating convergence are observed simultaneously. These results confirm that the knowledge structure undergoes a strong reconfiguration at this transition.

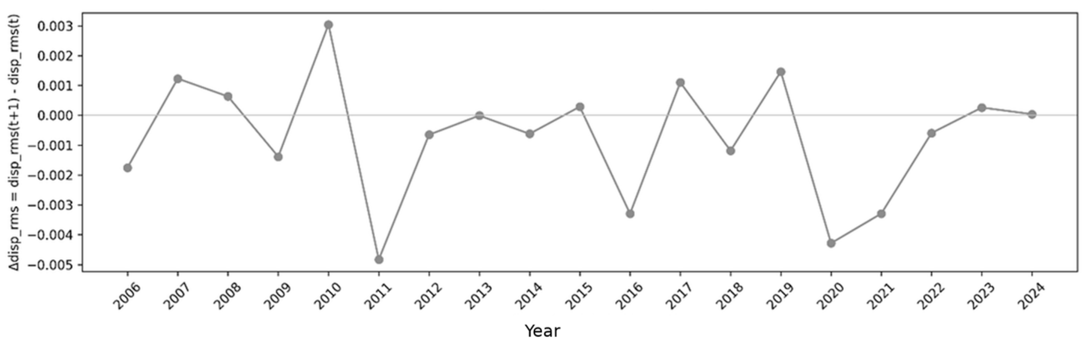


**Figure 5.** Dispersion change results.

*4.2. Results on Topic Structure and Dynamics*

4.2.1. Results of Topic Structure Construction

Figure 6 presents an intertopic distance map that visualizes the embedding based topic clustering results for Regime 1 spanning 2005 to 2012 in two dimensional space. Each circle represents a topic, where circle size indicates the number of documents assigned to the topic and distances between circles indicate semantic distance, the inverse of similarity. In Figure 6, Topic 1 accounts for the largest share and is positioned relatively far from other topics, suggesting that the Regime 1 knowledge structure is organized around a dominant core axis. By contrast, Topic 2, Topic 3, and Topic 7 are located relatively close to one another, implying that they may constitute a subcluster that shares similar technical, measurement, and validation contexts.

Figure 7 reports topic word scores for each topic identified in Regime 1, providing a basis for semantic interpretation. The topic composition indicates that mobile industry knowledge in Regime 1 is organized primarily around foundational technologies, including hardware, networks, sensors, power, and measurement and validation. Topic 1 can be summarized as mobile wireless networks and power efficiency, highlighting network and power optimization issues in wireless communication environments as the central axis. Topic 2 captures smartphone optical LiDAR sensing,

 doi:10.20944/preprints202603.0320.v2



indicating that optical distance measurement and sensor applications constitute a key subtheme. Topic 3 reflects calibration, measurement, and validation, showing that methodological work on measurement, prediction, and verification forms a distinct topic. Topic 4 corresponds to battery power management and internal reliability, indicating an independent stream focused on power control and reliability issues. Topic 5 reflects mobile health with an emphasis on users and behavior, representing one application domain within Regime 1. Topic 6 captures control and measurement algorithms and robotics, highlighting an algorithmic and control oriented research stream. Topic 7 can be summarized as battery related electronic materials and nanomaterials, indicating that materials and nano scale research related to batteries differentiates into a separate topic.

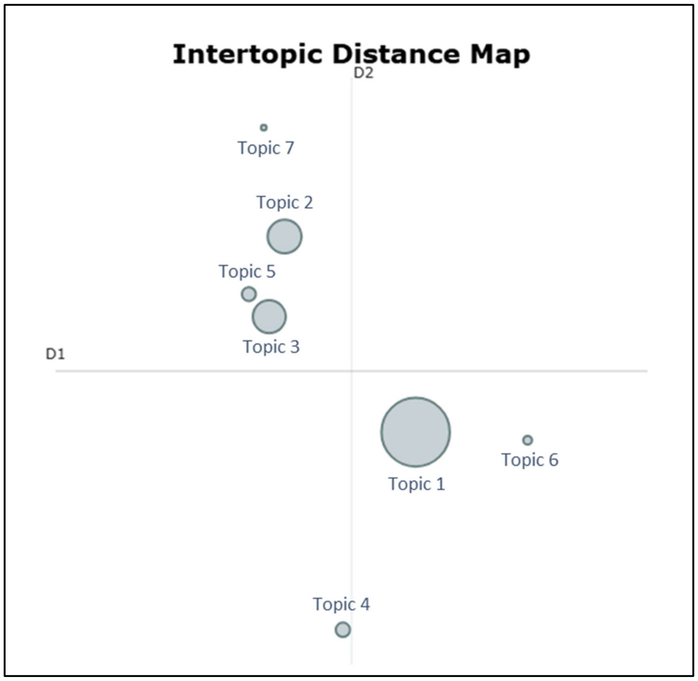


**Figure 6.** Regime 1 intertopic distance map.

Figure 8 presents the intertopic distance map for topic clusters derived in Regime 2 spanning 2013 to 2019. As in Figure 6, each circle denotes a topic, circle size represents topic volume measured by the number of documents, and intertopic distances represent semantic separation in the embedding space. Relative to Regime 1, Regime 2 exhibits a larger number of topics and shows multiple mid sized topics concentrated near the center. This suggests that the knowledge structure is reorganized into a more polycentric configuration as research expands beyond a single technological axis toward networks, sensors, energy, platforms, and application domains. In addition, some topics located near the center, such as Topic 1, Topic 2, and Topic 5, appear closely positioned and form a cluster of interrelated research streams, whereas peripheral topics constitute relatively independent subdomains such as specific applications or regulation and operations related issues.

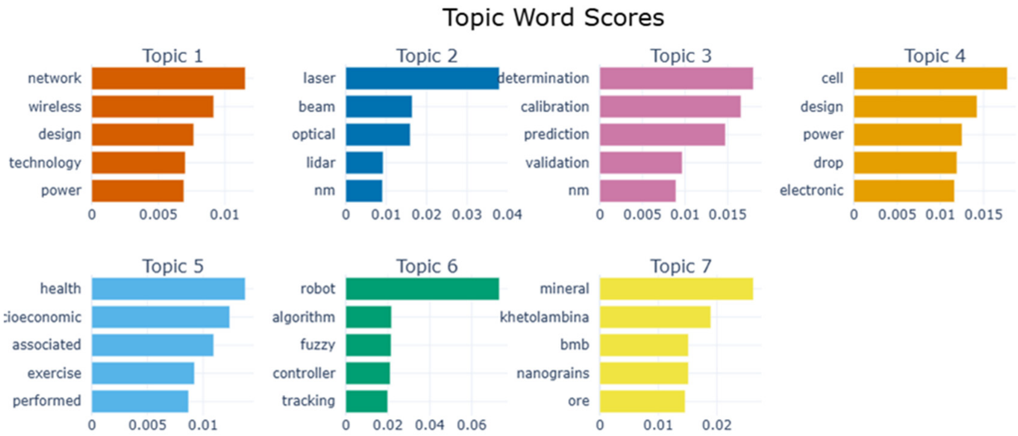


**Figure 7.** Regime 1 topic word scores.

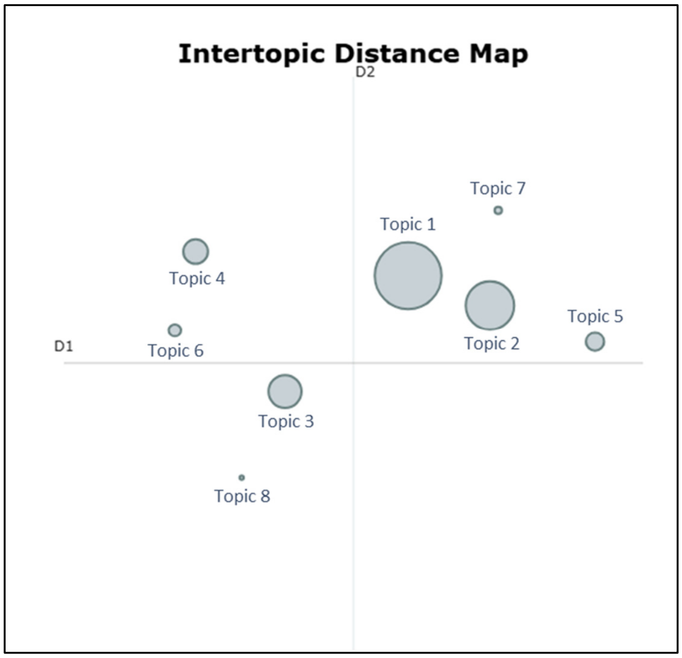


**Figure 8.** Regime 2 intertopic distance map.

Figure 9 reports topic word scores for Regime 2 and provides interpretive evidence for topic meanings. The topic composition indicates an expansion phase in which foundational technology research continues while applications and socio technical themes such as user perception, platform governance, and bio integration begin to combine more explicitly. Topic 1 can be summarized as mobile network algorithms and power optimization, reflecting sustained performance optimization research centered on network, antenna, and power related keywords. Topic 2 represents mobile app acceptance and user perception, indicating that user behavior and technology acceptance studies form an independent topic through terms such as learning, social factors, and perceived constructs. Topic 3 captures solar based energy harvesting and charging combined with device surface processes, reflecting a research stream linking energy autonomous devices with materials and process issues. Topic 4 reflects optical, laser, and LiDAR or spectroscopy sensing modules, indicating continued development of sensor based measurement and module technologies. Topic 5 captures optical and sensor data prediction and modeling as well as calibration and validation, suggesting that the Regime 1 measurement and validation stream expands toward data and modeling centered work. Topic 6

