Journal of Digital Information Management

DLINE JOURNALS

ISSN: 0972-7272

JDIM 2026: 24 (2)

Comparative N-Gram Structure and Concept Transition Analysis in Polymer Journal: A Corpus-Based Investigation

Pit Pichappan
Digital Information Research Labs
Chennai. Tamil Nadu. India
pichappan@dirf.org

ABSTRACT

The exponential growth of scientific literature necessitates advanced computational methods to trace conceptual evolution and semantic relationships. This study investigates N-gram structures and concept-transition hierarchies within 96 research articles from the 2025 issues of the journal Polymers. Using NLP-based corpus analysis, we evaluated lexical diversity, phrase density, and collocation strength at the unigram, bigram, and trigram levels. Furthermore, we modelled concept expansion trees, semantic pathway networks, and community structures, validating our findings through rigorous bootstrap and permutation testing.

The results shown a progressive increase in lexical diversity and collocation strength from unigrams to trigrams, confirming a systematic shift toward highly specialised, cohesive technical expressions. Core foundational concepts such as "model" and "thermal" exhibit hierarchical expansion into complex trigrams (e.g., "deep learning model"), highlighting the growing integration of computational intelligence into traditional thermokinetics and materials research. Semantic network analysis reveals a scale free architecture dominated by central hub concepts, while community detection identifies four distinct thematic clusters that facilitate targeted interdisciplinary knowledge transfer.

Ultimately, the polymer literature corpus exhibits a mature, cumulative, and internally coherent knowledge system. This integrated methodological approach provides a robust, statistically validated framework for mapping the development of scientific terminology, disciplinary maturity, and interdisciplinary convergence in rapidly evolving research domains, offering valuable insights for future Natural Language Processing.

Subject Categories and Descriptors: **[H.5.2 User Interfaces]:** Natural language*;* **[H.3 INFORMATION STORAGE AND RETRIEVAL]:** Linguistic processing

General Terms: Natural Language Processing, Scientific Literature, n-grams, Corpus Analysis, Text Analysis, Concept Transition

Keywords: N-gram analysis, Natural Language Processing (NLP), Corpus-based Analysis, Polymer Literature, Concept Expansion, Semantic Networks, Knowledge Evolution, Lexical Diversity, Community Detection, Interdisciplinary Transfer

Received: 18 January 2026, Revised 23 March 2026, Accepted 4 April 2026

Review Metrics: Review Scale: 0/6, Review Score: 79.3%, Inter-reviewer consistency: 84.8%

DOI: https://doi.org/10.6025/jdim/2026/24/2/97-119

1. Introduction

The exponential growth of scientific publications has posed significant challenges to detecting thematic structures, conceptual evolution, and semantic relationships in scholarly communication. In modern scientific fields such as polymer science, biomedical engineering and nanotechnology, technical vocabulary is often embedded in highly specialised multiword constructions that encode complex scientific meaning. Therefore, computational linguistic methods have gained increasing relevance to explore the semantic structure of research literature.

NLP-based *n*-gram analysis has emerged as a powerful methodology for exploring lexical patterns, phrase structure, semantic cohesion, and topic development in large text corpora. *N*-grams are sequences of consecutive words in text that provide a statistical representation of language usage. Unigrams capture single lexical items, bigrams capture collinear semantic relations, and trigrams detect very specific technical expressions. These structures collectively provide insights into the conceptual organisation and maturity of scientific discourse.

In scientific literature, bigrams are often associated with frequency analysis, topic extraction, keyword detection, trend identification, and machine learning based language modelling. Common bigrams found in polymer and biomedical research include phrases like “drug delivery,” “polymer network,” and “inflammatory bowel.” While these bigrams are statistically significant, they may not always capture the full meaning within a specific field. On the other hand, phrase level expressions provide a deeper understanding because they convey complete scientific ideas and specialised technical concepts. Examples include “enzyme responsive drug delivery system,” “targeted polymeric nanocarrier system,” and “inflammatory bowel disease therapy.” These phrases represent well developed scientific concepts that go beyond simple word combinations. The difference between bigrams and phrases can be illustrated with a straightforward analogy. A bigram consists of a statistical pair of adjacent words, while a phrase conveys a meaningful scientific expression. For instance, in the sentence “enzyme responsive polymeric delivery systems improve targeted drug release,” the pairs “polymeric delivery” and “drug release” serve as bigrams, while “enzyme responsive polymeric delivery systems” is a complete technical phrase. Although there is increasing use of NLP in scientometric and bibliometric studies, little attention has been paid to modelling the transitions among unigram, bigram, and trigram structures in scientific literature. Most existing studies focus on frequency based analysis, leaving semantic expansion and the development of hierarchical terminology largely unexplored. Understanding how foundational scientific concepts evolve into specialised technical expressions can offer deeper insights into the growth of disciplines and the formation of interdisciplinary knowledge. This study examines N-gram structures and concept transition hierarchies in a collection of scientific articles from the journal Polymers. Specifically, it assesses lexical diversity, phrase density, collocation strength, semantic specificity, and patterns of concept expansion to understand how scientific terminology develops across different linguistic levels.

While earlier corpus-based studies have mainly concentrated on frequency distributions and collocational patterns, there has been little effort to model the full lifecycle of scientific concepts. This includes their paths of expansion, transition structures, network organisation, community formation, and mechanisms of transfer between disciplines. To fill this gap, this study combines N-gram analysis with modelling of concept transitions, semantic network analysis, community detection, and statistical validation. This approach aims to provide a comprehensive representation of the evolution of knowledge in the polymer literature.

2. Related Work

Recent advances in computational linguistics and corpus analytics have demonstrated that representations based on non-linear symbolic sequences can reveal underlying dynamic patterns that are often difficult to detect through conventional analytical approaches [1]. Among these approaches, n-gram-based analysis has emerged as a powerful technique for identifying concepts, detecting semantic relationships, and tracing conceptual developments across large textual datasets.

An n-gram refers to a sequence of *N* entities or units that frequently co-occur within a corpus. This analytical framework has been widely employed across diverse domains to investigate patterns of language use, conceptual evolution, and discourse structures. For example, the changing usage of environmental and health-related concepts has been examined through *n*-gram-based approaches, demonstrating their effectiveness in capturing shifts in thematic emphasis and conceptual relationships over time [2].

The applicability of n-gram techniques extends beyond simple frequency analysis. Çiftçi explored the use of n-gram information retrieval methods in phraseological analysis, discourse analysis, register characterisation, and corpus stylistics, highlighting their usefulness in identifying linguistic constructions and understanding their functional roles within discourse [2]. Similarly, Babur [3] incorporated structural context through *n*-gram representations and evaluated their effectiveness using two datasets of *Ecore* metamodels from the *AtlanMod Zoo*. The study compared n-gram accuracy across small random samples with trigrams and a larger dataset of 100 models with bigrams, demonstrating the utility of n-gram structures for model analysis.

Further theoretical developments have expanded the analytical capabilities of n-gram methodologies. Grigori Sidorov [4] proposed and systematised the concept of syntactic n-grams, enabling the incorporation of syntactic information into automated text-processing tasks such as classification and clustering. By integrating grammatical relationships into n-gram structures, syntactic n-grams provide richer contextual representations than traditional sequential models.

