Large Knowledge Graphs for Construction, Manufacturing and Mining: A work in Progress

LitePaper in progress v.0, August 2025.

Santanu Banerjee, The SenseiAlgo Team.

The manufacturing industry has seen great revamps in recent times through the introduction IoT and AI integration. Autonomous dark factories are on the rise, not needing human intervention at all in many processes. From another end, breakthroughs in different AI products/frameworks/methodologies have made multiple tasks easier to automate. In this regard, Knowledge Graphs (KGs) have been playing an interesting role in different application area. The ability to store information accurately as Tripartite Graphs [1] have proved to be extremely useful especially in domains like biology and medicine, but also affecting interconnected industrial processes like construction, manufacturing, mining, etc.

Some important literature highlighting the evolution of knowledge graphs include discussions on how LLMs can be used for KG generation [2], triple-set prediction by LLMs [3], capabilities of LLMs for constructing KGs [4], usage of KG Embeddings in LLMs [5], information retrieval and reasoning [6], task-oriented KG reasoning [7]; additionally multiple other resources highlight structured reasoning with KGs [8], AI Agent usage to traverse through KGs to find innovative subgraphs for hypothesis generation [9] and more. Interestingly, not only can KGs be used for novel hypothesis generation by letting a swarm of AI Agents loose on the KG from nearby nodes with slightly different prompts/goals to develop starkly different sub-graphs, similar (backtracking/causal-analysis) approaches can be used in Root Cause Analysis (RCA); RCA could be useful not only in manufacturing or maintenance, but also in cybersecurity applications [10] [11] [12].

A recent paper in this regard focuses on utilizing KGs for smart manufacturing [13]. Apart from a whole lot of intriguing aspects which are discussed in the paper, the authors also focus on KG-reasoning mentioning that "*integrating structured ontologies with KGs enhances deductive reasoning, making implicit connections explicit and generating new insights to augment the existing knowledge base in" Smart Manufacturing*. They also discuss *"KG fusion at the conceptual layer … merging disparate ontologies and schemas from different KG"*s. The articles goes on to discuss applications of KGs in Engineering Design, Optimization and Scheduling, Quality Control, Supply Chain Management and Predictive Maintenance. In general KGs can be extremely useful for Risk Assessment, for usage alongside tools like (Design) Failure Modes and Effects Analysis, and allow novel frameworks to be developed on top of them for Safety Violation Assessment.

In SenseiAlgo, we understand the importance of being the first movers in using Large Knowledge Graphs (LKGs) for core industrial processes, especially to map and store complex indirect relationships in construction, mining and the manufacturing industries. In this regard we approach developing a LKG for Indian National Building Codes (INBC) 2016; the INBC can be found in different volume-wise PDFs (more than 1000 pages each). We use the GPT-4o model using OpenAI APIs to develop small KGs for each chunk of the PDF text (we use *text_splitter* from *langchain* to split the PDF files) and continue merging these chunks-wise KG; an intermediate LKG during this process is shown in Figure 1. It is observed that many clusters of good data connections gets formed, however additionally the LKG needs to be processed at regular intervals during its formation, to continually remove bad nodes and edges which somehow also clustered together among themselves. We will keep the community posted of our findings and look forward to novel methodology releases from academic labs focusing on KG post-processing.

