Guest editor

Michael Eiden

Contributing authors

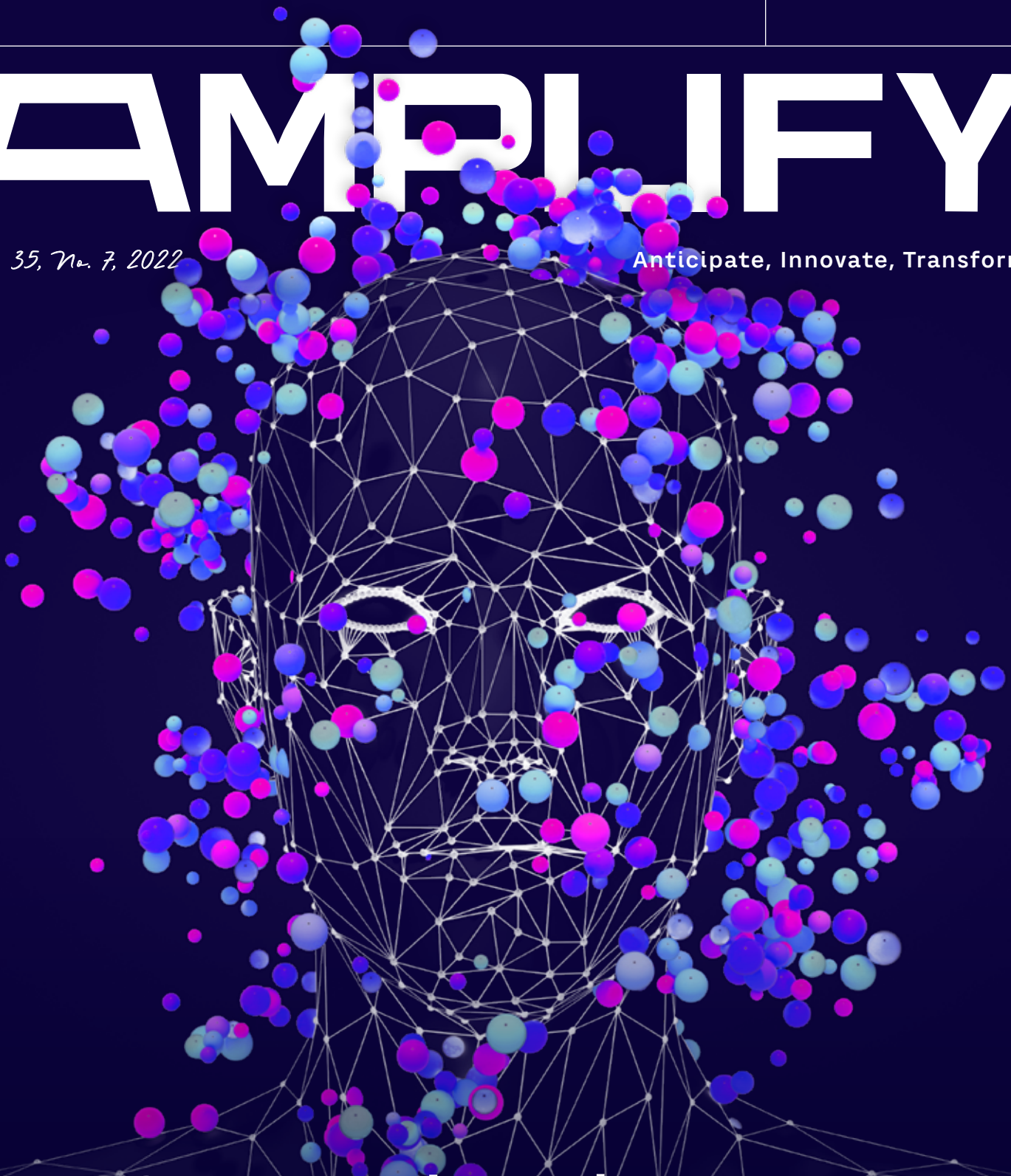
Cigdem Z. Gurgur
George F. Hurlburt
Phillipe Monnot
Lila Rajabion
Armand Rotaru
Andy E. Williams