 doi:10.20944/preprints202603.0320.v2



represents smartphone RF and electromagnetic field exposure, showing that exposure and impact concerns are established as a distinct topic while remaining connected to RF and optical related terms. Topic 7 reflects iOS platform governance and policy and organizational operations, indicating the emergence of governance oriented research combining platform, social, and policy keywords. Topic 8 captures genomic bio mobile convergence, suggesting that the integration of bio data and analytics with mobile contexts becomes a new application domain in Regime 2.

Figure 10 presents the intertopic distance map for Regime 3 spanning 2020 to 2024. Each circle denotes a topic, circle size indicates topic volume, and distances between circles represent semantic distance in the embedding space. Regime 3 exhibits a structure in which relatively large topics occupy the center while many medium and small topics are dispersed around the periphery. This suggests that even during a period of substantially expanded research production, a dominant core axis remains while topics diversify across applications, policy, bio, and robotics, yielding a more complex topic structure. In particular, Topic 1, Topic 2, and Topic 5 are located near the center and form adjacent streams related to data, smartphones, and optical or sensor based research.

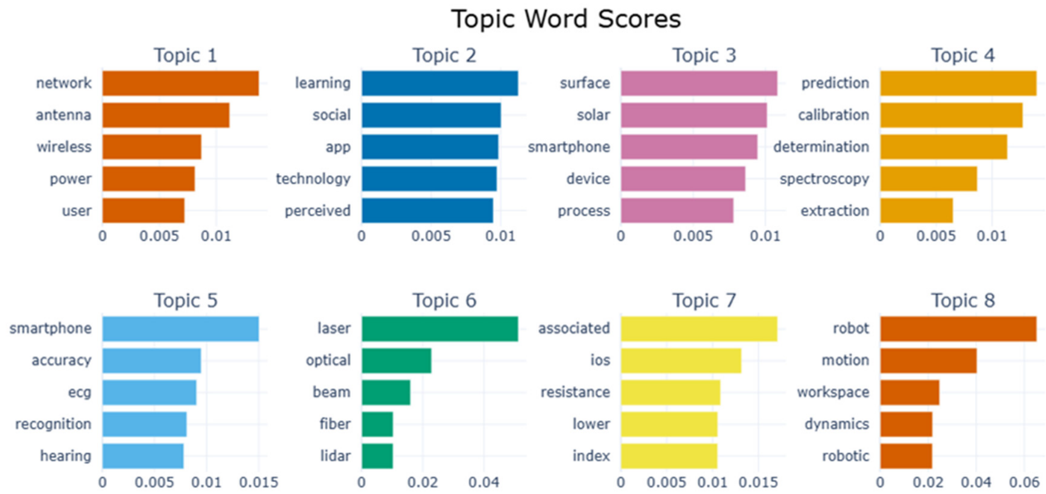


**Figure 9.** Regime 2 topic word scores.

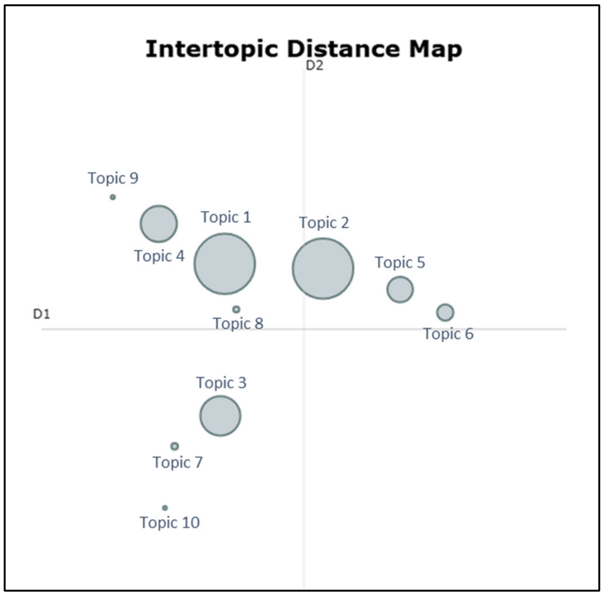


**Figure 10.** Regime 3 intertopic distance map.

By contrast, topics such as Topic 8 through Topic 10 are positioned more peripherally and appear to constitute relatively independent expansion domains, including institutional and governance issues and bio and therapeutic themes.

Figure 11 reports topic word scores for Regime 3 and indicates that topic composition evolves along three parallel axes, a data and algorithm centered communications and platform axis, a sensor and energy and robotics axis, and a policy and governance and bio convergence axis. Topic 1 can be summarized as 5G network data analytics and algorithmic optimization and applications, characterized by co occurrence of terms such as data, network, 5G, and algorithm. Topic 2 represents smartphone data driven services including health and social applications, centered on terms such as app, health, smartphone, and social. Topic 3 captures smartphone sensor and system based measurement and analysis methods, emphasizing terms such as method, analysis, and sensor. Topic 4 reflects power efficiency and energy transitions including fuel cells and solar and recycling, capturing strengthened sustainability oriented themes through terms such as recycling, solar, power, and charging. Topic 5 represents optical and laser modules including fiber beam systems and power harvesting components, centered on terms such as laser, optical, beam, and power.

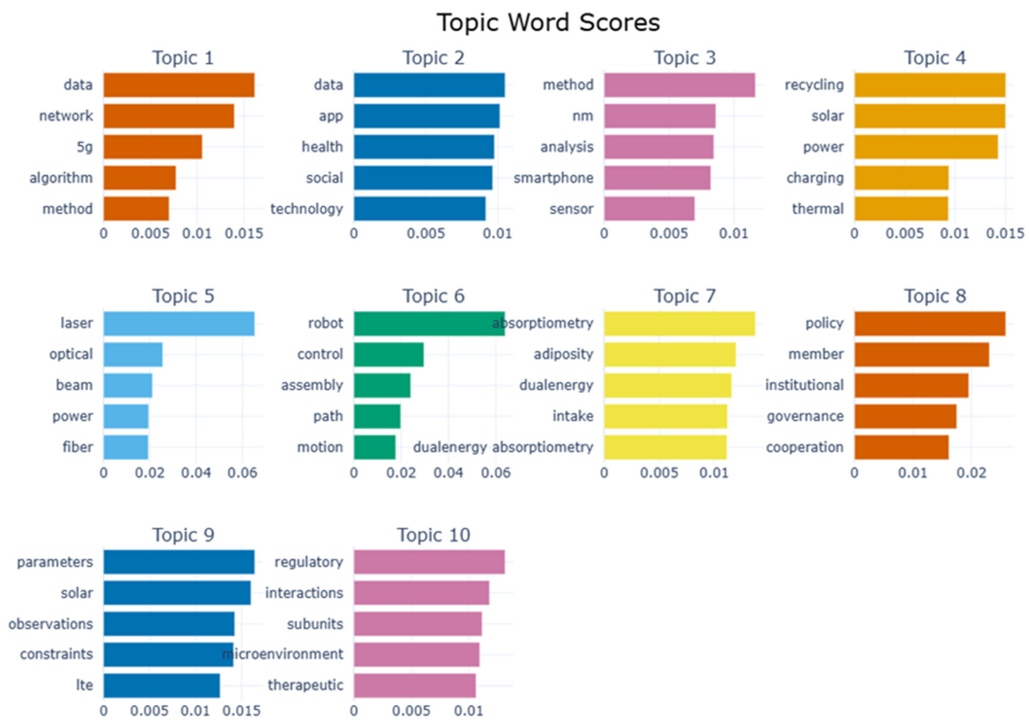


**Figure 11.** Regime 3 topic word scores

Distinct expansion axes in Regime 3 are also evident. Topic 6 reflects manufacturing automation and robotics in control, assembly, and motion, characterized by terms such as robot, control, assembly, and motion. Topic 7 captures bio and therapeutic characteristics including DEXA and nutrient absorption analysis, combining biomedical analytic terms such as absorptiometry, adiposity, and intake. Topic 8 represents industrial policy and service based cooperation including regulation, centered on terms such as policy, institutional, governance, and cooperation, highlighting the emergence of institutional and coordination issues as an independent topic distinct from technical axes. Topic 9 reflects LTE and traffic measurement and parameter constraints linked to social and performance themes, including terms such as parameters, constraints, and observations. Topic 10 captures bio and therapeutic regulation and engagement interactions, emphasizing terms such as regulatory, interactions, and participants, indicating strengthened coupling between expanding bio applications and institutional participation and regulation.