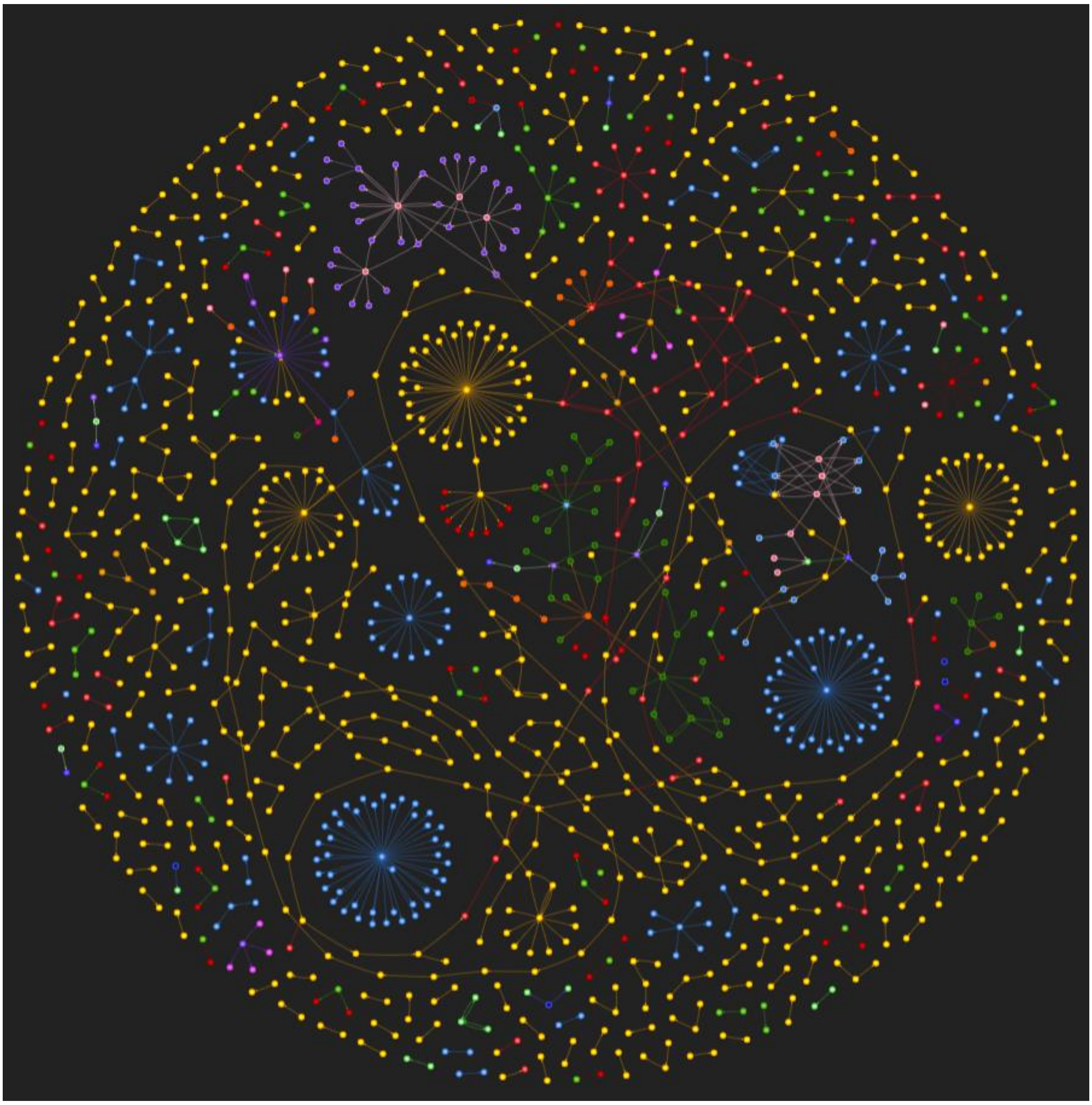
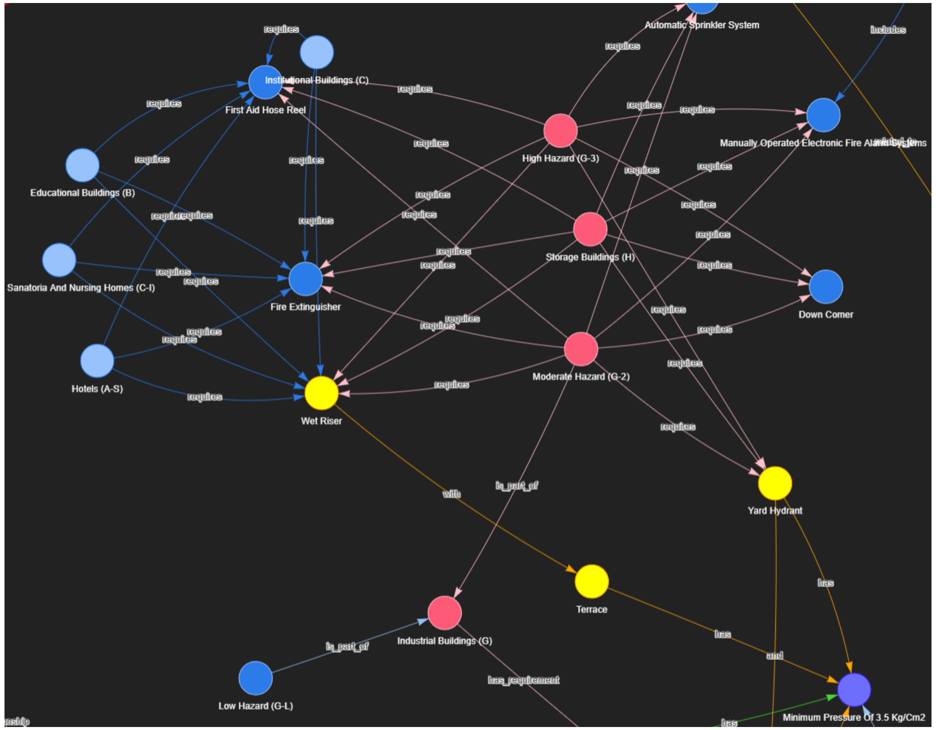