CUTTER
AN ARTHUR D. LITTLE COMMUNITY

AMPLIFY

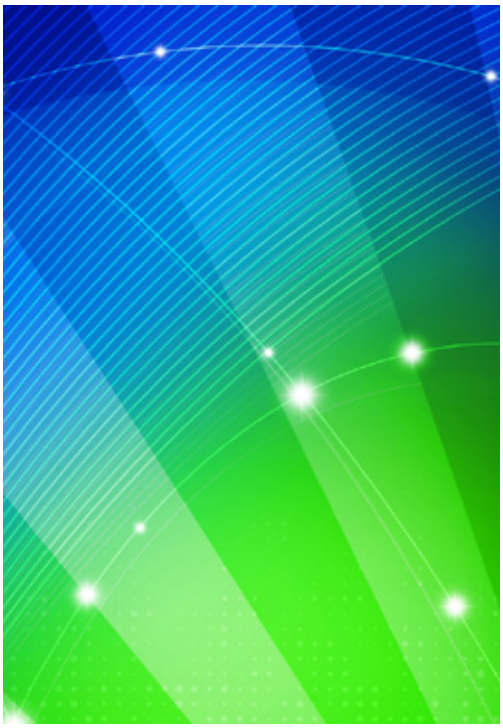
Vol. 35, No. 1, 2022

Anticipate, Innovate, Transform



Connecting the Dots with Knowledge Graphs

CONTENT



OPENING STATEMENT

BY MICHAEL EIDEN, GUEST EDITOR

The increasing realization that deep learning alone cannot be the solution to build robust, reliable artificial intelligence (AI) systems, coupled with the ever-increasing need to make use of heterogeneous data sources for decision making, has led to a recent resurgence of knowledge graphs (KGs). KGs are essentially graph-based representations of information that consist of three simple elements: nodes (which represent entities), edges (which encode a relationship between entities), and attributes that describe the relationships and entities. With this simple recipe, we can model any real-world problem as accurately as possible and thus encode domain knowledge into a system that is transparent for humans but can also be interpreted by computers.

KGs have been around for a while (research on them began in the 1980s, and Google announced it was using them in 2012), but they have often been solely used for knowledge representation. Today, even small companies have an amazing amount of data (often heterogeneous), and KGs are the perfect tool to leverage that data. Additionally, various technology platforms and open source tools now exist that make it much easier to design, build, and deploy KGs.

Use cases for KGs vary in range and cross many industries. Their most prominent applications are in product or content recommendation systems, but they have been successfully used in drug discovery research, for the estimation of passenger flows in transport hubs, and in the optimization of global supply chains. These are just a few examples; several others are included in the first article of this issue, authored by Lila Rajabion. Business leaders are discovering that KGs can provide meaningful insights into internal data, empower employees by serving up the right information at exactly the right time, and help managers and others make better decisions.

However, the most exciting KG area relates to AI. As discussed in the May 2021 issue of *Amplify* (and by Cigdem Gurgur in this issue), the lack of explainability (especially in deep learning systems) is a major challenge for more widespread adoption. In the May 2021 issue, Cutter Expert Claude Baudoin and Clayton Pummill told us:

> AI is mysterious. The vast majority of society does not understand how it works, and deep neural networks in particular can produce results that we cannot readily explain. People generally fear what they don't understand.[1]

KGs are now playing a seminal role in the emergent field of neuro-symbolic AI, which aims to integrate domain knowledge into AI systems. By combining AI's statistical/machine learning (ML) side with KGs, we get more effective, more explainable cognitive results and begin creating logic-based systems that get better with each application.[2] In other words, we can build the next generation of AI models, ones that support better human-machine collaboration, an idea taken to its very edge by Andy Williams in this issue with his article on general collective intelligence (GCI) and Industry 5.0.

IN THIS ISSUE

Our first article looks at a number of use cases for KGs, both general and specific. Rajabion provides four examples of how KGs can help leaders advance their understanding of the business environment in which their company sits. These include merging data silos to create a company overview across divisions, connecting different types of data in meaningful ways, aiding informed decision making by narrowing searches and contextualizing information, and showing interconnections that help leaders gain perspective. Next, Rajabion dives into how Google, LinkedIn, eBay, and IBM are using KGs and explains how other companies could follow suit. She then addresses four challenges currently faced by companies looking to leverage KGs, followed by a look at specific business efficiencies enabled by KGs, including making data more accessible for employees, helping leaders make data-driven decisions, and assisting companies in deploying AI technology.