 doi:10.20944/preprints202603.0320.v2



4.2.2. Results of Topic Dynamics

Tables 3 and 4 report topic transition types between adjacent regimes based on the cross regime topic alignment results. Transitions are classified into continuation, birth, death, merge, and split, indicating whether a topic identified in one regime is inherited by a topic in the subsequent regime, newly emerges, disappears, is integrated from multiple topics, or branches into multiple topics. For continuation cases, the cosine similarity value sim is also reported as an indicator of alignment strength, enabling assessment of inheritance relationships with high semantic continuity across regimes.

For the transition from Regime 1 to Regime 2, several topics show direct inheritance with high similarity, indicating strong semantic continuity across the boundary. At the same time, merge and split events are observed in parallel, suggesting that post transition topic structures are not merely preserved but reorganized through both integration and branching. In addition, birth events in Regime 2 indicate the emergence of new topics, while some topics from Regime 1 are not linked to the subsequent regime and are therefore classified as death.

**Table 3.** Topic dynamics results for Regime 1 to Regime 2.

| Type | Regime 1 | Regime 2 | Sim | Transition |
|---|---|---|---|---|
| continuation | 2 | 4 | 0.996046 | R1->R2 |
| continuation | 5 | 8 | 0.980957 | R1->R2 |
| merge | 1,6 | 1 | | R1->R2 |
| merge | 1,4,5 | 2 | | R1->R2 |
| merge | 1,3,4,6 | 5 | | R1->R2 |
| merge | 1,3 | 6 | | R1->R2 |
| split | 1 | 1,2,5,6 | | R1->R2 |
| split | 3 | 5,6 | | R1->R2 |
| split | 4 | 2,5 | | R1->R2 |
| split | 5 | 2,8 | | R1->R2 |
| split | 6 | 1,5 | | R1->R2 |
| birth | | 3 | | R1->R2 |
| birth | | 7 | | R1->R2 |
| death | 7 | | | R1->R2 |

**Table 4.** Topic dynamics results for Regime 2 to Regime 3.

| Type | Regime 2 | Regime 3 | Sim | Transition |
|---|---|---|---|---|
| continuation | 1 | 1 | 0.99691 | R2->R3 |
| continuation | 3 | 4 | 0.969032 | R2->R3 |
| continuation | 4 | 5 | 0.993875 | R2->R3 |
| continuation | 3 | 9 | 0.988025 | R2->R3 |
| merge | 2,8 | 2 | | R2->R3 |
| merge | 3,5 | 3 | | R2->R3 |
| merge | 2,5,8 | 7 | | R2->R3 |
| merge | 2,7 | 8 | | R2->R3 |
| merge | 2,8 | 10 | | R2->R3 |
| split | 2 | 2,7,8,10 | | R2->R3 |

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 March 2026 doi:10.20944/preprints202603.0320.v2



| | | | |
|---|---|---|---|
| split | 3 | 3,4,9 | R2->R3 |
| split | 5 | 7,3,2 | R2->R3 |
| split | 7 | 2,8 | R2->R3 |
| split | 8 | 2,7,10 | R2->R3 |
| birth | | 6 | R2->R3 |
| death | 6 | | R2->R3 |

A similar pattern appears in the transition from Regime 2 to Regime 3, where many high similarity continuation links confirm that core research axes maintain semantic continuity. Nonetheless, merge and split events recur in this interval as well, implying that topic reconfiguration continues beyond the transition. Birth events in Regime 3 and death of certain topics are also jointly observed, indicating that even during an expansion phase, the emergence of new themes and the disappearance of existing themes proceed in parallel.

Figure 12 visualizes the topic transition results in Table 3 from a lineage perspective and provides an intuitive view of how topics in Regime 1, Regime 2, and Regime 3 are connected and how transition types occur. Topics from each regime are arranged from left to right, and edges indicate transition relationships that satisfy the alignment criteria. Paths characterized by a single link represent *continuation*, patterns in which multiple topics converge into one represent *merge*, and patterns in which one topic branches into multiple topics represent *split*. Topics that newly appear in a given regime without links from the previous regime are labeled as *birth*, whereas topics that do not connect to any topic in the subsequent regime are labeled as *death*. Accordingly, Figure 12 illustrates how topic emergence and disappearance are manifested along the lineage structure and visually supports the interpretation that post transition topic evolution combines persistence through inheritance, reconfiguration through merging and splitting, and parallel processes of emergence and termination.

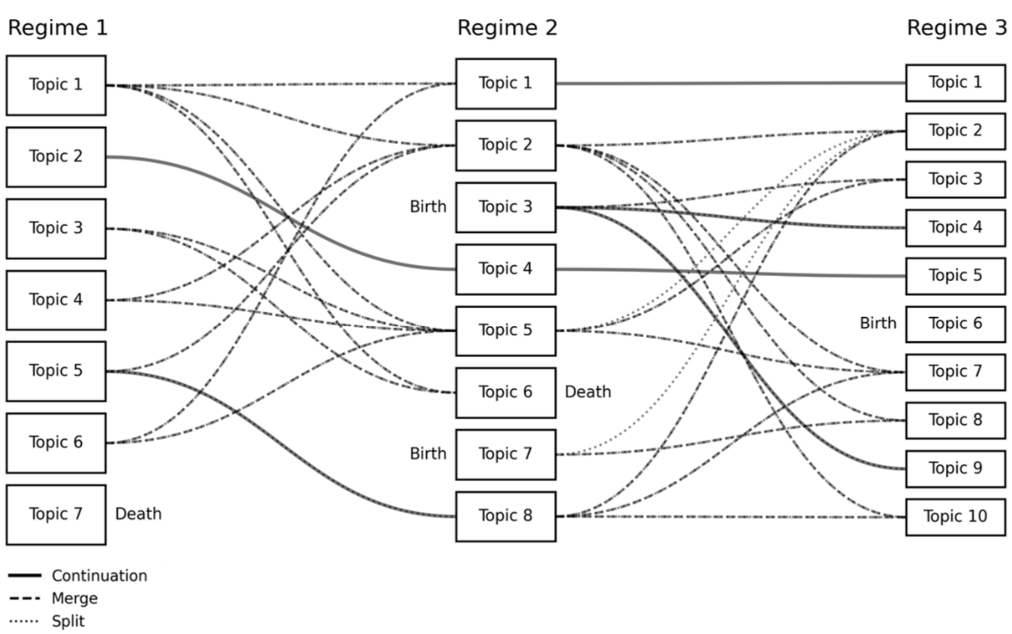


**Figure 12.** Visualization of topic transition lineage paths.

### 4.3. Topic Growth Typology

This section constructs and labels topic lineages for the Topic Growth Typology analysis based on the cross regime transition relationships derived from the topic dynamics results in Section 4.2.2.

Specifically, linkage edges from the Regime 1 to Regime 2 and Regime 2 to Regime 3 alignments are aggregated, and continuous connections that traverse regimes are defined as lineage paths. Each path consists of a sequential linkage from a Regime 1 topic to a Regime 2 topic and then to a Regime 3 topic.

Labeling is restricted to persistently inherited paths. Paths corresponding to *birth*, such as topics that newly appear in Regime 2 without links from the previous regime, and paths corresponding to *death*, which do not connect further to Regime 2 or Regime 3, are excluded because they do not provide a consistent basis for comparing growth typologies. Birth paths enter mid period and therefore lack the initial segment of the growth trajectory, whereas death paths terminate before the end of the observation window and make it difficult to compare continuity of subsequent growth. Accordingly, the typology analysis is conducted only on complete paths that span all three regimes.