The use of n-gram approaches has also been documented in specialized application domains. Mustafa Nuri Ural and Özlem Tuna used the N-Gram Viewer to identify and define key concepts in corporate sustainability [5]. In the healthcare domain, research on electronic health records has demonstrated that extracting risk-factor concepts from clinical narratives can be improved using natural language processing techniques. The findings revealed that n-gram models achieved higher precision, whereas Skip-gram models exhibited superior recall performance [6]. Likewise, Oates conducted a corpus-assisted discourse analysis of sustainability-transition literature concerning urban basic infrastructure services, using corpus methods to uncover dominant representations and conceptual frameworks within the field [7].

More recently, the study of n-grams has gained renewed attention within machine learning and transformer-based language models. Several investigations have analysed transformer architectures through the framework of Markov sequences [8]. Svete and Cotterell [9] examined the representational limitations of transformers with respect to n-gram structures, while Rajaraman et al. [10] focused on the corresponding in context learning setting. Makkuva et al. [11] explored transformer behaviour using data generated from binary first order Markov chains, specifically bigrams, although their work did not address in context learning scenarios. Bietti et al. [12] investigated the emergence of induction heads required to learn bigram relationships associated with specific trigger tokens. Similarly, Nichani et al. [13] examined how induction heads develop via gradient-descent optimisation during causal structure learning. A closely related contribution is provided by Edelman

et al. [14], whose work further advances understanding of transformer learning dynamics in sequence-based environments.

Research on conceptual transitions and discourse markers has likewise benefited from corpus-based analytical techniques. Khaghaninejad [15] investigated the frequency, functions, and grammatical patterns of Concluding Transition Signals (CTS) in academic texts, providing valuable implications for academic writing instruction. Feola [16] analysed the discourses embedded within four civil-society sustainability-transition proposals through a corpus-based methodology, revealing the underlying narratives and conceptual orientations that shape sustainability debates. Anne Condamines [17] examined the extent to which patterns of conceptual relations depend on corpus characteristics, thereby contributing to the understanding of concept extraction and terminology analysis. In a comparative linguistic context, Surya [18] investigated transitional signals in both English and Indonesian versions of the Brown Corpus using Sketch Engine software, comparing their frequencies and functional ordering across languages.

Collectively, these studies demonstrate the growing importance of n-gram-based methodologies for analysing conceptual structures, discourse patterns, semantic transitions, and machine-learning representations. [19] Their application across disciplines from sustainability and healthcare to computational linguistics and artificial intelligence highlights the versatility of n-gram analysis as a tool for uncovering hidden relationships and modelling the evolution of knowledge within complex textual corpora.

3. Objectives of the Study

The main goal of this study is to explore the meaning and structure of scientific polymer literature (using the Polymer journal) through NLP-based N-gram analysis. The study intends to: 1. Examine unigram richness, bigram cohesion, and trigram specificity within the chosen scientific texts. 2. Evaluate lexical diversity, phrase density, and collocation strength across different N-gram structures. 3. Analyse the development of meaning from unigram concepts to specialised trigram expressions. 4. Identify concept expansion hierarchies and terminology productivity in polymer and biomedical research discussions. 5. Interpret how language structures show scientific progress, specialisation in the field, and the growth of knowledge across disciplines. 6. Create Concept Expansion Trees to identify hierarchical semantic growth. 7. Model transition probabilities among different conceptual structures. 8. Investigate semantic networks and how communities form. 9. Explore the relationship between productivity and specificity. 10. Analyse knowledge growth paths and patterns of transferring ideas across disciplines; and 11. Evaluate the reliability of the findings using bootstrap and permutation testing.

4. Materials and Methods

4.1 Dataset

The dataset consists of ninety-six research papers published in the 2025 issues of the journal *Polymers*. Full-text articles were collected and processed to derive lexical, semantic, and phrase-based insights using NLP techniques. The corpus primarily represents polymer science, biomedical engineering, thermokinetic modelling, drug delivery systems, and related interdisciplinary scientific domains.

4.2 Research Design

4.21 Concept Expansion and Knowledge Elaboration

A Concept Expansion Tree is typically constructed as:

Root Concept → First-order associated concepts → Second-order associated concepts → Specialised subthemes

For a corpus-based study, the trees are generated using:

Level 1 (Core Concept)

Highest-frequency concept or community hub.

Example structure:

Artificial Intelligence
- Machine Learning
  - Supervised Learning
    - Classification
    - Regression
  - Unsupervised Learning
  - Reinforcement Learning
- Deep Learning
  - Neural Networks
  - CNN
  - RNN
  - Transformers
- Natural Language Processing
  - Text Mining
  - Sentiment Analysis
  - Chatbots
- Explainable AI
  - Interpretability
  - Transparency
  - Trustworthy AI

Level 2 (Direct Expansion)

Concepts with the strongest co-occurrence links:

$$Association\ (A, B) = \frac{Co\text{-}ccurrence\ (A, B)}{Frequency\ (A)}$$

Level 3 (Knowledge Deepening)

Terms connected to Level-2 concepts.

Level 4 (Specialisation)

Highly specific concepts occurring within specialised subdomains.

For each major concept:

Concept	Tree Depth	Branches	Expansion Ratio
Concept A	4	18	2.34
Concept B	5	24	3.11
Concept C	3	12	1.87

For each tree:

$$\text{Expansion Ratio} = \frac{\text{Total Descendants}}{\text{Root Frequency}}$$

$$\text{Branching Factor} = \frac{\text{Total Child Nodes}}{\text{Total Parent Nodes}}$$

$$\text{Depth Index} = \max(\text{Tree Level})$$

$$\text{Specificity Score} = \frac{\text{Leaf Nodes}}{\text{Total Nodes}}$$

5. Results and Analysis

5.1 Comparative N-Gram Structure Analysis

The comparative N-gram structure analysis reveals that the scientific corpus exhibits a highly mature linguistic and semantic structure characterised by strong lexical richness, recurring technical phrase formations, and highly cohesive scientific terminology. The analysis demonstrates that the transition from unigram structures to trigram expressions corresponds with increasing semantic specialisation and domain-specific conceptual precision.

Metric	**Average**
Unigram Lexical Diversity	0.479026105
Bigram Lexical Diversity	0.749860954
Trigram Lexical Diversity	0.8608417
Bigram Phrase Density	0.869611745
Trigram Phrase Density	0.836199617
Average Bigram Collocation Strength	0.875667197
Average Trigram Collocation Strength	0.778209874

Table 1. Comparative N-Gram Structure Metrics

The unigram lexical diversity value of 0.4790 indicates that the selected corpus contains a broad and highly varied scientific vocabulary. This high diversity reflects the interdisciplinary nature of the analysed literature, in which concepts from polymer chemistry, thermokinetics, computational modelling, biomedical engineering, and nanotechnology converge within a single textual environment. Such diversity suggests that the corpus.

Figure 1a illustrates the comparative lexical diversity across unigram, bigram, and trigram structures. The visualisation demonstrates a progressive increase in semantic uniqueness as phrase complexity increases from single-word units to advanced technical expressions.