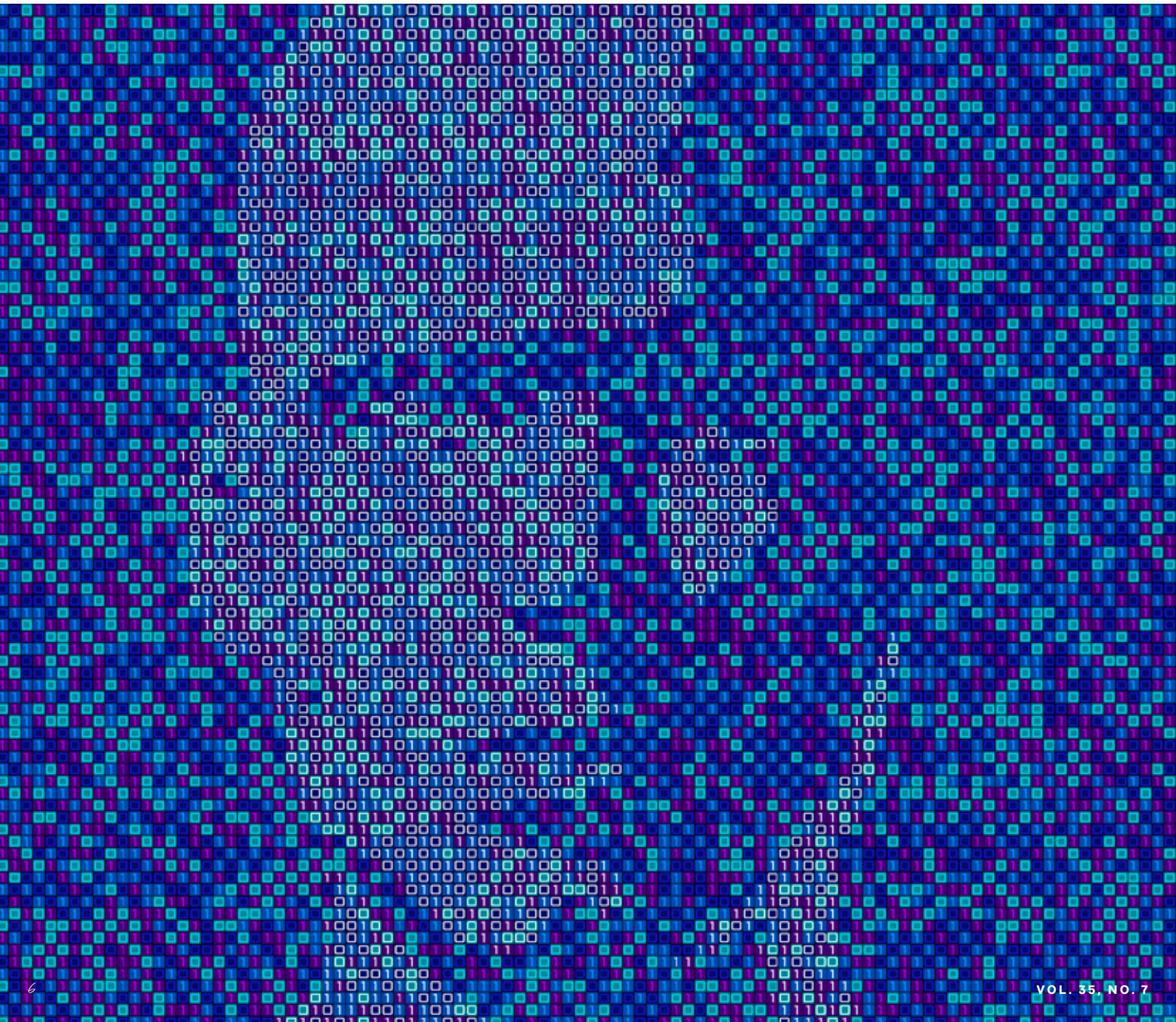
Next, Gurgur looks at KGs in the context of blockchain. The article begins with background information on how KGs have been used in advanced analytics and their role in helping AI developers. Gurgur then shows how blockchain's immutability and verifiability offer designers a way to advance KGs to produce more reliable results. The blockchain/KG combination is an ideal one to build more explainable AI systems, she says. Finally, Gurgur explains how KG-enabled information systems can be used in industrial settings to enhance product development lifecycles, improve factory safety, and enhance information systems to the point where employees need less technical knowledge to perform their duties.

> **BY COMBINING AI'S STATISTICAL/ML SIDE WITH KGs, WE GET MORE EFFECTIVE, MORE EXPLAINABLE COGNITIVE RESULTS AND BEGIN CREATING LOGIC-BASED SYSTEMS THAT GET BETTER WITH EACH APPLICATION**

Our third article is from George Hurlburt, who details how a KG was used to assist a regional center of a major university system in its course selection process. The KG helped leaders more clearly see the array of educational pathways from K-12 to community college (CC) coursework that are the results of articulation agreements between universities and CCs. Hurlburt shares five figures from the KG that demonstrate its meaningful visualizations. He also explains how the KG was built, including limiting the number of arcs and emphasizing node unambiguity. Finally, Hurlburt concludes with five key academic relationships and trends that are clearly demonstrated by the regional center's KG.

Our fourth article, by Williams, looks at how human-centric functional modeling (a way to allow computers to solve general problems) could be used to create KGs capable of providing compete semantic models of systems, enabling us to transition to Industry 5.0. He defines Industry 5.0 as a world in which far greater integration is possible, including functional computing approaches like GCI. Although the emergence of GCI isn't guaranteed (it could end up in a technology gravity well, says Williams), it would bridge type 1 and type 2 reasoning and lead to a radical increase in our ability to solve every problem.

Our final article — written by myself and my colleagues at Arthur D. Little, Philippe Monnot and Armand Rotaru — demonstrates KG use in the real world. We illustrate several prominent, real-world KG applications, then detail how we designed a KG to ensure vertical traceability within a systems engineering context. We began by extracting relevant entities from 20,000 heterogeneous files with the help of natural language processing (NLP) technologies and proceeded to define a suitable ontology that incorporated concepts from the field of systems engineering. We then developed an ML model that consumed features derived from the KG and mimicked the way an independent safety assessment auditor would work in practice.



Using precision and recall to evaluate the model's accuracy resulted in finding previously incorrectly labeled software requirement specifications. We also found that combining graph-based features with text-based ones boosts the classification accuracy significantly, thus showing significant promise in augmenting human safety assessors in the future. We end the article with some specific advice on using KGs, including unlocking new insights, extracting more from the data you have, and starting small with the intention of scaling quickly.

We hope you enjoy reading this issue (and viewing *Amplify's* brand-new design); we certainly enjoyed putting it together. We're hoping KGs' potential to take important processes and technologies to new levels will help business leaders better connect the dots.