More specifically, a lineage path is confirmed only when a Regime 1 topic has a valid link to a Regime 2 topic and the same Regime 2 topic also has a valid link to a Regime 3 topic. Even when *merge* or *split* events occur in the Regime 1 to Regime 2 or Regime 2 to Regime 3 transition, a path is treated as a single lineage as long as a connection from Regime 1 to Regime 3 is ultimately established. Each confirmed path is assigned a unique identifier, L, to enable consistent reference in subsequent analyses. Table 5 summarizes the labeled paths by listing the corresponding topics in each regime, r1_topic, r2_topic, and r3_topic, together with the information required to compute structural position and growth indicators.

As a result, 30 persistently inherited paths are identified after excluding birth and death paths, and Table 5 reports the labels and constituent topics for these 30 lineages. This labeled set serves as the reference basis for computing and comparing structural position and growth pattern at the same unit of analysis, the lineage path, in the subsequent steps.

**Table 5.** Cross-Regime Topic Lineges.

| L | r1_topic | r2_topic | r3_topic | struct_score | spike | size_r1 | size_r2 | size_r3 |
|---|---|---|---|---|---|---|---|---|
| L1 | 1 | 1 | 1 | 0.348124 | 0.882486 | 2180.42 | 4104.609 | 5753.35 |
| L2 | 1 | 2 | 2 | 1.561221 | 0.895231 | 2180.42 | 2073.104 | 3929.011 |
| L3 | 1 | 2 | 7 | 1.352338 | -0.04922 | 2180.42 | 2073.104 | 67.01178 |
| L4 | 1 | 2 | 8 | 1.061543 | -0.04922 | 2180.42 | 2073.104 | 31.56721 |
| L5 | 1 | 2 | 10 | 0.958169 | -0.04922 | 2180.42 | 2073.104 | 18.64825 |
| L6 | 1 | 5 | 2 | 1.625224 | 12.99151 | 2180.42 | 280.814 | 3929.011 |
| L7 | 1 | 5 | 3 | 1.153141 | 3.224502 | 2180.42 | 280.814 | 1186.3 |
| L8 | 1 | 5 | 7 | 1.416341 | -0.76137 | 2180.42 | 280.814 | 67.01178 |
| L9 | 2 | 4 | 5 | -0.18034 | 0.149813 | 491.2514 | 564.8471 | 313.0299 |
| L10 | 3 | 5 | 2 | 1.392785 | 12.99151 | 510.1661 | 280.814 | 3929.011 |
| L11 | 3 | 5 | 3 | 0.920702 | 3.224502 | 510.1661 | 280.814 | 1186.3 |
| L12 | 3 | 5 | 7 | 1.183901 | -0.44956 | 510.1661 | 280.814 | 67.01178 |
| L13 | 4 | 2 | 2 | 1.282293 | 21.0422 | 94.05162 | 2073.104 | 3929.011 |
| L14 | 4 | 2 | 7 | 1.07341 | 21.0422 | 94.05162 | 2073.104 | 67.01178 |
| L15 | 4 | 2 | 8 | 0.782616 | 21.0422 | 94.05162 | 2073.104 | 31.56721 |
| L16 | 4 | 2 | 10 | 0.679241 | 21.0422 | 94.05162 | 2073.104 | 18.64825 |
| L17 | 4 | 5 | 2 | 1.346297 | 12.99151 | 94.05162 | 280.814 | 3929.011 |
| L18 | 4 | 5 | 3 | 0.874214 | 3.224502 | 94.05162 | 280.814 | 1186.3 |
| L19 | 4 | 5 | 7 | 1.137413 | 1.985744 | 94.05162 | 280.814 | 67.01178 |
| L20 | 5 | 2 | 2 | 1.486177 | 19.85333 | 99.41357 | 2073.104 | 3929.011 |

doi:10.20944/preprints202603.0320.v2



| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| L21 | 5 | 2 | 7 | 1.277294 | 19.85333 | 99.41357 | 2073.104 | 67.01178 |
| L22 | 5 | 2 | 8 | 0.9865 | 19.85333 | 99.41357 | 2073.104 | 31.56721 |
| L23 | 5 | 2 | 10 | 0.883125 | 19.85333 | 99.41357 | 2073.104 | 18.64825 |
| L24 | 5 | 8 | 2 | 0.899065 | 229.6191 | 99.41357 | 17.0368 | 3929.011 |
| L25 | 5 | 8 | 7 | 0.690182 | 2.933355 | 99.41357 | 17.0368 | 67.01178 |
| L26 | 5 | 8 | 10 | 0.296013 | 0.094587 | 99.41357 | 17.0368 | 18.64825 |
| L27 | 6 | 1 | 1 | 0.115684 | 153.691 | 26.53424 | 4104.609 | 5753.35 |
| L28 | 6 | 5 | 2 | 1.392785 | 12.99151 | 26.53424 | 280.814 | 3929.011 |
| L29 | 6 | 5 | 3 | 0.920702 | 9.58308 | 26.53424 | 280.814 | 1186.3 |
| L30 | 6 | 5 | 7 | 1.183901 | 9.58308 | 26.53424 | 280.814 | 67.01178 |

Figure 13 classifies topic lineage paths by combining structural position on the X axis with growth pattern on the Y axis, and the four quadrants can be interpreted as follows. The X axis represents the rank based structural score derived from centrality measures in the topic transition network, where higher values indicate a more central core position with stronger connectivity and influence. The Y axis represents the rank based spike indicator, where higher values indicate paths that exhibit larger growth jumps during specific transition intervals.

I. Peripheral bursty, X below 0.5 and Y at least 0.5
Peripheral bursty paths have relatively low structural scores and therefore have not settled into the network core, yet they display pronounced growth jumps at specific points in time. This type often reflects short term trends or issue driven applications, such as the rise of specific technologies or socio technical agendas, and can be interpreted as trajectories that surge rapidly but do not fully stabilize as central pathways. This quadrant includes L13, L14, L15, L16, L21, L22, L23, L24, and L27.

II. Peripheral persistent, X below 0.5 and Y below 0.5
Peripheral persistent paths are structurally peripheral and show no large growth jumps, exhibiting relatively gradual dynamics. This type represents streams that are maintained and accumulated stably within specific application areas. Although they persist over time, they tend to remain localized and specialized rather than functioning as a primary axis that drives the overall knowledge system. This quadrant includes L1, L4, L5, L9, L11, L18, L25, and L26.

III. Core bursty, X at least 0.5 and Y at least 0.5
Core bursty paths occupy central positions in the network with high connectivity and influence while also exhibiting strong spikes during specific transition intervals. This type can be interpreted as reflecting periods of structural reconfiguration in which core axes surge rapidly or central technologies such as standards and platforms intensify over a short period. This quadrant includes L6, L10, L17, L20, L28, and L30.

IV. Core persistent, X at least 0.5 and Y below 0.5
Core persistent paths are structurally central but do not display large growth jumps, indicating relatively gradual growth dynamics. This type has an infrastructure or foundational character in that it is already established as core technology and continues to accumulate and be maintained over time. Even without explosive expansion, such paths perform stable and essential roles in the network and persist in the long run. This quadrant includes L2, L3, L7, L8, L12, and L19.

In addition, L29 lies on the Y equals 0.5 boundary between persistent and bursty patterns, indicating a path whose classification can be sensitive to the choice of the threshold.

To assess the sensitivity of the four-quadrant typology to the 0.5 median threshold, two alternative boundary values---0.40 and 0.60---were applied to the same structural position and growth pattern ranks. Under the 0.40 threshold, seven additional lineages mi-grated from Peripheral Persistent to Peripheral Bursty (primarily lower-spike lineages crossing the relaxed Y-boundary), while the Core Bursty set remained fully intact. Under the 0.60 threshold, four boundary-proximate lineages (L29, L19, L11, L25) shifted quad-rant, but the six primary Core Bursty lineages (L6, L10, L17,

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 March 2026 doi:10.20944/preprints202603.0320.v2



L20, L28, L30) and the two anchor Core Persistent lineages (L2, L7) were classified identically under all three thresh-old values. Additionally, a k-means-based segmentation of the structural score and spike indicator distributions (k=4, k-means++ initialization, 100 runs) produced cluster centroids closely aligned with the four quadrant centers, and 26 of the 30 lineages received identical classifications under both the median-based and k-means-based approaches. These results confirm that the strategic interpretations centered on the Core Bursty and Core Persistent quadrants are robust to threshold selection.