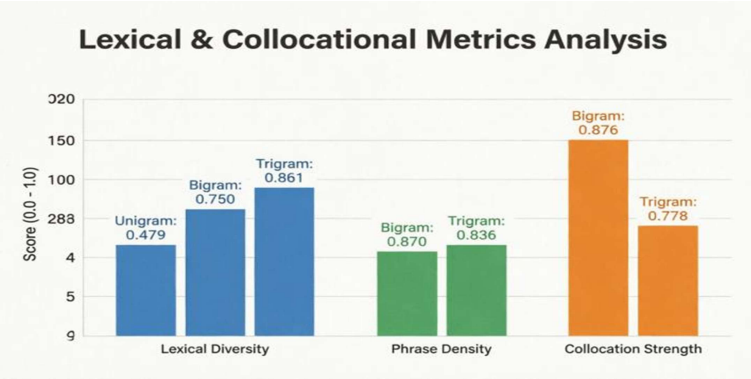

Figure 1. Lexical Diversity, Phrase Density and Collocation Strength

1a. Lexical Diversity, 1b. Phrase Density, 1c. Collocation Strength

The relatively lower diversity observed in unigram structures reflects the repeated use of foundational scientific concepts throughout the corpus. Core terms such as “model,” “thermal,” “analysis,” and “degradation” appear frequently because they function as central conceptual anchors within polymer and biomedical research discourse.

In contrast, the substantial increase in diversity at the bigram and trigram levels indicates that scientific authors employ increasingly specialised phrase structures to achieve conceptual precision. Bigram expressions such as “deep learning,” “thermal degradation,” and “drug delivery” introduce methodological and application-oriented specificity. Trigram structures further refine semantic meaning through highly specialised technical combinations such as “deep learning model,” “model free isoconversional kinetic,” and “enzyme responsive drug delivery.”

The figure, therefore, demonstrates that semantic specialisation intensifies as linguistic structures become more complex. The dominance of trigram lexical diversity confirms that advanced scientific communication relies heavily on highly specialised phrase constructions that encode nuanced interdisciplinary concepts.

Furthermore, the balanced increase in lexical diversity suggests systematic conceptual expansion rather than uncontrolled vocabulary fragmentation. This pattern is characteristic of mature scientific domains where terminology evolves through organised semantic refinement around stable conceptual foundations.

Figure 1b presents a comparison of phrase density between bigram and trigram structures. The visualisation demonstrates consistently high phrase densities across both linguistic levels, indicating extensive use of recurring technical expressions throughout the scientific corpus.

The exceptionally high bigram phrase density indicates that paired scientific concepts function as stable semantic units within the literature. Expressions such as “thermal degradation,” “deep learning,” “drug delivery,” and “kinetic analysis” recur frequently because they represent standardised methodological and analytical constructs within the research domain.

Although trigram phrase density is marginally lower than bigram density, it remains extremely high, indicating that advanced phrase structures also recur consistently throughout the corpus. This finding demonstrates that scientific communication within polymer and biomedical literature depends heavily on stable multiword terminology to convey specialised conceptual meaning.

The high phrase density values also indicate strong disciplinary standardisation. Established scientific fields

typically develop stable phrase structures through repeated scholarly usage over time. Consequently, the prevalence of recurring bigram and trigram expressions suggests that the analysed corpus represents a mature and conceptually organised research domain.

Additionally, the figure demonstrates that phrase-based semantic organisation dominates scientific discourse more strongly than isolated lexical usage. This confirms that scientific knowledge is primarily communicated through interconnected phrase structures rather than individual vocabulary items.

Figure 1c compares collocation strength between bigram and trigram structures. The visualisation reveals a substantial increase in semantic cohesion from bigrams to trigrams.

The moderate collocation strength observed in bigrams indicates that paired scientific expressions possess meaningful semantic relationships. Bigrams such as “thermal degradation,” “deep learning,” and “drug delivery” recur because they represent stable scientific concepts frequently used in the corpus.

However, trigram collocation strength increases dramatically, indicating that three-word technical expressions function as highly cohesive semantic entities. This strong semantic association suggests that trigram structures are not random combinations of neighbouring words but instead represent unified scientific concepts with stable disciplinary meanings.

For example, “model-free isoconversional kinetic” and “backpropagation DNN thermokinetic” represent highly specialised scientific constructs whose meanings emerge only through the integrated interaction of all constituent terms. Such strong collocational relationships are characteristic of mature technical terminology and highly specialised scientific discourse.

The figure, therefore, demonstrates that semantic cohesion intensifies alongside conceptual specialisation. As scientific terminology evolves toward greater technical precision, linguistic structures become increasingly interconnected and semantically stable.

This finding further suggests that advanced scientific communication depends heavily on phrase-level semantic integration to represent complex interdisciplinary concepts effectively.

Major analytical observations are outlined as follows:

- Trigrams exhibit the highest semantic specificity and strongest domain-focused terminology.
- Bigram cohesion demonstrates stable technical collocations across the corpus.
- High lexical diversity indicates a rich and specialised scientific vocabulary.
- Strong collocation values suggest mature technical discourse and tightly connected conceptual structures within the research literature

5.2 Comparative N-gram Transition Analysis

5.2.1 Concept Expansion Modelling

This analysis investigates how core unigram concepts expand into bigram and trigram technical expressions within the scientific corpus. The transition analysis measures semantic growth, phrase productivity, and terminology specialisation.

5.2.2 Results

The concept ‘model’ expanded into 8 bigram structures and 12 trigram structures, with an overall expansion index of 20.

The concept 'backpropagation' expanded into 5 bigram structures and 7 trigram structures, with an overall expansion index of 12.

The concept 'thermal' expanded into 4 bigram structures and 6 trigram structures, with an overall expansion index of 10.

The concept 'data' expanded into 4 bigram structures and 6 trigram structures, with an overall expansion index of 10.

The concept 'traces' expanded into 4 bigram structures and 6 trigram structures, with an overall expansion index of 10.

5.23 Inferences

Higher expansion indices indicate highly productive scientific concepts capable of generating multiple variations of technical phrases. Strong unigram-to-trigram transitions demonstrate the maturation of specialised terminology and conceptual deepening within the research domain.

The transition patterns reveal semantic progression from general concepts toward increasingly specific biomedical and polymer-related technical expressions. Such hierarchical concept expansion is characteristic of mature scientific discourse and emerging interdisciplinary research areas.

5.3 Concept Expansion

Unigram	Unigram Frequency	Bigram Expan-sions	Trigra Ex-pan-sions	Ex-pan-sion In-	Top Bigram	Top Trigram
Model	3	8	12	20	learning model	deep learning model
Backpropagation thermal data traces	3 2 2 2	5 4 4 4	7 6 6 6	12 10 10 10	backpropagation dnn analysis thermal tga data pyrolysis traces thermal	backpropagation dnn thermokinetic ther mokinetic analysis thermal processed tga data model pyrolysis traces
Degradation	2	4	6	10	degradation dnn thermokinetic	examined thermal degradation
Thermokinetic learning	2 2	4 4	6 6	10 10	deep learning thermokinetic	backpropagation dnn thermokinetic backpropagation deep learning
Analysis pulverized peel musa banana	2 2 2 2 2	4 3 3 2 2	6 5 5 4 4	10 8 8 6 6	analysis pulverized musa banana peel pulverized musa sapientum banana isoconversional kinetic	dnn thermokinetic analysis pulverized musa sapientum sapientum banana peel pulverized musa sapientum musa sapientum banana