REFERENCES

[1] Baudoin, Claude, and Clayton Pummill. "Bridging the AI Trust Gap." *Cutter Business Technology Journal* (renamed *Amplify*), Vol. 34, No. 5, 2021.

[2] Aasman, Jans. "Neuro-Symbolic AI: The Peak of Artificial Intelligence." AiThority, 16 November 2021.

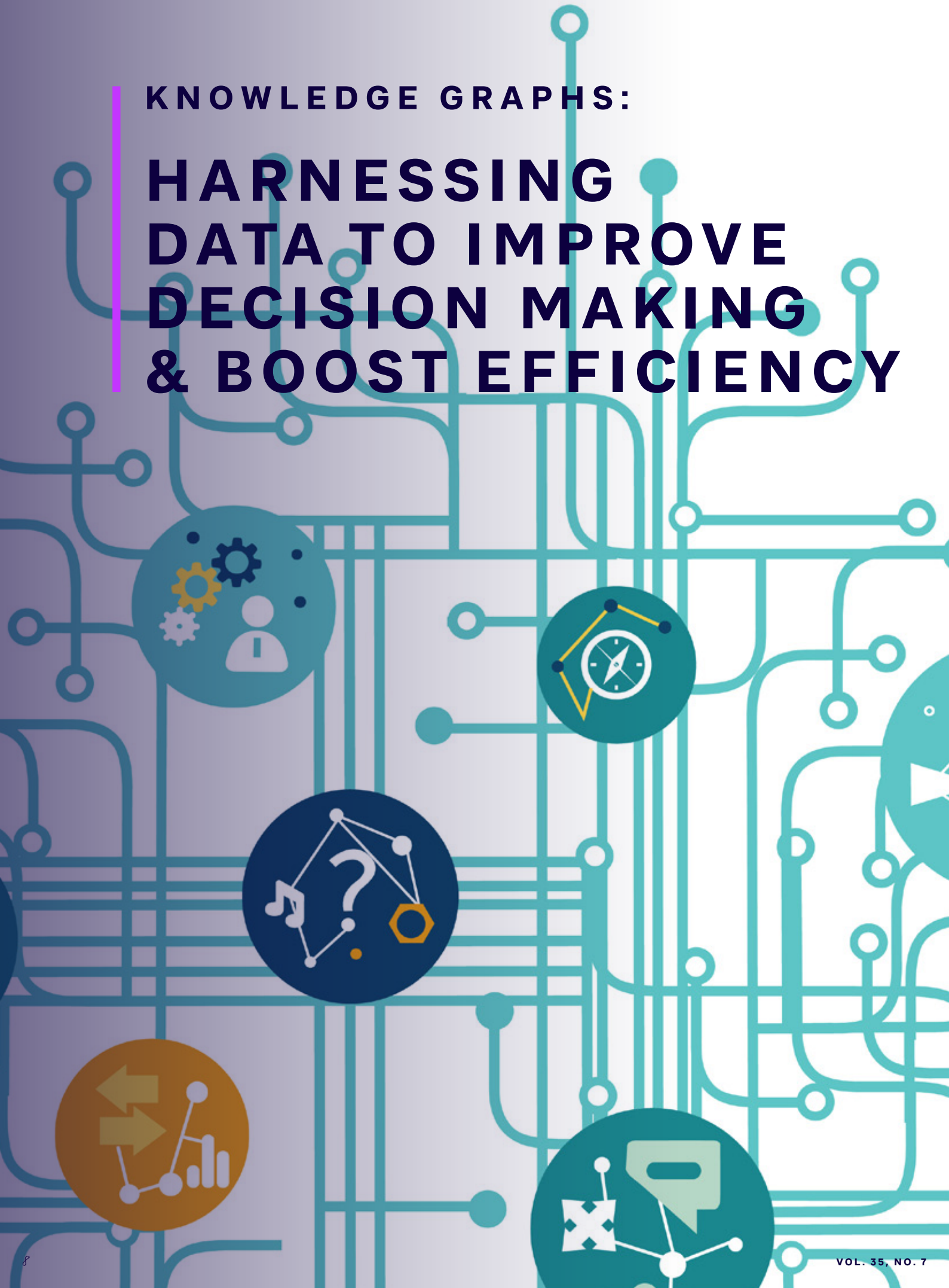
About the guest editor

MICHAEL EIDEN

Michael Eiden is a Cutter Expert and a member of the Arthur D. Little (ADL) AMP open consulting network. Dr. Eiden serves as Partner and Head of AI at ADL and is an expert in machine learning (ML) and artificial intelligence (AI) with more than 15 years' experience across different industrial sectors. He has designed, implemented, and productionized ML/AI solutions for applications in medical diagnostics, pharma, biodefense, and consumer electronics. Dr. Eiden brings along deep expertise in applying supervised, unsupervised, as well as reinforcement ML methodologies to a very diverse set of complex problem types. He has worked in various global technology hubs, such as Heidelberg (Germany), Cambridge (UK), and Silicon Valley (US), with clients ranging from small and medium-sized enterprises to globally active organizations. Dr. Eiden earned a doctorate in bioinformatics. He can be reached at experts@cutter.com.