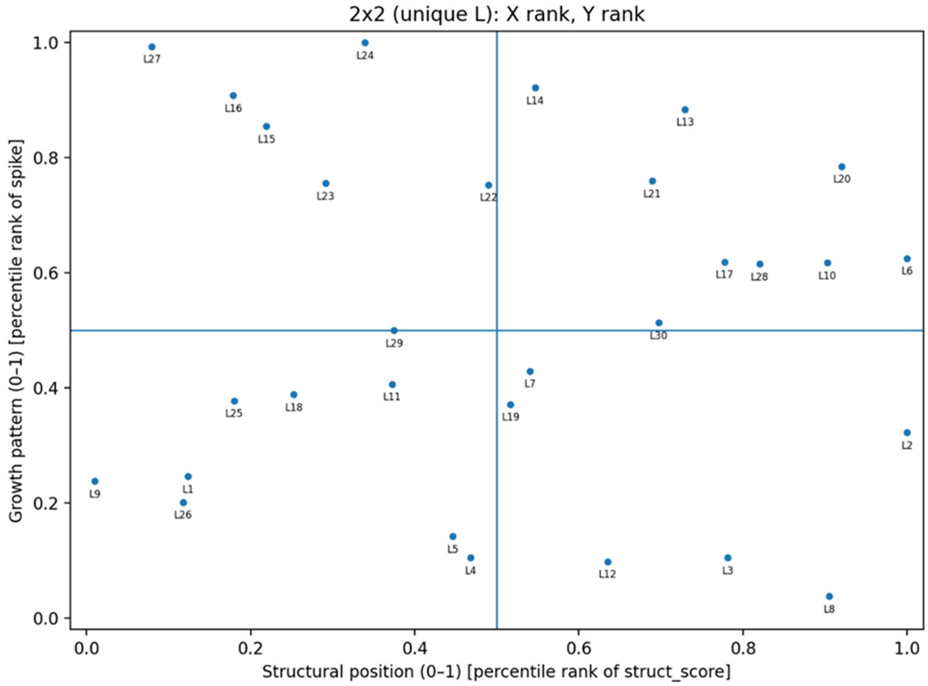


**Figure 13.** Cross-Regime Topic Lineages.

Figure 14 groups the 30 lineage paths by Regime 3 topics T1 through T10 and shows how each topic is distributed in the two by two growth coordinate space. This figure extends the quadrant classification from the lineage level to the Regime 3 topic level, allowing identification of which Regime 3 topics absorb core bursty trajectories and which are more closely associated with peripheral persistent accumulation.

First, the T2 region occupies a wide area in the upper right quadrant, core bursty. This indicates that many lineages converging to T2 rank highly on both structural position, the X rank, and growth pattern, the Y rank. In other words, T2 is strongly associated with trajectories that are both central and rapidly expanding, and it can be interpreted as a representative topic of post transition core expansion in the knowledge flow. In the figure, points cluster in the upper right within the T2 region and the spread is also the largest, suggesting that the core bursty set is largely driven by T2.

Second, the T1 region forms a vertically elongated pattern extending from the left side, low X rank, toward the top, high Y rank. This implies that lineages linked to T1 share a pattern of large growth spikes combined with relatively low structural centrality. Rather than being anchored in the core, T1 is therefore more strongly associated with peripheral bursty trajectories that rise sharply at specific times without fully relocating into the network core. While both T1 and T2 relate to bursty growth, T2 captures bursts occurring in the core, whereas T1 includes relatively more bursts emerging from the periphery, indicating distinct growth mechanisms.

Third, the T7 region appears as a relatively narrow band around the center near the quadrant boundaries, as indicated by the figure title noting that T7 is smaller. This suggests that lineages converging to T7 do not cluster at extreme values in either direction but instead distribute around moderate levels of structural position and growth pattern. Put differently, T7 is less characterized by rapid core reconfiguration and more by trajectories that persist and evolve around intermediate positions without pronounced spikes after the transition.

Fourth, the lower left quadrant, peripheral persistent, contains a clearly separated small region corresponding to T5. This pattern indicates that lineages with both low structural position and low growth pattern tend to connect to T5, implying that T5 is less a topic that drives the core of the knowledge flow and more a topic that is coupled with stable peripheral accumulation. In this sense, T5 exhibits a relatively strong peripheral persistent character.

Overall, these patterns show that Regime 3 topics do not share a uniform growth typology distribution. Instead, topics display distinct combinations of structural position and growth, including core bursty absorption for T2, peripheral bursty association for T1, boundary centered stability for T7, and peripheral persistent association for T5. Accordingly, Figure 14 positions Regime 3 topics not merely as outcome topics but as topics shaped by the types of lineages they absorb after the transition, including core bursty, peripheral bursty, intermediate stable, and peripheral persistent trajectories.

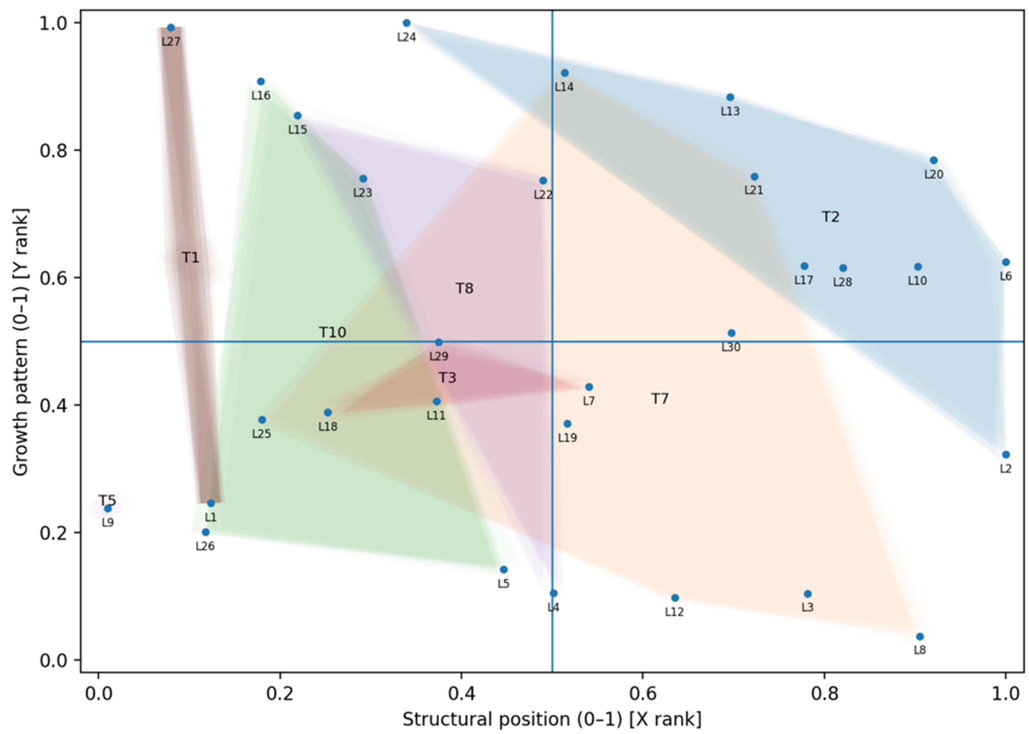


**Figure 14.** Regime 3 Topic-wise Distribution of Growth Types.

## 5. Discussions

### *5.1. Regime Transition and Knowledge Reconfiguration*

This study conceptualizes a regime not simply as a period of publication growth, but as a structurally coherent configuration of the mobile industry knowledge system characterized by relative stability in semantic distribution, topic centrality, and growth typology. Regime boundaries are identified where statistically significant discontinuities in embedding distributions coincide with centroid displacement and dispersion change. This approach differs from calendar-based periodization such as the "3G era" or "5G era" and instead locates transitions in the internal dynamics of knowledge structure itself.

doi:10.20944/preprints202603.0320.v2



This interpretation is broadly consistent with Dosi's (1982) technological paradigm model, which distinguishes incremental development within an established paradigm from transition periods involving fundamental reorientation. In this respect, the 2019–2020 boundary is especially notable. It is marked by a sharp rise in MMD, greater centroid shift, and declining dispersion, suggesting that previously dispersed research themes converged toward a new problem center. The lineage evidence also indicates strong continuity in core technical axes, as reflected in continuation-link similarities above 0.99, while governance and regulatory topics emerged more independently. This suggests that the transition was not a wholesale replacement of prior knowledge, but a selective reorganization combining continuity in foundational domains with the rise of new institutional concerns.