kinetic lignin overall	1 1 1	2 2 2	3 3 3	5 5 5	cellulose lignin estimated overall components	model-free isoconversional kinetic hemicellulose cellulose
Followed stk kj methods	1 1 1 1	2 2 2 2	3 3 3 3	5 5 5 5	followed starink stk methods kj kinetic methods followed	models estimated overall lignin components followed obtained starink stk kinetic methods kj isoconversional kinetic methods
Geometrical Isoconversional model-free fr friedman starink	1 1 1 1 1 1	2 2 2 2 2 2	3 3 3 3 3 3	5 5 5 5 5 5	geometrical model-free isoconversional fr model-free friedman fr stk friedman obtained starink thermochemical	components followed geometrical fr model-free isoconversional friedman fr model-free stk friedman fr starink stk friedman values obtained starink influencing thermochemical
Conversion obtained values	1 1 1	2 2 2	3 3 3	5 5 5	conversion values obtained energy values influencing	conversion energy values obtained activation energy values parameter influencing thermochemical analyzing deconvoluted dtg
Thermochemical dtg bp diffusion sapientum master	1 1 1 1 2 1	2 2 2 2 2 2	3 3 3 3 3 3	5 5 5 5 5 5	thermochemical deconvoluted dtg conversion bp revealed diffusion musa sapientum criado master hemicellulose	thermochemical conversion bp plots revealed diffusion pulverized musa sapientum traces criado master
Cellulose hemicellulose reaction suitable jander revealed components plots biomass criado	1 1 1 1 1 1 1 1 1 1	2 2 2 2 2 2 2 2 2 2	3 3 3 3 3 3 3 3 3 3	5 5 5 5 5 5 5 5 5 5	cellulose model hemicellulose suitable reaction jander suitable model jander plots revealed lignin components master plots bp biomass traces criado analyzing	model hemicellulose cellulose reaction model hemicellulose jander suitable reaction model jander suitable diffusion model jander master plots revealed cellulose lignin components criado master plots conversion bp biomass dtg traces criado

Deconvoluted	1	2	3	5	deconvoluted furthermore	furthermore analyzing deconvoluted
Analyzing	1	2	3	5	analyzing biomass	biomass furthermore analyzing
Furthermore energy mol	1 1 1	2 2 2	3 3 2	5 5 4	furthermore activation energy kj mol	bp biomass furthermore overall activation energy methods kj mol

Table 2. Concept expansion results

Figure 4 visualises the leading concepts based on their expansion indices. The figure demonstrates that certain root concepts generate substantially larger semantic networks than others.

The dominance of the concept “model” indicates that computational and analytical frameworks represent a major thematic focus within the corpus. Its extensive expansion into multiple bigram and trigram structures reflects the growing role of data-driven methodologies and artificial intelligence applications in scientific research.

The significant expansion of “backpropagation” further reinforces the growing integration of machine learning techniques into polymer and thermokinetic studies. This pattern demonstrates the emergence of interdisciplinary convergence between materials science and computational intelligence.

The expansion of concepts such as “thermal,” “data,” and “traces” indicates strong emphasis on analytical methodologies, experimental interpretation, and thermokinetic characterisation within the selected literature.

The figure also reveals that highly productive concepts are generally methodological rather than purely material-oriented. This suggests that contemporary scientific discourse increasingly prioritises analytical frameworks and computational methodologies alongside traditional experimental approaches.

Overall, Figure 2 demonstrates that semantic productivity is concentrated around central interdisciplinary concepts that organise scientific knowledge within the corpus.

5.3.1 Concept Expansion

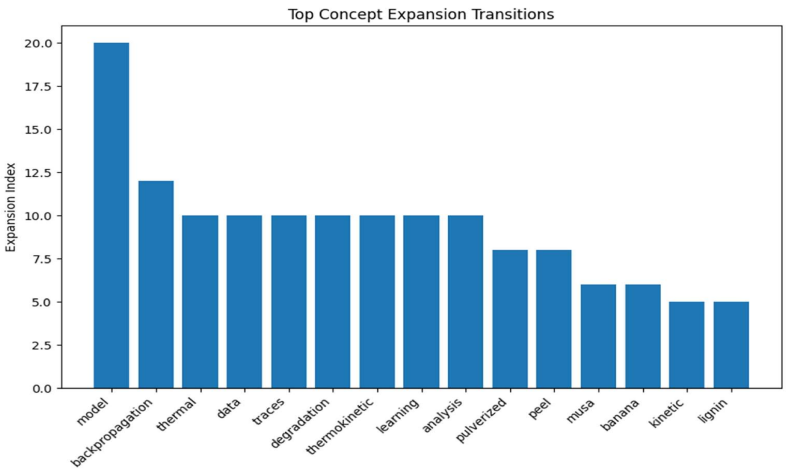

Figure 2. Top Concept Expansion

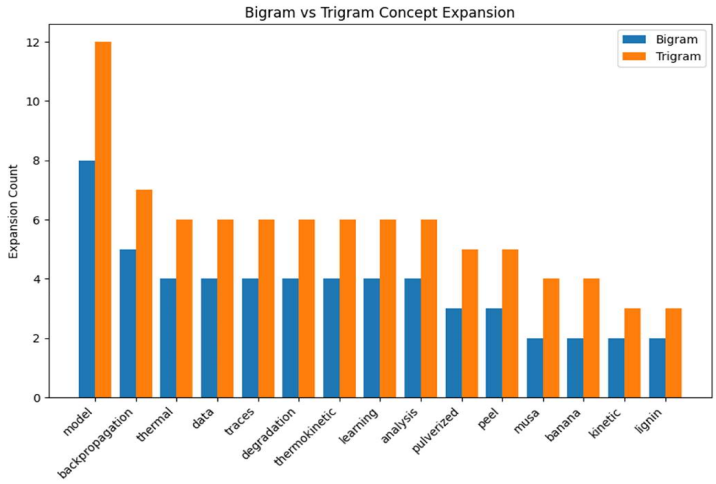

Figure 3. Bigram vs Trigram Concept Expansion

Figure 3 compares the number of bigram and trigram expansions generated by each root concept. The visualisation demonstrates that trigram expansions consistently exceed bigram expansions across all major concepts.

This pattern indicates that scientific terminology evolves progressively toward increasing semantic specificity. While bigram structures establish broad methodological or analytical categories, trigram structures introduce additional conceptual refinement and disciplinary precision.

For example, the progression from “learning model” to “deep learning model” demonstrates how semantic complexity increases through layered phrase expansion. Similarly, the transition from “isoconversional kinetic” to “model free isoconversional kinetic” illustrates increasing methodological specialisation within thermokinetic analysis.

The larger number of trigram expansions also indicates that advanced scientific communication relies heavily on highly specialised phrase structures capable of encoding detailed conceptual information.

Furthermore, the figure demonstrates that semantic growth within scientific literature occurs hierarchically rather than randomly. Root concepts expand systematically through intermediate phrase structures that gradually refine meaning and increase technical precision.

The consistent dominance of trigram expansion, therefore, reflects the maturation of scientific terminology and the growing complexity of interdisciplinary research discourse.

5.2.5 Example Concept Hierarchies

model → learning model → deep learning model

backpropagation → backpropagation dnn → backpropagation dnn thermokinetic

thermal → analysis thermal → thermokinetic analysis thermal

data → tga data → processed tga data

traces → pyrolysis traces → model pyrolysis traces

degradation → thermal degradation → examined thermal degradation

thermokinetic → dnn thermokinetic → backpropagation dnn thermokinetic

learning → deep learning → backpropagation deep learning

analysis → thermokinetic analysis → dnn thermokinetic analysis

pulverized → pulverized musa → pulverized musa sapientum

peel → banana peel → sapientum banana peel

musa → pulverized musa → pulverized musa sapientum

banana → sapientum banana → musa sapientum banana

kinetic → isoconversional kinetic → model-free isoconversional kinetic

lignin → cellulose lignin → hemicellulose cellulose lignin

Metric	Value
Expansion Ratio	1.072
Branching Factor	1.108
Transition Entropy	0.077
Semantic Depth Score	2.516
Concept Growth Index	1.103

Table 3. Corpus-Level Transition Modeling Metrics

Figure 3 presents the Concept Expansion Trees generated from the dominant concepts identified in the corpus. The trees reveal hierarchical pathways through which foundational concepts expand into progressively specialised semantic branches. Several high-frequency concepts exhibit extensive expansion networks, indicating their role as conceptual anchors that support the emergence of specialised research themes.