KNOWLEDGE GRAPHS:

HARNESSING DATA TO IMPROVE DECISION MAKING & BOOST EFFICIENCY



Author

Lila Rajabion

A majority of businesses collect and store a substantial volume of data, but many don't adequately harness it to enhance their decision making or fuel new opportunities.[1] The sheer volume of data makes it difficult for companies to manage; this is compounded by the multiple silos in which data is stored. In this article, we'll look at how knowledge graphs (KGs) can help solve that problem, opening up avenues to improved decision making, better employee data access, and easier deployment of artificial intelligence (AI) technology. We'll also examine some real-world examples of KGs (including Google and others) and look at some of the challenges faced by companies as they develop KGs.

KGs have been around for quite a while, but they didn't receive much attention until Google began integrating them into its search engines. Today, large companies like Google, LinkedIn, and Amazon use KGs to optimize searches, but companies of any size can use them to improve data accessibility and searchability.

Today's emphasis on searchability is forcing content marketing and search engine optimization (SEO) experts to create rich networks of informative and instructional materials to satisfy customers during the buyer journey. Companies that don't excel at searching and retrieving data for their customers have trouble remaining competitive.[2] Using a methodological system like KGs to more efficiently manage that data thus becomes a strategic advantage.

For example, if a person wants to search Google for his or her favorite place to eat but only knows the location and not the name of the restaurant, Google, with the help of its KG, can provide relevant suggestions in real time. Similarly, KGs can improve a company's content marketing and SEO by: (1) unambiguously defining content for search engines and (2) building robust information environments around products and services for prospects and customers.[3]

BUSINESS USES

One of the most important KG functions is creating linkages across multiple data sets. By providing a visual representation of the underlying connections between data nodes, KGs help leaders advance their understanding of their environment so they can make intelligent business choices.[4] Here are four examples:

1. By providing a way to merge data silos, KGs create a valuable overview of all knowledge in a company, both within departments/divisions and across them. This is helpful for companies with multiple divisions, especially if they're located in different regions or countries.
2. KGs have the ability to connect different kinds of data in meaningful ways.[5] For example, academic graphs include people, papers, research topics, and conferences to help users detect connections between researchers and pieces of research.
3. By narrowing searches and contextualizing information, KGs can help business leaders make more informed decisions faster.[6]
4. By having each topic or item represented just once (with all its connections) in context with all other subjects and their relationships, KGs clearly show how each node is interconnected. This helps leaders gain perspective on how important ideas relate to one another.

REAL-WORLD EXAMPLES

The benefits of KGs are not limited to large tech companies. In fact, any company with a significant amount of data can benefit from them. Following are some examples of how companies are using KGs to improve content management and user-centric services — and how other companies could follow suit.

RATHER THAN CRAWLING THROUGH OR INDEXING WEBSITES, GOOGLE USES ITS KG TO ORGANIZE THE WORLD'S INFORMATION BY TOPIC

GOOGLE

The search results page on Google responds to questions the company has already addressed with the help of its KG. Since Google does not develop content, the results it displays originate from credible sources that are organized and linked, yet dispersed over the Internet.[7] Voice-activated assistants Google Assistant and Google Home use the same KG to answer verbal inquiries.

In other words, Google's KG is a knowledge base designed to improve its search engine results using information acquired from a variety of sources. Following its launch in 2012, Google's KG saw tremendous growth, more than tripling in a matter of months to reach 570 million entities and 18 billion facts by its most recent count.[8]

Rather than crawling through or indexing websites, Google uses its KG to organize the world's information by topic; advantages for the company include scale, data integrity, and speed. Google can easily harness user behavior data to understand what topics are significant to individuals and suggest topics based on user history. Other companies could use this approach, leveraging data to better understand customer behavior in order to improve products and/or marketing.

AWS

Amazon Web Services (AWS) KGs are a mechanism for modeling and conveying knowledge about the company's services. This concept has been around for a while, but the development of scalable graph databases has made it more applicable.[9] Compared to data management systems like relational databases, KGs are extraordinarily adaptable, capable of accounting for the variety and heterogeneity of data in the real world.

Using a collection of ideas, the properties of those concepts, the interactions between those concepts, and the logical constraints that are expected to hold, AWS KGs can capture the semantics of a specific domain.[10] Because this model includes logic, we can reason about graphs and the information included within them, making the information implicit in the graph readily available. The process of information asset consolidation includes integrating an organization's information assets and making them easily accessible to all members of an organization.[11]

AWS KGs open the door to a variety of applications, most of which are helpful on their own, not only for the company but for its clients. For example, Amazon could turn the data it gathers into a more helpful resource by using an enterprise KG. Furthermore, it could develop corporate knowledge graphs by using the built-in federated query functionalities of the Amazon Neptune graph database.[12] Public data from the Internet could be used to enrich the information already included within these graphs. Other companies can similarly use KGs to help them organize information from dissimilar data sources to enable more intelligent search. Ultimately, KGs can help organizations make their data more understandable by using business terms rather than ambiguous codes.