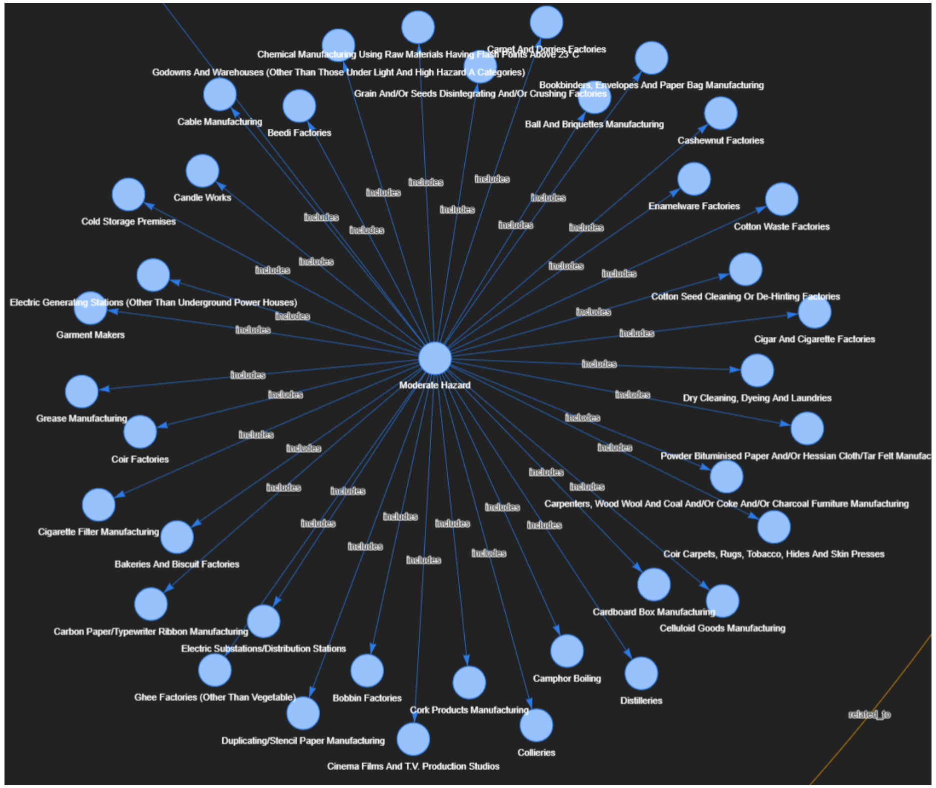