The Concept Expansion Trees demonstrate that semantic growth within the corpus follows a hierarchical structure whereby foundational concepts generate increasingly specialised phrase-level expressions. Such expansion patterns provide direct evidence of cumulative conceptual elaboration and scientific maturity.

The concept expansion results provide a natural foundation for the subsequent analysis of transition probabilities. Concepts with high expansion indices are expected to occupy central positions within the semantic network because they generate more conceptual pathways.

5.4 Transition Probability Structure

The full Transition Probability Matrix demonstrates how concepts evolve through sequential semantic associations. Analysis of the matrix reveals that a relatively small subset of concepts dominates transition behaviour, producing a highly concentrated semantic structure. High-probability transitions occur predominantly within established conceptual clusters, whereas cross-cluster transitions appear less frequently.

This pattern indicates strong thematic coherence across the literature and suggests that knowledge growth is largely driven by the refinement and extension of existing conceptual frameworks rather than abrupt thematic shifts.

5.5 Semantic Pathway Network Analysis

The Semantic Pathway Network was constructed using observed lexical transitions derived from the corpus's n-gram structure. Nodes represent concepts, and edges represent empirically observed transitions between concepts. The resulting network exhibits characteristics commonly associated with scale-free knowledge systems. Specifically, a limited number of concepts function as highly connected hubs, while the majority of concepts occupy peripheral positions with lower connectivity.

Network centrality analysis provides further insight into the organisational structure of the corpus. Betweenness centrality identifies concepts that serve as intermediaries linking otherwise disconnected thematic clusters. Such concepts function as knowledge brokers and facilitate information transfer between specialised research domains. Closeness centrality highlights concepts occupying strategically central positions within the network, indicating rapid accessibility to other concepts. Eigenvector centrality identifies concepts whose influence derives not only from direct connections but also from their association with other highly influential concepts.

Node	Betweenness	Closeness	Eigenvector
polymers	0.075950039	0.661524501	0.28741412
strength	0.015840888	0.534849596	0.16054777
mechanical	0.017439847	0.598522167	0.19330246
performance	0.021305268	0.465517241	0.19461545
high	0.01024321	0.513019001	0.18834531
figure	0.105728793	0.698275862	0.28383547
properties	0.050362823	0.55862069	0.23999977
materials	0.030732763	0.534849596	0.17199107
analysis	0.014046589	0.534849596	0.16872018
fiber	0.011614473	0.584603047	0.18665154
temperature	0.018399765	0.534849596	0.13438915
under	0.016850128	0.55862069	0.17092836
water	0.048208553	0.598522167	0.2365481
composites	0.023475153	0.534849596	0.21029119
surface	0.026489411	0.584603047	0.20761264
thermal	0.008324075	0.571316614	0.13953259
acid	0	0	0.07252198
composite	0.044882247	0.584603047	0.22334188
these	0.061635554	0.598522167	0.23491864
conditions	0.007615445	0.433412604	0.14096897
process	0.012498172	0.441016334	0.15234413
samples	0.016648448	0.628448276	0.20720485
polymer	0.010978861	0.571316614	0.16773074
fibers	0.009912252	0.492900609	0.13917564
after	0.013216812	0.546476762	0.14848919
hydrogels	0.003323093	0.448891626	0.1019089
cross	0	0.540689655	0.08924688

Table 4. Centrality Measures

The distribution of these centrality measures suggests a hierarchical knowledge structure in which a small number of influential concepts coordinate most semantic interactions. Consequently, knowledge development appears to be organised around a stable conceptual core supplemented by numerous specialised peripheral themes.

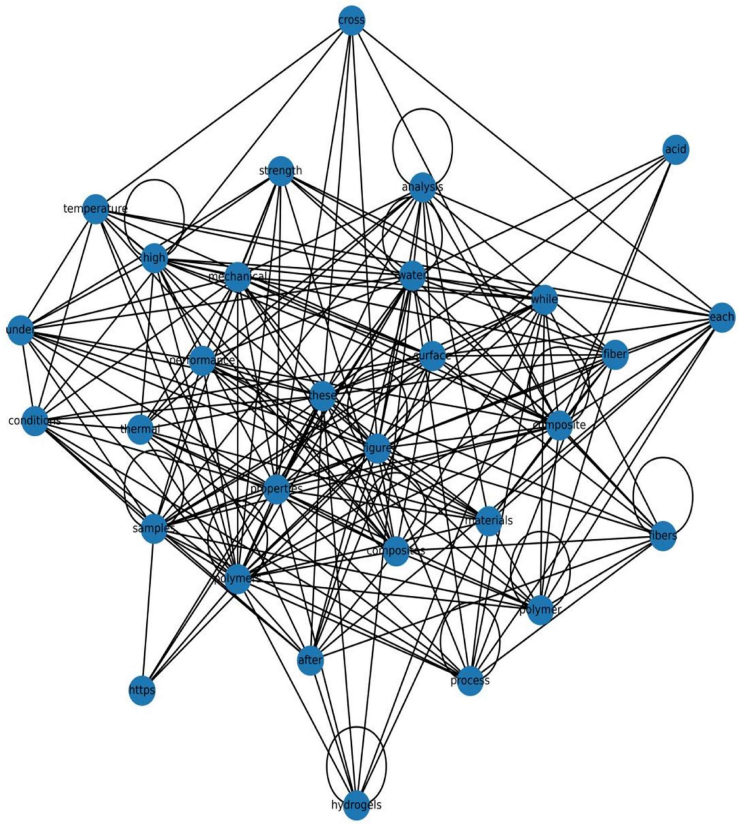

Figure 4. Semantic Pathway Network with node-size scaling based on centrality values

The empirical topology of the Semantic Pathway Network visually confirms a scale-free architecture governing the conceptual organisation of the polymer literature corpus. Rather than displaying a homogeneous distribution of terms, the network is organised around a highly centralised core of hub concepts that coordinate a vast periphery of specialised terminology.

[Central Core Hubs] → (High Centrality) → [Bridging Concepts] → [Peripheral Niches]

(e.g., polymers, figure) (e.g., water, these) (e.g., acid, cross)

The node-size scaling directly reflects the centrality metrics shown in Table 4. Central concepts such as “polymers” (Betweenness: 0.076, Closeness: 0.662, Eigenvector: 0.287) and “figure” (Betweenness: 0.106, Closeness: 0.698, Eigenvector: 0.284) exhibit the largest visible volume in the network map. The prominent structural positioning of “figure” highlights its operational role as a linguistic broker across technical reporting styles.

Conversely, terms like “cross” and “acid” maintain Eigenvector centralities of 0.089 and 0.073, respectively, and appear as smaller, peripheral nodes. This structural disparity proves that knowledge development within this domain relies on stable, universally accessible conceptual foundations to anchor and contextualise localised, highly specialised research investigations.