LINKEDIN

LinkedIn's KG is an enormous knowledge base constructed from entities such as members, jobs, titles, skills, companies, geographical locations, schools, and the connections between them.[13] LinkedIn uses this ontology to improve its recommendation system; search, monetization, and consumer product offerings; and business and consumer analytics.

Developing this type of comprehensive knowledge base proved extremely challenging. Websites like Wikipedia and Freebase are almost entirely dependent on user contributions.[14] LinkedIn took a different approach. LinkedIn's KG is primarily derived from the large quantity of content provided by corporate administrators, recruiters, advertisers, and other users.[15]

The KG grows constantly as individuals sign up for the platform, employment opportunities become available, new companies join, new skills are added, and new titles surface in user profiles and job ads.

Moreover, the company uses machine learning (ML) methods to help find solutions to its KG network challenges.[16] This is essentially a process of data standardization on user-generated content and external data sources. ML is applied to entity taxonomy construction, entity-relationship inference, data representation for downstream consumers, insight extraction from the graph, and interactive data acquisition from users to validate inferences.[17]

New entities are continuously added to the KG, and new connections are forged between existing entities. Alterations to existing partnerships are also possible. For instance, when a member gets a new position, the mapping from her previous title to her present one is updated accordingly. It is necessary to perform real-time updates on the LinkedIn KG network whenever member profiles undergo modifications or when entities are added. Other companies could similarly take advantage of ML to help them improve their data quality and KGs.

EBAY

eBay's product knowledge graph encodes semantic knowledge about items, entities, and their connections. This information is vital to eBay's marketplace technology, which automatically connects sellers and buyers. eBay uses KGs to describe products, schedule deliveries, and service customers through virtual assistants. eBay's KG sometimes links items to real-world entities, establishing a product's identity and value to a customer.

The KG also links goods. For example, if a person looks for Lionel Messi memorabilia, and the KG shows he plays football (soccer) for FC Barcelona, that person may also be interested in FC Barcelona items or items like signed jerseys from other Barcelona players.

For eBay, understanding product connections is as important as entity interactions, and the knowledge network must answer a search query in milliseconds. Because large graph queries can take hours to complete, eBay engineers built a flexible, universal architecture. The KG keeps track of every entry and change, and the data is organized in a log. This enables a variety of back-end data storage options, such as low-latency document storage and a graph store for long-running analysis. To keep the graph in chronological order, each store adds its operations to the write log, resulting in more consistent results for customers.

Other e-commerce companies could similarly use KGs, leveraging entity relations to better understand their products' relationships (e.g., suggesting an iPhone case to someone who just purchased an iPhone and successfully modeling various phone sizes and cases in order to offer a case that fits the phone bought).

> **KGS VARY GREATLY IN SCOPE AND DESIGN, BUT THE CHALLENGES IN CREATING THEM ARE SIMILAR FOR MOST IMPLEMENTATIONS**

IBM

Watson Discovery services uses IBM's KG framework in two ways. First, the framework directly supports Watson Discovery, leveraging structured and unstructured knowledge to discover new information. Second, it allows individuals to construct KGs based on the prebuilt KG. Discovery creates knowledge not present in existing documents or available data sources. Examples include connections between entities (e.g., drug side effects, acquisition targets, and sales leads), new important entities in the domain (e.g., an investor for a specific investment area), or changes in the significance of an existing entity (e.g., an increasing interaction between a person of interest and a criminal).[18] Other companies could similarly leverage KGs to identify prospects, current customers who might be interested in other products, and potential investors.

KG CHALLENGES

KGs have been used to improve search results across a variety of search engines, including Google and Bing, and to provide support for a large number of applications. Amazon is developing a product graph that will serve as an official KG for all the items in the world. The thousands of product verticals we need to model, the vast number of data sources we need to extract knowledge from, the enormous volume of new products we need to handle every day, and the number of applications (search, discovery, personalization, and voice) we wish to support present significant challenges when it comes to the construction of such a graph. KGs vary greatly in scope and design, but the challenges in creating them are similar for most implementations.

DISAMBIGUATION & CONTROL OF INDIVIDUAL IDENTITIES

Resolving ambiguity between entities is a significant difficulty in Semantic Web and KGs. Problems arise when an entity's name or mention is not given its own normalized identity and type in the context of a conversation. Many autonomously generated things, such as people with similar names and book/movie titles, have similar surface forms to each other. Likewise, comparable products may be listed under various headings. A lack of appropriate linking and disambiguation can lead to inaccurate judgments about entities. The difficulty of identity management rises exponentially when dealing with many contributors on a large scale.

RESOLUTIONS & MEMBERSHIP

There are several types of entities in most KGs. For instance, Angelina Jolie is a human, an actor, and a humanitarian; she's better known for acting than her humanitarian efforts. A KG might employ a particular set of attributes based on a user's job, and early on, the criteria for being a class member might be easy to understand. As the number of instances grows, it becomes harder to enforce these criteria while maintaining semantic stability. For example, e-sports did not exist when Google set up the category for sports in its KG. So how does Google keep the sports category separate from e-sports while still including them?

KNOWLEDGE MANAGEMENT

A good entity-linking system must grow naturally based on the data it receives, which is constantly changing. Companies can merge or split, and new scientific discoveries can make one thing into more than one. KG frameworks are getting better at storing and managing changes, but they still cannot manage highly dynamic information.