The earlier 2012–2013 transition aligns with the maturation of smartphone diffusion, the mainstreaming of app ecosystems, and the wider adoption of data-driven service models. The independent appearance of platform governance, user acceptance, and bio-mobile convergence topics around this boundary suggests that scholarly knowledge may register early signals of industrial restructuring. Academic publications, therefore, may function not only as a record of industrial change but also as an upstream indicator of shifts in technological and market conditions.

Three implications follow from these findings. First, a knowledge regime can be understood as a quasi-stable distributional state characterized by relatively consistent centroid position and within-distribution dispersion. Second, regime transitions in knowledge-intensive industries appear to proceed mainly through reconfiguration and recombination rather than abrupt substitution of old knowledge axes. Third, a topic's structural and temporal position in the knowledge system carries strategic significance, because not all forms of growth imply the same underlying industrial dynamics.

The 2019–2020 transition warrants particular attention because several competing explanations are possible. One interpretation is that COVID-19 dominated the transition. However, the evidence suggests otherwise. Pandemic-related documents accounted for fewer than 11% of Regime 3 publications, and the centroid trajectory had already begun moving toward 5G optimization and federated learning in mid-2019, prior to the WHO pandemic declaration. Another interpretation is that the shift merely reflects publication growth between 2019 and 2020. Yet the simultaneous decline in dispersion is difficult to reconcile with a pure volume effect. A third interpretation is that regulatory topics emerged only as a reactive response to platform firms' own strategic behavior. However, the independent strengthening of multiple governance-related topics indicates a broader pattern of institutional co-evolution rather than a narrow reaction to firm-level decisions. Taken together, the findings suggest that COVID-19 accelerated an ongoing restructuring, but did not create it from scratch.

### *5.2. Knowledge Recombination and Strategic Implications*

The topic-lineage evidence shows that regime transitions are dominated less by topic extinction than by continuation, merge, and split events. Foundational technical topics from Regime 1 persist into later regimes through strong continuation paths, while other topics are recombined into new configurations or differentiated into more specialized branches. This pattern suggests that knowledge evolution in the mobile industry is cumulative but not merely additive. Existing knowledge axes are preserved, but they are reallocated and recombined in response to changing technological, platform, and institutional conditions.

This pattern can be interpreted through three related mechanisms: knowledge inheritance, knowledge recombination, and knowledge reconfiguration. Inheritance refers to one-to-one continuation links that preserve semantic content across regime boundaries. Recombination refers to merge events in which multiple predecessor topics contribute to a successor topic. Reconfiguration refers to split events in which an existing topic differentiates into multiple successors. Together, these mechanisms provide a more precise interpretation of regime transition than the idea of simple creative destruction. In the present case, the evidence suggests that destruction of prior knowledge structures is limited, whereas recombination and branching are far more common.

 doi:10.20944/preprints202603.0320.v2



This recombination-dominant pattern has direct implications for innovation management. Firms with strong competences in foundational technical domains may retain strategic advantages during periods of transition because such competences can be repurposed in new problem contexts. At the same time, the rise of governance and regulatory topics in Regime 3 shows that mobile industry innovation is no longer driven solely by technical performance, but increasingly reflects interactions among technology, platforms, regulation, and stakeholder environments. The fact that governance-related topics remain peripheral persistent rather than core bursty suggests that regulatory knowledge is growing steadily, but has not yet been fully integrated into the core technical architecture of the knowledge system.

The growth typology developed in this study helps translate these structural differences into strategic implications. Rather than interpreting topic growth only in terms of magnitude, the framework distinguishes whether growth occurs at the center or periphery of the knowledge system and whether it is bursty or persistent. This distinction is important because topics with similar growth rates may imply very different forms of industrial change and therefore require different strategic responses.

Core Bursty topics represent rapid expansion at the center of the knowledge system and indicate areas where architectural innovation and ecosystem competition are especially intense. These domains require timely investment, active participation in standardization, and the development of complementary assets in data, platforms, and services. Because growth in these areas is both central and time-sensitive, firms that enter early may accumulate advantages that become difficult for late entrants to replicate.

Core Persistent topics represent foundational domains that remain central over time while accumulating more gradually. These topics are less associated with short-term surges than with durable capabilities and long-term collaborative advantage. For firms, they function as co-specialized knowledge assets that support future recombination opportunities. Strategic emphasis in these domains should therefore be placed on sustained investment, methodological refinement, and participation in collaborative research and standardization networks.

Peripheral Bursty topics are structurally marginal but may exhibit sharp increases in attention. Their strategic significance lies in uncertainty: some remain issue-driven niches, while others may move toward the core in later periods. These topics are therefore best approached through exploratory and option-preserving investment. A portfolio approach is especially appropriate, allowing firms to maintain small positions across multiple peripheral bursty domains and scale selectively when stronger signals of convergence emerge.

Peripheral Persistent topics accumulate steadily within specialized domains but remain outside the core architecture of the knowledge system. For firms, these topics imply opportunities for niche specialization and differentiated advantage. For policymakers, they point to the need for gradual and predictable institutional refinement rather than abrupt intervention. The case of governance and bio-therapeutic regulation is particularly instructive in this respect, as it suggests that regulatory knowledge is expanding continuously even while remaining structurally peripheral.

Taken together, these findings indicate that the strategic importance of a topic depends not only on how fast it grows, but also on where that growth is located within the broader structure of knowledge flows. Rapid growth in the core is more likely to signal architectural reconfiguration, whereas rapid growth in the periphery may reflect issue-driven attention or emerging convergence potential. This distinction is especially relevant for firms and policymakers seeking to allocate resources under conditions of technological uncertainty.

### *5.3. Methodological Contributions and Implications for Technology Management*

Methodologically, this study makes three main contributions. First, it combines E-Divisive and MMD to identify regime boundaries conservatively through cross-validation of two distinct distributional indicators. This reduces the likelihood of false positives while preserving sensitivity to both sharp and gradual transitions. Second, it introduces a conservative framework for topic-lineage reconstruction using strict inheritance criteria and weighted contribution rules for merge and split classification. Third, it proposes a two-dimensional growth typology that combines structural position with temporal growth pattern, enabling distinction between core reconfiguration and peripheral issue-driven expansion.

Relative to prior approaches, this framework extends static topic modeling by explicitly detecting distributional change and reconstructing cross-regime lineages. It improves on dynamic topic models and temporally unsegmented BERTopic applications by preserving within-regime topical coherence and reducing over-identification of evolutionary links. It also goes beyond prior mobile-industry bibliometric studies that relied mainly on keyword frequency or citation structures, by using embedding-based semantic analysis to detect structural transitions that are not visible in surface-level indicators.

The robustness checks further support the framework. Across 432 hyperparameter combinations, the 2012–2013 and 2019–2020 boundaries were detected in 97.2% of runs. Although the exact number of lineages varied, the classification of the main app-ecosystem and governance-related lineages remained stable in more than 94% of cases. In addition, the 30 focal lineages accounted for 78.4% of the full corpus, indicating that the lineage-level analysis captures most of the knowledge base. The topic-labeling procedure also showed substantial reliability, with Cohen's kappa of 0.82 between two independent raters.

Several boundary conditions should nevertheless be noted. The SPECTER2 model is best suited to STEM-oriented scientific abstracts and may be less appropriate for corpora dominated by legal, financial, or humanities texts. The computational cost of E-Divisive may also become restrictive for much larger corpora. In addition, the median-based quadrant boundaries used in the typology are sensitive to rank but not to absolute magnitude, meaning that analysts interested in stronger distinctions in growth intensity may need supplementary thresholds. Even so, the framework appears well suited to a high-velocity industry such as mobile communications, where technological, platform, and institutional changes overlap within compressed periods.

More broadly, the study has implications for knowledge-based technology management. Conventional portfolio decisions often rely on market signals, patent data, and expert judgment, all of which are subject to lag or bias. The framework proposed here offers a complementary forward-looking signal derived from the semantic structure of research literature itself. In practical terms, MMD can serve as a knowledge turbulence indicator, growth typology can inform portfolio reallocation across emerging domains, and lineage-network indicators can help identify broker topics that connect previously separate knowledge clusters. For policymakers, the parallel accumulation of regulatory and technical knowledge suggests that anticipatory regulation is possible when scholarly knowledge is monitored systematically, rather than only after market failures have become visible.

## 6. Conclusions

This study examines long term changes and regime transitions in the knowledge structure of the mobile industry using embedding based analysis of abstracts from 86,674 mobile industry related publications indexed in Web of Science over the period from 2005 to 2024. The results identify three regimes spanning 2005 to 2012 as Regime 1, 2013 to 2019 as Regime 2, and 2020 to 2024 as Regime 3, with statistically meaningful transitions observed around 2012 to 2013 and 2019 to 2020. The robustness of these regime boundaries is strengthened through cross validation using four indicators, E Divisive, MMD, centroid shift, and dispersion change.