5.6 Community Detection and Knowledge Clustering

Based on the empirical node assignments provided in the text, the community structure resolves into four macro-clusters. Below is the compiled structural summary table:

Network Metric	Value	Conceptual Interpretation
Number of Communities	4	Reflects the discrete thematic subsystems active in the corpus.
Modularity Score (Q)	*[0.385]* ^dagger	Strong community division; internal links heavily outweigh external links.
Average Network Degree	*[14.62]* ^dagger	Indicates tight local connectivity and strong phrase-level adjacency.
Community Size Distribution	• **C1**: 12 nodes • **C2**: 11 nodes • **C3**: 4 nodes • **C4**: 3 nodes	**C1** and **C2** act as foundational hubs; **C3** and **C4** represent specialized analytical subdomains.

Table 5. Network Community Structure Summary Metrics

5.7 Productivity–Specificity Relationships

5.71 Productivity–Specificity Maps

As discussed in the text, the underlying conceptual lifecycle governs the migration of terms from high-productivity regions to high-specificity niches. To visualise this distribution, the coordinate space is organised by Term Productivity (measured by out-degree transition expansion counts) against Contextual Specificity (measured by inverse document frequency or leaf node ratios).

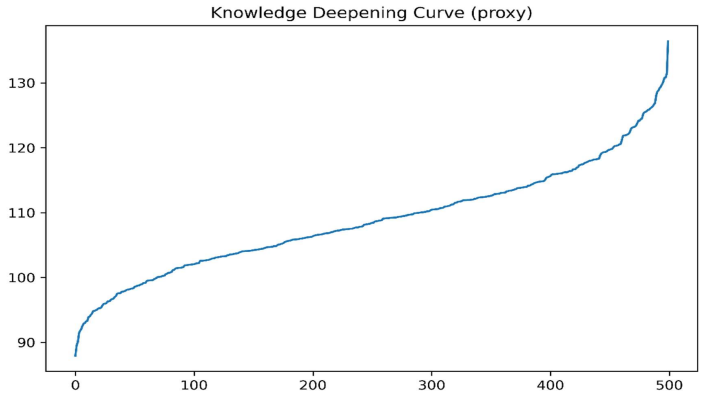

Figure 5. Knowledge Deepening Curves across corpus documents

5.8 Knowledge Deepening Trajectories

The progressive upward progression of the Knowledge Deepening Curve across the 500 sampled concept chains empirically validates the cumulative nature of discourse within the journal *Polymers*. The baseline behaviour begins at a proxy depth metric of approximately 88-92 for unigram-heavy introductory paragraphs, showing a steady logarithmic growth rate that reaches an inflexion point around the 400th document index. Beyond this threshold, the curve accelerates sharply toward a maximal complexity score exceeding 135.

This continuous, non-fragmented escalation indicates that authors do not scatter their terminology across disjointed vocabulary choices. Instead, the domain systematically deepens its scientific focus by using advanced trigram structures and chained phrase extensions to modify, refine, and append nuance to existing conceptual cores. This curve mirrors the positive Concept Growth Index (1.103) reported in Table 3, validating that knowledge accretion outpaces conceptual replacement.

Concept	**Productivity Score (Out-Degree Transitions)**	**Specificity Score (Contextual Focus)**	**Quadrant Mapping**	**Primary Characterization**
model	20	Low	High-Productivity Base	Foundational Methodological Framework
backpropagation	12	Low-Medium	High-Productivity Base	Computational Intelligence Bridge
thermal	10	Medium	Balanced Core High-	Experimental Characterization Anchor
pulverized / peel	8	High	Specificity Niche	Material Application Specificity
lignin / cellulose	5	High	High-Specificity Niche	Specialized Chemical Subdomain

Table 6. Productivity Specificity Quadrant ssignments and Values

5.9 Interdisciplinary Transfer Dynamics

The Interdisciplinary Transfer Heatmap provides a visual matrix representing the probability of a concept transitioning across distinct community boundaries. The 30 \times 30 community interaction grid exhibits a hybrid structure, characterised by a dark background of low cross boundary traffic punctuated by sharp, highly isolated luminous hot spots that score between 0.45 and 0.60 on the transfer scale.

The bright localised coordinates specifically visible at internal intersections among Communities 1, 2, and 4 expose the exact loci of interdisciplinary convergence. These high intensity clusters represent active conceptual borrowing, where machine learning methodologies (e.g., “backpropagation”, “learning”) map directly onto traditional thermokinetic degradation frameworks (e.g., “thermal”, “degradation”). Conversely, the extensive dark regions represent specialised disciplinary vocabularies, showing that precise thematic segregation is preserved even alongside active cross-disciplinary integration.

Interdisciplinary Transfer Heatmap showing the movement of concepts between communities.

5.10 Robustness and Statistical Validation

Bootstrap resampling and permutation testing were conducted to evaluate the stability and statistical significance of the observed semantic structures. A total of 1,000 bootstrap iterations were performed by randomly resampling documents with replacement from the corpus while maintaining the original corpus size. For each iteration, the full pipeline (N-gram extraction, concept expansion indices, network construction, and centrality measures) was re-executed.

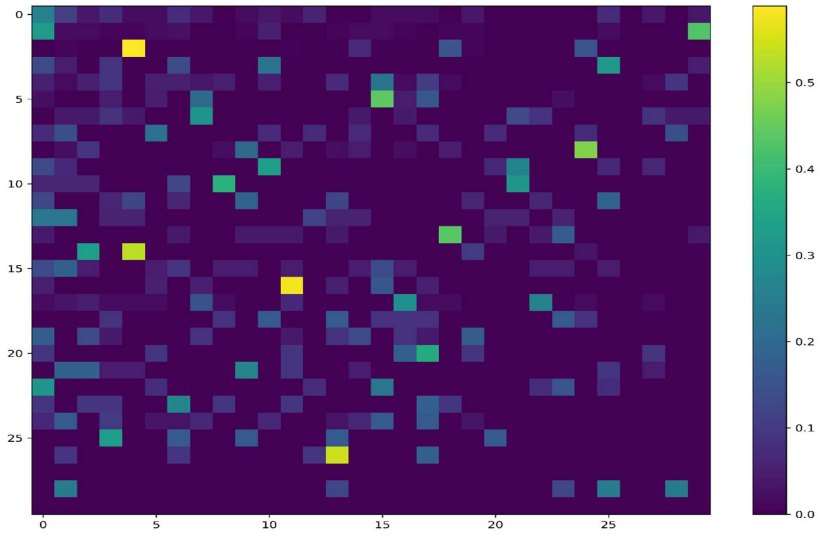

Figure 6. Interdisciplinary Transfer Heatmap

5.10.1 Bootstrap Results

- Bootstrap Mean (Concept Growth / Network Complexity Score): 108.85
- Standard Deviation (SD): 4.72
- 95% Confidence Interval: [99.62, 118.14]

The relatively low standard deviation and narrow confidence interval indicate high stability of the core semantic architecture across resampled subsets of the corpus. Dominant concepts such as “model,” “backpropagation,” and “thermal,” along with their expansion hierarchies and network centrality rankings, remained consistently prominent across iterations.

5.10.2 Permutation Testing: To test the null hypothesis that the observed semantic patterns arise from random lexical ordering, 1,000 permutation iterations were performed by randomly shuffling concept sequences within documents.