Maintaining several stores (e.g., IBM's polymorphic stores) may be difficult. There are many things to think about regarding the integrity of the update process, its eventual consistency, updates that conflict, and runtime performance in general. The answer may be different kinds of distributed data stores that are already set up to handle incremental cascade updates. It's also essential to keep track of changing schemas and type systems without making the knowledge already in the system inconsistent. Google solves this problem by thinking of the metamodel layer as comprised of several layers. The lower layers stay mostly the same while the higher levels are made up of meta types, which are just instances of types that can be used to improve the type system.[19]

EXTRACTION OF INFORMATION FROM A VARIETY OF ORGANIZED & UNORGANIZED SOURCES

Although there have been recent improvements in understanding natural language, it's still hard to pull out structured knowledge, which includes entities, their types, attributes, and relationships. To grow KGs at scale, we must use manual methods alongside unsupervised and semi-supervised methods to extract knowledge from unstructured data in open domains. For example, eBay's product knowledge graph gets many of its graph relationships from the unstructured text in listings and seller catalogs. The IBM Discovery KG gets its facts from documents.

Training knowledge-extraction systems in traditional supervised ML frameworks is difficult and time-consuming. This can be mitigated by using fully unsupervised or semi-supervised approaches. Entity recognition, classification, text, and entity embeddings are all useful ways to connect unstructured text to graph entities.

CAN KGs IMPROVE BUSINESS EFFICIENCY?

There are several ways KGs can improve business efficiency. The first is by creating an advanced way for business leaders to merge, sort, and view data.[20] KGs create a web of information on a subject, pulling from multiple sources and merging various data types to help leaders better understand their company's reality and make data-driven decisions.

The second way is helping employees quickly gain access to the information they need. KGs make it easier to understand internal assets, such as benefits, tax information, organizational structure, and more.

The third way is helping companies deploy AI technology, such as chatbots and advanced search. KGs can act as inputs for ML, since ML algorithms achieve better results if they include domain knowledge. KGs help capture domain knowledge, but ML algorithms require that any discrete structure, such as a graph, first be converted to a numerical format.

> **RESOLVING AMBIGUITY BETWEEN ENTITIES IS A SIGNIFICANT DIFFICULTY IN SEMANTIC WEB AND KGs**

A KG could help chatbots memorize, associate, and reason about the semantic connections between entities, bridging the gap from perceived intelligence to cognitive intelligence. However, we must keep in mind that logical reasoning remains a challenge for KGs. For example, a medical chatbot could collect symptoms and offer basic medical advice but is not intended to replace a physician's diagnosis or advice.

CONCLUSION

KGs can play an important role in business, particularly in improving decision making and boosting efficiency. By merging data silos, KGs create a valuable overview of all the knowledge in a company, both within and across departments/divisions. Similarly, by narrowing searches and contextualizing information, KGs can help business leaders make more informed decisions faster. Because each topic or item in a KG is represented just once in context with all other subjects and their relationships, KGs show node interconnections that help leaders gain perspective on how important ideas relate to one another.

Companies such as Google, AWS, LinkedIn, eBay, and IBM are already using KGs to improve searches, make data more accessible to leaders and employees, improve product suggestions made to customers, and much more. KGs can also help companies with their AI deployments, including chatbots. KGs act as ML inputs, adding domain knowledge to help ML algorithms achieve better results.

As customers become ever more accustomed to fast, accurate product searches and expect to consistently receive useful suggestions for ancillary purchases, KGs are an important tool for companies hoping to successfully satisfy them during the buyer journey.

REFERENCES

[1] Nickel, Maximilian, et al. "A Review of Relational Machine Learning for Knowledge Graphs." *Proceedings of the IEEE,* Vol. 104, No. 1, December 2015.

[2] Collarana, Diego, et al. "FuhSen: A Federated Hybrid Search Engine for Building a Knowledge Graph On-Demand (Short Paper)." *On the Move to Meaningful Internet Systems: OTM 2016 Conferences*. Springer, 2016.

[3] Nickel et al. (see 1).

[4] Kejriwal, Mayank. "Knowledge Graphs." In *Applied Data Science in Tourism: Tourism on the Verge*, edited by Roman Egger. Springer, 2022.

[5] Grangel-González, Irlán, Felix Lösch, and Anees ul Mehdi. "Knowledge Graphs for Efficient Integration and Access of Manufacturing Data." *Proceedings of the 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*. IEEE, 2020.

[6] Galkin, Mikhail, et al. "Enterprise Knowledge Graphs: A Semantic Approach for Knowledge Management in the Next Generation of Enterprise Information Systems." *Proceedings of the 19th International Conference on Enterprise Information Systems (ICEIS 2017)*. SciTePress, 2017.

[7] Heist, Nicolas, et al. "Knowledge Graphs on the Web — An Overview." Cornell University, March 2020.

[8] Balog, Krisztian, and Tom Kenter. "Personal Knowledge Graphs: A Research Agenda." *Proceedings of the 2019 ACM SIGIR International Conference on Theory of Information Retrieval (ICTIR 2019)*. ACM, 2019.