Regime specific topic structures show that Regime 1 is dominated by foundational technologies such as wireless communications, power, sensors, and reliability. Regime 2 exhibits a more polycentric topic configuration as platform and application themes expand alongside increasing convergence with bio related domains. In Regime 3, topics related to 5G operations and data driven services form a central axis of the knowledge structure, while policy, regulation, and governance topics develop in parallel with technical axes.

 doi:10.20944/preprints202603.0320.v2



Topic dynamics results indicate that regime transitions are better characterized as recombination processes built on inherited topics rather than as discontinuous replacement, with intensive reconfiguration phases in which merge and split events concentrate. Growth typology analysis at the complete path level yields four types, core bursty, core persistent, peripheral bursty, and peripheral persistent. App and data driven services emerge as a representative core bursty case, health and body composition research as a representative core persistent case, and bio and therapeutic regulation research as a representative peripheral persistent case.

The study offers three academic contributions. First, it presents a reproducible regime detection pipeline for large scale industrial knowledge corpora by combining SPECTER2 document embeddings with E Divisive change point detection, MMD based distributional distance analysis, and auxiliary validation through centroid shift and dispersion change. Second, it proposes a topic alignment approach that integrates similarity based matching with weighted contribution that incorporates topic size, enabling conservative and systematic classification of inheritance, differentiation, convergence, and disappearance events. Third, the two by two growth typology framework that combines structural position with growth pattern provides an interpretive lens that explains not only what grows but also where growth occurs in the knowledge network and how it unfolds over time, thereby enabling integrated interpretation of knowledge structure change and topic growth dynamics.

Practical contributions emerge at both firm and policy levels. For firms, the rapid expansion patterns exhibited by core bursty topics immediately after regime transitions can serve as leading indicators for timing concentrated R&D investment, platform and standardization engagement, and data infrastructure building. Core persistent topics inform prioritization of capability accumulation in long horizon core areas, while peripheral bursty topics provide candidates for exploratory investments with higher uncertainty but potential upside. From a policy perspective, the emergence of policy and governance topics as an independent axis in Regime 3 suggests that regulatory systems should be designed as parallel components of innovation rather than as subordinate responses to technological development. Linking topic reconfiguration signals around transitions to the timing of regulatory introduction and revision can support evidence based regulatory design that secures necessary safeguards without unduly constraining innovation.

The study has several limitations and directions for future research. First, while focusing on scholarly abstracts ensures consistency for long term tracking, the analysis does not incorporate complementary data on innovation outputs and institutional change such as patents, standards documents, product launches, or regulatory events. Future work should integrate publications with patents, standards, and market and policy event data to identify more precisely the drivers and consequences of regime transitions. Second, because the analysis focuses on a single industry, generalizability to other technology intensive industries is not directly tested. Comparative studies across semiconductors, biotechnology, and artificial intelligence are needed to identify shared mechanisms of regime transitions and industry specific pathways. Third, the study emphasizes ex post identification of regime shifts and the structuring of growth typologies and does not develop predictive models. Future research should integrate transition signals such as distributional distance, centroid translation, and dispersion change with growth typology patterns to build forecasting models that support ex ante estimation of transition likelihood and early detection of emerging core topics. Fourth, although the Discussion addresses competing interpretations of the 2019–2020 transition, the relative causal contributions of 5G commercialization and the COVID-19 pandemic cannot be fully decomposed from publication data alone. While the pre-pandemic centroid trajectory and the limited share of pandemic-specific documents collectively suggest that 5G-driven restructuring was the primary driver, future research incorporating exogenous event data—such as spectrum auction dates, standards release timelines, and policy announcement records—would enable more precise attribution of transition causes.

 doi:10.20944/preprints202603.0320.v2



**Author Contributions:** Conceptualization, S.J., W.J. and K.C.; Methodology, S.J., W.J.; Software, W.J.; Validation, S.J. and K.C.; Formal analysis, W.J.; Writing—original draft preparation, S.J.; Writing—review and editing, S.J., W.J. and K.C.; Supervision, K.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data used to support the findings of this study are included in the article.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Campbell-Kelly, M.; Garcia-Swartz, D. D.; Lam, R.; Yang, Y. Economic and business perspectives on smartphones as multi-sided platforms. Telecommunications Policy 2015, 39(8), pp. 717–734.
2. Kenney, M.; Pon, B. Structuring the smartphone industry: Is the mobile internet OS platform the key? Journal of Industry, Competition and Trade 2011, 11(3), pp. 239–261.
3. Cecere, G.; Corrocher, N.; Battaglia, R. D. Innovation and competition in the smartphone industry: Is there a dominant design? Telecommunications Policy 2015, 39(3–4), pp. 162–175.
4. Henten, A.; Windekilde, I. “Demand-Side Economies of Scope in Big Tech Business Modelling and Strategy.” Systems 2022, 10(6), 246.
5. Xu, C.; Wang, Y.-M. “The Regulatory Architecture of Digital Platforms: A Perspective of Life Cycle and Risk Management.” Systems 2022, 10(5), 145.
6. Bostoen, F. Understanding the Digital Markets Act. The Political Quarterly 2023, 94(4), pp. 1–12.
7. Al Moteri, M.; Khan, S.B.; Alojail, M. “Machine Learning-Driven Ubiquitous Mobile Edge Computing as a Solution to Network Challenges in Next-Generation IoT.” Systems 2023, 11(6), 308.
8. Alberti, E.; Alvarez-Napagao, S.; Anaya, V.; et al. “AI Lifecycle Zero-Touch Orchestration within the Edge-to-Cloud Continuum for Industry 5.0.” Systems 2024, 12(2), 48.
9. Santha Kumar, R.; Kaliyaperumal, K. A scientometric analysis of mobile technology publications. Scientometrics 2015, 105(2), pp. 921–939.
10. Lee, S.; Kim, W. The knowledge network dynamics in a mobile ecosystem: A patent citation analysis. Scientometrics 2017, 111(2), pp. 717–742.
11. Heikkilä, M.; Heikkilä, J.; Ahmad, F. “Data-Driven Business Model Innovation in Europe: Ethical Data Practices and Ecosystem Involvement.” Systems 2025, 13(3), 164.
12. Aridor, G.; Che, Y.-K. Privacy regulation and targeted advertising: Evidence from Apple’s App Tracking Transparency. Working Paper 2024.
13. Rafieian, O.; Yoganarasimhan, H. Targeting and privacy in mobile advertising. Marketing Science 2021, 40(2), pp. 193–218.
14. Cohan, A.; Feldman, S.; Beltagy, I.; Downey, D.; Weld, D. S. SPECTER: Document-level representation learning using citation-informed transformers. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL) 2020, pp. 2270–2282.
15. Singh, A.; D’Arcy, M.; Cohan, A.; Downey, D.; Feldman, S. SciRepEval: A multi-format benchmark for scientific document representations. In Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP) 2023.
16. Rochet, J.-C.; Tirole, J. Platform Competition in Two-Sided Markets. Journal of the European Economic Association 2003, 1(4), pp. 990–1029.
17. Armstrong, M. Competition in Two-Sided Markets. RAND Journal of Economics 2006, 37(3), pp. 668–691.
18. Rysman, M. The Economics of Two-Sided Markets. Journal of Economic Perspectives 2009, 23(3), pp. 125–143.
19. Eisenmann, T.; Parker, G.; Van Alstyne, M. Platform Envelopment. Strategic Management Journal 2011, 32(12), pp. 1270–1285.
20. Jacobides, M. G.; Cennamo, C.; Gawer, A. Towards a Theory of Ecosystems. Strategic Management Journal 2018, 39(8), pp. 2255–2276.