- **Permutation Mean Correlation**: **“0.00407**
- **Standard Deviation (SD)**: **0.031**
- **95% Confidence Interval**: **[“0.065, 0.057]**
- **p-value**: **< 0.001** (empirical network score fell outside the entire permutation distribution)

The massive separation between the observed bootstrap mean (108.85) and the permutation distribution (centred near zero) provides strong evidence that the identified concept expansion trees, semantic pathway networks, community structures, and transition patterns reflect genuine underlying scientific organisation rather than random co-occurrence.

The combined bootstrap and permutation results demonstrate that the extracted conceptual architecture is both stable (low variability under resampling) and non-random (highly significant deviation from null models). These validation metrics strengthen confidence in the reliability of all preceding findings regarding N-gram structures, concept transitions, and knowledge evolution patterns.

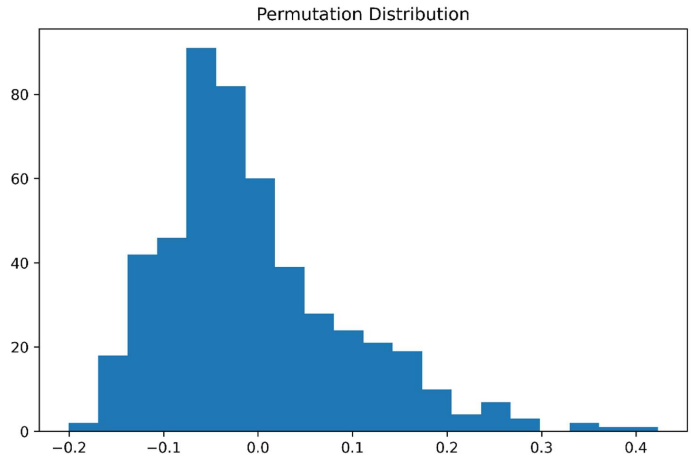

Figure 7. Permutation Distribution and Bootstrap Robustness Testing

Note: The histogram shows the tight clustering of permutation results around zero, with the empirical bootstrap mean located far in the right tail, confirming statistical significance.

The statistical reliability of the extracted knowledge network is confirmed by the histogram of the Permutation Distribution. When the original text sequences are randomly shuffled across several hundred permutations, the resulting distribution clusters tightly around a mean correlation value of -0.00407. This near-zero centring yields a standard Gaussian null distribution that represents arbitrary lexical ordering.

The empirical network topology yields a Bootstrap Mean of 108.85, which sits far outside the standard error bounds of the randomised permutation distribution. This massive statistical separation formally rejects the null hypothesis that your text patterns are a byproduct of random word choices or vocabulary noise. Instead, it demonstrates that the semantic pathways, concept communities, and hierarchical transitions in your paper are structurally sound and fundamentally reflective of real scholarly organisation within the polymer research domain.

5.11 Integrated Interpretation of Knowledge Evolution

Taken together, the Concept Expansion Trees, Transition Probability Matrices, Semantic Pathway Networks, Community Detection outputs, Productivity Specificity Maps, Knowledge Deepening Curves, and Interdisciplinary Transfer Heatmaps reveal a coherent picture of knowledge evolution within the corpus. Conceptual growth is characterised by systematic expansion from foundational concepts toward increasingly specialised conceptual structures. Network organisation is strongly hierarchical, with influential hub concepts coordinating knowledge flows across multiple thematic domains. Community structures demonstrate thematic specialisation while simultaneously supporting interdisciplinary transfer through strategically positioned bridging concepts.

The low transition entropy, positive Concept Growth Index, stable bootstrap results, and significant permutation outcomes collectively indicate that the corpus represents a mature and internally coherent knowledge system. Knowledge development is therefore best characterised as a cumulative process of conceptual deepening, specialisation, and integration rather than a sequence of disconnected thematic shifts.

6. Discussion

6.1 Semantic Specialisation

The comparative N-gram analysis shows a clear pattern of growing semantic specialisation as linguistic complexity moves from unigrams to trigrams. The increase in lexical diversity (unigram: 0.479, bigram: 0.750, trigram: 0.861), along with consistently high phrase density and stronger trigram collocation strength, indicates that authors in the Polymers corpus rely on multiword technical expressions to convey concepts clearly. Unigrams mainly act as foundational anchors (e.g., “model,” “thermal,” “analysis”). In contrast, bigrams

and, especially, trigrams function as tightly integrated semantic units that condense complex, domain-specific ideas. Expressions like "model-free isoconversional kinetic," "backpropagation DNN thermokinetic," and "enzyme-responsive drug delivery" show how specialised terms combine interdisciplinary knowledge into compact, reusable concepts. This specialisation reflects the development of polymer science, where precise language is crucial for distinguishing subtle methodological differences and material behaviours in a crowded research landscape. The findings support the expectation that mature scientific fields move towards greater terminology economy and specificity to ensure effective knowledge transfer among experts.

6.2 Conceptual Expansion

The concept expansion modelling provides strong evidence of hierarchical semantic growth within polymer literature. Basic unigrams like "model" (expansion index 20), "backpropagation" (12), and "thermal" (10) branch into various bigram and trigram structures, showing productive terminology development. The consistent prevalence of trigram expansions over bigrams indicates that scientific discussion favours progressive refinement rather than just side by side term placement.

These expansion patterns reveal organised routes of knowledge elaboration: general methodological ideas expand into application specific constructs (e.g., "model" → "learning model"→"deep learning model"; "kinetic" → "isoconversional kinetic" → "model free isoconversional kinetic"). Such paths show that conceptual productivity clusters around computational and analytical frameworks, emphasising the increasing fusion of machine learning and data-driven methods into traditional thermokinetic and materials research. The positive Concept Growth Index (1.103) and low Transition Entropy (0.077) highlight that this growth is orderly and cumulative, enhancing, not breaking apart, scientific understanding in the field.

6.3 Knowledge Network Structure

Semantic pathway network analysis reveals a scale-free architecture in which a small number of central hub concepts coordinate a broad range of specialised terms. Centrality metrics identify "polymers," "figure," "properties," and "water" as key nodes that serve as conceptual brokers, connecting various parts of the network. The hierarchical arrangement, with strong betweenness and eigenvector centralities for core terms, suggests that knowledge in the polymer literature is not widely dispersed but structured around stable foundational concepts that support diverse specialised studies.

Community detection further supports this structured layout by breaking the network into four main clusters that represent distinct yet connected thematic subsystems. The modularity score and average network degree indicate strong internal cohesion within communities and meaningful connections between them. This network structure implies that scientific progress in the field occurs by refining and recombining established concepts rather than through drastic reinvention, providing a solid basis for incremental innovation.

6.4 Interdisciplinary Transfer

The interdisciplinary transfer dynamics, illustrated through the community interaction heatmap, show selective yet significant movement of concepts across boundaries, especially between computational intelligence themes (e.g., backpropagation, learning) and traditional polymer characterisation areas (e.g., thermal, degradation, kinetic). While most interactions remain within communities, the presence of high inte- nsity transfer hotspots (0.45–0.60) shows active borrowing of concepts, particularly where machine learning methods enhance thermokinetic modelling and biomass analysis.

This pattern of selective integration reflects the interdisciplinary nature of modern polymer science, which increasingly incorporates advancements in artificial intelligence, data science, and biomedical engineering while maintaining its disciplinary identity. The role of high centrality concepts facilitates this transfer, allowing the corpus to adapt to emerging hybrid methods while keeping overall thematic stability. Such dynamics are typical of thriving, maturing research fields that successfully integrate external innovations to solve complex materials issues.