Figure 1: An evolving LKG (under-development) from INBC 2016 (Volumes 1 and 2) followed by two zoomed in views of different portions of this LKG

These images are hyperlinked to view the entire graph (HTML) in any browser window (warning: it takes a long time to load, but is worth the patience)


Automatic Sprinkler System
Institutional Buildings (C)
First Aid Hose Reel
Manually Operated Electronic Fire Alarm Systems
High Hazard (G-3)
Educational Buildings (B)
Storage Buildings (H)
Sanatoria And Nursing Homes (C-I)
Fire Extinguisher
Down Comer
Hotels (A-S)
Moderate Hazard (G-2)
Wet Riser
Yard Hydrant
Terrace
Industrial Buildings (G)
Low Hazard (G-L)
Minimum Pressure Of 3.5 Kg/Cm2

Chemical Manufacturing Using Raw Materials Having Flash Points Above 23°C
Carpet And Durries Factories
Godowns And Warehouses (Other Than Those Under Light And High Hazard A Categories)
Bookbinders, Envelopes And Paper Bag Manufacturing
Grain And/Or Seeds Disintegrating And/Or Crushing Factories
Cable Manufacturing
Beedi Factories
Ball And Briquettes Manufacturing
Cashewnut Factories
Candle Works
Cold Storage Premises
Enamelware Factories
Cotton Waste Factories
Electric Generating Stations (Other Than Underground Power Houses)
Cotton Seed Cleaning Or De-Hinting Factories
Garment Makers
Cigar And Cigarette Factories
Moderate Hazard
Dry Cleaning, Dyeing And Laundries
Grease Manufacturing
Coir Factories
Carpenters, Wood Wool And Coal And/Or Coke And/Or Charcoal Furniture Manufacturing
Cigarette Filter Manufacturing
Coir Carpets, Rugs, Tobacco, Hides And Skin Presses
Bakeries And Biscuit Factories
Cardboard Box Manufacturing
Celluloid Goods Manufacturing
Carbon Paper/Typewriter Ribbon Manufacturing
Electric Substations/Distribution Stations
Camphor Boiling
Ghee Factories (Other Than Vegetable)
Bobbin Factories
Cork Products Manufacturing
Distilleries
Duplicating/Stencil Paper Manufacturing
Collieries
Cinema Films And T.V. Production Studios

References

[1] M. Banf and J. Kuhn, *A Tripartite Perspective on GraphRAG,* 2025.

[2] Y. Yao, X. Wang, Z. Zhang, Y. Qin, Z. Zhang, X. Chu, Y. Yang, W. Zhu and H. Mei, *Exploring the Potential of Large Language Models in Graph Generation,* 2024.

[3] Y. Yuan, Y. Xu and W. Zhang, *Is Large Language Model Good at Triple Set Prediction? An Empirical Study,* 2024.

[4] Y. Zhu, X. Wang, J. Chen, S. Qiao, Y. Ou, Y. Yao, S. Deng, H. Chen and N. Zhang, *LLMs for Knowledge Graph Construction and Reasoning: Recent Capabilities and Future Opportunities,* 2024.

[5] J. Barmettler, A. Bernstein and L. Rossetto, *ConceptFormer: Towards Efficient Use of Knowledge-Graph Embeddings in Large Language Models,* 2025.

[6] Y. Gao, Y. Xiong, Y. Zhong, Y. Bi, M. Xue and H. Wang, *Synergizing RAG and Reasoning: A Systematic Review,* 2025.

[7] G. Niu, B. Li and Y. Lin, *A Survey of Task-Oriented Knowledge Graph Reasoning: Status, Applications, and Prospects,* 2025.

[8] *ADD LLM TO Knowledge-Graph: NEW GIVE Method (Berkeley),* 2025.

[9] *From Knowledge to Hypotheses: How AI is Changing the Way Science Begins,* 2025.

[10] Z. Yan and J. Liu, "A Review on Application of Knowledge Graph in Cybersecurity," in *2020 International Signal Processing, Communications and Engineering Management Conference (ISPCEM)*, Montreal, Quebec, Canada, 2020.

[11] L. F. Sikos, "Cybersecurity knowledge graphs," *Knowledge and Information Systems,* vol. 65, no. 9, pp. 3511-3531, 2023.

[12] Y. Ma, Y. Chen, Y. Wang, J. Yu, Y. Li, J. Lu and Y. Wang, "The Advancement of Knowledge Graphs in Cybersecurity: A Comprehensive Overview," in *Computational and Experimental Simulations in Engineering. ICCES 2023. Mechanisms and Machine Science*, Cham, 2024.

[13] Y. Wan, Y. Liu, Z. Chen, C. Chen, X. Li, F. Hu and M. Packianather, "Making knowledge graphs work for smart manufacturing: Research topics, applications and prospects," *Journal of Manufacturing Systems,* vol. 76, pp. 103-132, 2024.