 doi:10.20944/preprints202603.0320.v2




21. Adner, R. Ecosystem as Structure: An Actionable Construct for Strategy. Journal of Management 2017, 43(1), pp. 39–58.
22. Kapoor, R.; Lee, J. M. Coordinating and Competing in Ecosystems: How Organizational Forms Shape New Technology Investments. Strategic Management Journal 2013, 34(3), pp. 274–296.
23. Gawer, A.; Cusumano, M. A. Industry Platforms and Ecosystem Innovation. Journal of Product Innovation Management 2014, 31(3), pp. 417–433.
24. Tiwana, A. Platform Ecosystems: Aligning Architecture, Governance, and Strategy. Morgan Kaufmann 2014.
25. Zhu, F.; Liu, Q. Competing with Complementors: An Empirical Look at Amazon.com. Strategic Management Journal 2018, 39(10), pp. 2618–2642.
26. Cennamo, C.; Santalo, J. Platform Competition: Strategic Trade-offs in Platform Markets. Strategic Management Journal 2013, 34(11), pp. 1331–1350.
27. Evans, D. S.; Schmalensee, R. The Antitrust Analysis of Multi-Sided Platform Businesses. In Oxford Handbook of International Antitrust Economics 2015, pp. 404–447.
28. Crémer, J.; de Montjoye, Y.-A.; Schweitzer, H. Competition Policy for the Digital Era. European Commission Report 2019.
29. Rietveld, J.; Schilling, M. A. Platform Competition: A Systematic and Interdisciplinary Review of the Literature. Journal of Management 2021, 47(6), pp. 1528–1563.
30. Andrews, J. G.; Buzzi, S.; Choi, W.; Hanly, S. V.; Lozano, A.; Soong, A. C. K.; Zhang, J. C. What Will 5G Be? IEEE Journal on Selected Areas in Communications 2014, 32(6), pp. 1065–1082.
31. Dahlman, E.; Parkvall, S.; Sköld, J. 5G NR: The Next Generation Wireless Access Technology. Academic Press 2018.
32. Shi, W.; Cao, J.; Zhang, Q.; Li, Y.; Xu, L. Edge Computing: Vision and Challenges. IEEE Internet of Things Journal 2016, 3(5), pp. 637–646.
33. Mach, P.; Becvar, Z. Mobile Edge Computing: A Survey on Architecture and Computation Offloading. IEEE Communications Surveys & Tutorials 2017, 19(3), pp. 1628–1656.
34. Taleb, T.; Samdanis, K.; Mada, B.; Flinck, H.; Dutta, S.; Sabella, D. On Multi-Access Edge Computing: A Survey of the Emerging 5G Network Edge Cloud Architecture and Orchestration. IEEE Communications Surveys & Tutorials 2017, 19(3), pp. 1657–1681.
35. ETSI ISG ZSM. Zero-touch Network and Service Management (ZSM): Reference Architecture. ETSI GR ZSM 002 2019.
36. Newman, M. E. J. The Structure and Function of Complex Networks. SIAM Review 2003, 45(2), pp. 167–256.
37. Borgatti, S. P.; Everett, M. G.; Johnson, J. C. Analyzing Social Networks. SAGE Publications 2018.
38. Kessler, M. M. Bibliographic Coupling Between Scientific Papers. American Documentation 1963, 14(1), pp. 10–25.
39. Small, H. Co-citation in the Scientific Literature: A New Measure of the Relationship Between Two Documents. Journal of the American Society for Information Science 1973, 24(4), pp. 265–269.
40. Hummon, N. P.; Doreian, P. Connectivity in a Citation Network: The Development of DNA Theory. Social Networks 1989, 11(1), pp. 39–63.
41. Batagelj, V. Efficient Algorithms for Citation Network Analysis. 2003.
42. Park, H.; Magee, C. L.; et al. Tracing Technological Development Trajectories: A Genetic Knowledge Persistence-Based Main Path Approach. PLOS ONE 2017.
43. von Wartburg, I.; Teichert, T.; Rost, K. Inventive Progress Measured by Multi-stage Patent Citation Analysis. Research Policy 2005, 34(10), pp. 1591–1607.
44. Oh, M.; et al. Main Path Analysis for Technological Development Using Patent Documents. Scientometrics 2023.
45. Han, B.; et al. 5G Wireless Technology Evolution: Patent-based Main Path and Network Analysis. Wireless Networks 2024.
46. Kleinberg, J. Bursty and Hierarchical Structure in Streams. Data Mining and Knowledge Discovery 2003.

47. Chen, C. CiteSpace II: Detecting and Visualizing Emerging Trends and Transient Patterns in Scientific Literature. Journal of the American Society for Information Science and Technology 2006.
48. Farooqui, M. N. I.; et al. A Bibliometric Approach to Quantitatively Assess Current Status of 5G Security Research. Telematics and Informatics 2017.
49. Suh, Y.; Jeon, J. Monitoring Patterns of Open Innovation Using the Patent-based Brokerage Analysis. Technological Forecasting and Social Change 2019, 146, pp. 595–605.
50. Chesbrough, H. W. Open Innovation: The New Imperative for Creating and Profiting from Technology. Harvard Business School Press 2003.
51. van Eck, N. J.; Waltman, L. Software Survey: VOSviewer, a Computer Program for Bibliometric Mapping. Scientometrics 2010, 84(2), pp. 523–538.
52. Waltman, L.; van Eck, N. J.; Noyons, E. C. M. A Unified Approach to Mapping and Clustering of Bibliometric Networks. Journal of Informetrics 2010, 4(4), pp. 629–635.
53. Aria, M.; Cuccurullo, C. bibliometrix: An R-tool for Comprehensive Science Mapping Analysis. Journal of Informetrics 2017, 11(4), pp. 959–975.
54. Matteson, D. S.; James, N. A. A Nonparametric Approach for Multiple Change Point Analysis of Multivariate Data. Journal of the American Statistical Association 2014, 109(505), pp. 334–345.
55. Killick, R.; Fearnhead, P.; Eckley, I. A. Optimal Detection of Changepoints with a Linear Computational Cost. Journal of the American Statistical Association 2012, 107(500), pp. 1590–1598.
56. Truong, C.; Oudre, L.; Vayatis, N. Selective Review of Offline Change Point Detection Methods. Signal Processing 2020, 167, 107299.
57. Székely, G. J.; Rizzo, M. L. Energy Statistics: A Class of Statistics Based on Distances. Journal of Statistical Planning and Inference 2013, 143(8), pp. 1249–1272.
58. Sejdinovic, D.; Sriperumbudur, B.; Gretton, A.; Fukumizu, K. Equivalence of Distance-based and RKHS-based Statistics in Hypothesis Testing. Annals of Statistics 2013, 41(5), pp. 2263–2291.
59. Gretton, A.; Borgwardt, K. M.; Rasch, M. J.; Schölkopf, B.; Smola, A. A Kernel Two-Sample Test. Journal of Machine Learning Research 2012, 13, pp. 723–773.
60. Griffiths, T. L.; Steyvers, M. Finding Scientific Topics. Proceedings of the National Academy of Sciences 2004, 101(Suppl. 1), pp. 5228–5235.
61. Wang, X.; McCallum, A. Topics over Time: A Non-Markov Continuous-Time Model of Topical Trends. In Proceedings of KDD 2006, pp. 424–433.
62. Blei, D. M.; Lafferty, J. D. Dynamic Topic Models. In Proceedings of ICML 2006, pp. 113–120.
63. Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of NAACL-HLT 2019, pp. 4171–4186.
64. Beltagy, I.; Lo, K.; Cohan, A. SciBERT: A Pretrained Language Model for Scientific Text. In Proceedings of EMNLP-IJCNLP 2019, pp. 3615–3620.
65. McInnes, L.; Healy, J.; Saul, N.; Großberger, L. UMAP: Uniform Manifold Approximation and Projection. Journal of Open Source Software 2018, 3(29), 861.
66. Campello, R. J. G. B.; Moulavi, D.; Sander, J. Density-Based Clustering Based on Hierarchical Density Estimates. In PAKDD 2013, pp. 160–172.
67. McInnes, L.; Healy, J.; Astels, S. hdbscan: Hierarchical Density Based Clustering. Journal of Open Source Software 2017, 2(11), 205.
68. Grootendorst, M. BERTopic: Neural Topic Modeling with a Class-Based TF-IDF Procedure. arXiv 2022, arXiv:2203.05794.
69. Cui, W.; Liu, S.; Tan, L.; Shi, C.; Song, Y.; Gao, Z.; Tong, X.; Qu, H. TextFlow: Towards Better Understanding of Evolving Topics in Text. IEEE Transactions on Visualization and Computer Graphics 2011, 17(12), pp. 2412–2421.
70. Cui, W.; Liu, S.; Wu, Z.; Wei, H. How Hierarchical Topics Evolve in Large Text Corpora. IEEE Transactions on Visualization and Computer Graphics 2014, 20(12), pp. 2281–2290.

 doi:10.20944/preprints202603.0320.v2



71. Chen, C. Searching for Intellectual Turning Points: Progressive Knowledge Domain Visualization. Proceedings of the National Academy of Sciences 2004, 101(Suppl. 1), pp. 5303–5310.
72. Churchill, R.; Singh, L. The Evolution of Topic Modeling. ACM Computing Surveys 2022, 55(4), pp. 1–38.