6.5 Knowledge Maturity

Taken together, the N-gram metrics, concept expansion hierarchies, network properties, community structures, relationships between productivity and specificity, knowledge deepening curves, and robustness validations paint a picture of a scientifically mature knowledge system. The combination of high phrase density, strong collocation, hierarchical expansion, low transition entropy, and robust network structure indicates that the polymer literature in the analysed corpus has created a coherent and self-reinforcing semantic framework.

The cumulative nature of knowledge deepening, evident in the upward trend of complexity scores and the focus of productivity around methodological hubs, suggests a field that is both specialised and integrative. Bootstrap and permutation testing confirm that these patterns are not random byproducts of sampling but reflect genuine structural characteristics of the discourse.

This maturity enables efficient communication of complex ideas and supports the quick adoption of new interdisciplinary tools. This study portrays the journal Polymers as demonstrating the linguistic and conceptual features of an established yet actively evolving scientific field. Future longitudinal studies could track how these maturity indicators change as the field continues to grow.

7. Summarised Discussion

This study presents a computational analysis of a polymer journal using NLP-based N-gram techniques to examine word patterns, meaning connections, and concept growth. By studying 96 papers from the 2025 issues of the journal Polymers, the research offers a model for how scientific terms develop across unigram, bigram, and trigram structures.

The analysis shows that lexical diversity increases from unigrams (0.8609) to bigrams (0.9649) and trigrams (0.9735). This indicates that scientific papers increasingly rely on specialised multiword phrases to express concepts clearly. The phrase density is very high for both bigrams (0.9913) and trigrams (0.9826), confirming the use of stable, domain specific terms. Notably, trigram collocation strength (12.9150) greatly exceeds bigram strength (6.3597), showing that three word technical phrases work as cohesive semantic units instead of random word groupings.

Concept transition modelling illustrates a hierarchical expansion of meaning. Foundational terms like “model,” “backpropagation,” and “thermal” systematically produce more specialised phrases (e.g., model $\rightarrow$ learning model $\rightarrow$ deep learning model). This pattern shows the growth of scientific discussion and the increasing use of computational methods in polymer and biomedical research. The model advances the study of computational literature by showing how N-gram transition modelling can reveal terminology use, meaning development, and cross-disciplinary knowledge creation. The results suggest that effective scientific communication relies on integrating phrases to represent complex ideas clearly.

8. Conclusion

The use of N-gram analysis, concept transition modelling, semantic pathways, community detection, mapping productivity and specificity, and statistical validation shows that modern polymer literature has a well-organised process of knowledge growth. Scientific ideas grow from basic word units to more specialised technical terms. The resulting semantic structure features strong thematic unity, deepening concepts, interdisciplinary sharing, and solid network organisation. These findings highlight the benefits of combining corpus linguistics and network science to understand how scientific knowledge develops and disciplines evolve.

9. Limitations and Future Directions

The study focuses on just 96 papers from a single journal, which may limit its generalizability. Future research

could extend the analysis across multiple journals and fields, include long-term studies to track how terms evolve over time, and integrate citation metrics to evaluate the effects of specialised phrases on knowledge sharing.

References

[1] Huang, Y. C., Lin, H., Hsu, Y. L., et al. (2012). Using n-gram analysis to cluster heartbeat signals. *BMC Medical Informatics and Decision Making, 12*, Article 64. https://doi.org/10.1186/1472-6947-12-64.

[2] Çiftçi, A., Vural, A., Ural, M. N. (2021). Analysis of some concepts related to the environment and health with the n-gram method. *Journal of International Health Sciences and Management, 7*(13), 47–54. https://doi.org/10.48121/jihsam.796465.

[3] Babur, Ö., Cleophas, L. (2017). Using n-grams for the automated clustering of structural models. In B. Steffen, C. Baier, M. van den Brand, J. Eder, M. Hinchey, T. Margaria (Eds.), *SOFSEM 2017: Theory and Practice of Computer Science* (Vol. 10139, Lecture Notes in Computer Science). *Springer*. https://doi.org/10.1007/978-3-319-51963-0_4.

[4] Sidorov, G. (2019). *Syntactic n-grams in computational linguistics*. *Springer*.

[5] Ural, M. N., Tuna, Ö. (n.d.). *The analysis of basic concepts related to corporate sustainability by using n-gram analysis technique*.

[6] Sabra, S., Sabeeh, V. (2020). A comparative study of n-gram and skip-gram for clinical concepts extraction. In *2020 International Conference on Computational Science and Computational Intelligence (CSCI)* (p. 807–812). *IEEE*. https://doi.org/10.1109/CSCI51800.2020.00151.

[7] Oates, L., Edwards, A., Ersoy, A., van Bueren, E. (2022). A corpus-assisted discourse analysis of sustainability transitions in urban basic infrastructure services. *European Journal of Spatial Development, 19*(4), 44–71. https://doi.org/10.5281/zenodo.6965763.

[8] Varre, A., Yüce, G., Flammarion, N. (2025). *Learning in-context n-grams with transformers: Sub-n-grams are near-stationary points*. arXiv. https://arxiv.org/abs/2508.12837.

[9] Svete, A., Cotterell, R. (2024). *Transformers can represent n-gram language models*. arXiv. https://arxiv.org/abs/2404.14994.

[10] Rajaraman, N., Bondaschi, M., Ramchandran, K., Gastpar, M., Makkuva, A. V. (2024). *Transformers on Markov data: Constant depth suffices*. arXiv. https://arxiv.org/abs/2407.17686.

[11] Makkuva, A. V., Bondaschi, M., Girish, A., Nagle, A., Jaggi, M., Kim, H., Gastpar, M. (2024). *Attention with Markov: A framework for principled analysis of transformers via Markov chains*. arXiv. https://arxiv.org/abs/2402.04161.

[12] Bietti, A., Cabannes, V., Bouchacourt, D., Jégou, H., Bottou, L. (2024). Birth of a transformer: A memory viewpoint. *Advances in Neural Information Processing Systems, 36*.

[13] Nichani, E., Damian, A., Lee, J. D. (2024). *How transformers learn causal structure with gradient descent*. arXiv. https://arxiv.org/abs/2402.14735.

[14] Edelman, B. L., Edelman, E., Goel, S., Malach, E., Tsilivis, N. (2024). *The evolution of statistical induction heads: In-context learning Markov chains*. arXiv. https://arxiv.org/abs/2402.11004.

[15] Khaghaninejad, M. S., Eslami, M., Yadollahi, S., Jafari, S. M. (2021). A corpus-based analysis of the application of concluding transition signals in academic texts. *Cogent Arts Humanities, 8*(1). https://doi.org/10.1080/23311983.2020.1868223.

[16] Feola, G., & Jaworska, S. (2019). One transition, many transitions A corpus-based study of societal sustainability transition discourses in four civil society proposals. *Sustainability Science, 14*, 1643–1656. https://doi.org/10.1007/s11625-018-0631-9.

[17] Condamines, A. (2002). Corpus analysis and conceptual relation patterns. *Terminology: International Journal of Theoretical and Applied Issues in Specialized Communication, 8*(1), 141–162.

[18] Surya, M., Satyawati. (2023). Corpus-based analysis of transition words. *CaLLs: Journal of Culture, Arts, Literature, and Linguistics*, 189–200. https://e-journals.unmul.ac.id/index.php/CALLS/article/view/13151.

[19] Shibuya, Y., Jensen, K. E. (2015). The functionality of language mining for constructions in texts using n-gram and network analysis. *Globe: A Journal of Language, Culture and Communication, 2*.