Wikipedia Knowledge Graphs & Science Gateways

Navigating the Nano-Future: Forecasting Emerging Concepts through Wikipedia Link Architectures

1. RESEARCH QUESTION

"What will be the emerging concept or technology in the field of Nanotechnology based on the structural evolution of the Wikipedia knowledge graph?"

2. ABSTRACT

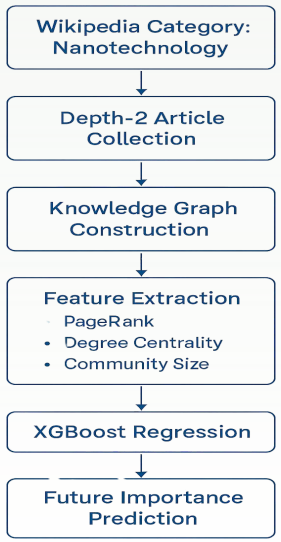
Identifying emerging scientific concepts is difficult in rapidly evolving fields such as nanotechnology. WinMaxxers addresses this challenge through a Science Gateway that forecasts concept growth using only Wikipedia data. Unlike black-box approaches, their method treats Wikipedia as a knowledge graph, where articles are nodes and hyperlinks represent relationships, providing transparent and explainable predictions.

Their dataset contains 1,530 nanotechnology-related Wikipedia articles collected through a depth-2 traversal of the domain's root category. Historical snapshots from 2020–2026 were used to extract graph features including PageRank, degree centrality, and community size. These features were used to train an XGBoost regression model that predicts future concept importance.

The gateway enables researchers, students, and educators to explore concept trends without managing large-scale data collection, graph processing, or machine learning infrastructure. The system was built using Python, Wikipedia APIs, NetworkX, Pandas, and XGBoost, creating a reproducible workflow for knowledge graph analysis.

Key lessons included the importance of narrowing the project scope to achieve meaningful results and implementing robust checkpointing and retry systems for large-scale API data collection. Early findings suggest that interpretable graph-based features can effectively identify concepts gaining momentum, demonstrating the potential of Science Gateways to make advanced forecasting tools more accessible to the scientific community.

3. METHODOLOGY



4. Dataset Statistics

Metric	Value
Articles	1,530
Year Analyzed	2020-2026
Prediction Target	Future Importance Score
Data Source	Wikipedia
Domain	Nanotechnology

5. GOALS

Structure-Only Prediction: Leverage internal link architecture as predictive markers.

Explainable AI Modeling: Shun black-box modeling; prioritize actionable, weighted structural features.

Temporal Dynamics: Target growth surges across historical 7-year network intervals.

6. SOLUTION ARCHITECTURE

Layer	Technologies Employed
Data	MediaWiki API, Wikidata
Backend	Python, Pandas, NetworkX, MatplotLib
ML Engine	XGBoost Regressor, Scikit-Learn
Hosting	GitHub Pages Portal

XGBoost Pandas NetworkX

7. OUTCOMES

- **Standard Clean Dataset:** Produced clean master list of 1,530 distinct nodes mapped across 7-year time intervals (2020–2026).
- **Model Performance:** Evaluated multi-tier models; selected XGBoost for performance.
- **Emergent Topics Page:** Developed a tab predicting PageRank trajectories of all articles analyzed

Authors

Damone Washington
WSSU - Rising Senior

Quintin Ampofo
HU - Rising Junior

Cole Scott
HU - Rising Sophomore

Darell Barnes
VU - Rising Sophomore