[9] Zheng, Da, et al. "DGL-KE: Training Knowledge Graph Embeddings at Scale." *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 2020.

[10] Zheng et al. (see 9).

[11] Baclawski, Ken, et al. "Ontology Summit 2020 Communiqué: Knowledge Graphs." *Applied Ontology*, Vol. 16, No. 2, 2021.

[12] Szekely, Pedro, et al. "Building and Using a Knowledge Graph to Combat Human Trafficking." *Proceedings of the Semantic Web — International Semantic Web Conference (ISWC) 2015*. Springer, 2015.

[13] Chen, Xi, et al. "How LinkedIn Economic Graph Bonds Information and Product: Applications in LinkedIn Salary." *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD)*. ACM, 2018.

[14] Dong, Xin, et al. "Knowledge Vault: A Web-Scale Approach to Probabilistic Knowledge Fusion." *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*. ACM, 2014.

[15] Auradkar, Aditya, et al. "Data Infrastructure at LinkedIn." *Proceedings of the IEEE 28th International Conference on Data Engineering*. IEEE, 2012.

[16] Auradkar et al. (see 15).

[17] Auradkar et al. (see 15).

[18] Kejriwal (see 4).

[19] Fensel, Dieter et al. "Introduction: What Is a Knowledge Graph?" In *Knowledge Graphs*. Springer, 2020.

[20] Yahya, Muhammad, John G. Breslin, and Muhammad Intizar Ali. "Semantic Web and Knowledge Graphs for Industry 4.0." *Applied Sciences*, Vol. 11, No. 11, 31 May 2021.

About the author

Lila Rajabion is Assistant Professor and coordinator of the Master of Science in Information Technology (MSIT) program at SUNY Empire State College, where she teaches and develops the MSIT curriculum with a concentration in cybersecurity and Web design. She has more than 20 years' experience conducting research and providing consulting in various dimensions of IT combined in the academia and private sectors. Dr. Rajabion also has significant professional experience in providing leadership in the areas of systems analysis and design, cybersecurity, enterprise software application development, and IT project management for local and global projects. In addition, she has conducted various needs-based training programs. Dr. Rajabion has written many publications and has participated in research grants for the National Center for Women & Information Technology, which focuses on women in STEM. She is a long-time advocate for increasing participation and retention of women and minorities in the IT workforce. Dr. Rajabion has worked on many projects in this domain. She can be reached at lrajabion@sar.usf.edu.

KNOWLEDGE GRAPHS MEET BLOCKCHAIN:

BOOSTING PRODUCTIVITY IN INDUSTRIAL PRODUCTS WITH TRUSTWORTHY & EXPLAINABLE ML

Author

Cigdem Z. Gurgur

Advances in high-tech sensing, the proliferation of electronic manufacturing records and mobile sensors, and the development of the Industrial Internet of Things (IIoT) are causing manufacturing data to accumulate exponentially. Although this data is often stored in heterogeneous formats and distributed, it's an important source of the information we need to deploy intelligent production management tools. The process involves knowledge extraction and prediction processes using artificial intelligence (AI) models, the success of which is mainly due to advances in machine learning (ML).

ML, a subset of AI, enables learning from observations (data) and experience (repeated training) and is key to transforming large manufacturing data sets (often called "big industrial data") into actionable knowledge. This is stimulated by large data sets involving various real-world features and an increase of the computational gains generally attributed to powerful graphic-processing cards.

Knowledge graphs (KGs) that leverage AI and ML technology, particularly deep learning, are now being widely studied for use in manufacturing because of their ability to easily handle large amounts of data and model complex relationships.[1] Deep learning, a subset of ML, learns without human supervision. KGs are a powerful data science technique created to mine information from diverse data formats.

KGs' ability to handle connected data and embrace relationships in a flexible, graphic form makes them highly efficient in domains where data structures are constantly changing and evolving, as in manufacturing. Flexible KG schemas can handle dynamic, uncertain variables and quickly encode domain and application knowledge. KGs complement ML methods, enabling accurate data collection that facilitates faster, more precise AI application development.

For example, a recent *Technological Forecasting and Social Change* article showed how smart manufacturing could use advanced manufacturing technology and data-mining techniques like KGs to improve product quality while shortening production cycles, enhancing production efficiency, and reducing costs.[2]

THE HISTORY OF KGs IN ADVANCED ANALYTICS

In the last 10 years or so, KGs have emerged as an important area in advanced analytics and the AI domain, helping to connect data sources and solve large-scale enterprise problems. They are at the core of human-facing technologies, such as search, question answering, dialogue, fraud prevention and investigation, product recommenders, and autonomous systems.

In addition to business problems, KGs have been used to solve social problems involving difficult technical challenges, such as human trafficking. A recent article in *IEEE Transactions on Big Data* described a KG effort that resulted in an effective semantic search engine to assist analysts and investigative experts in the human-trafficking domain.[3]

Because data sets are so often scattered, AI developers struggle to discover, share, and manage data from different systems in different formats. This requires understanding, structuring, integrating, and verifying data each time new features or applications are built based on it. One of the main benefits of structuring knowledge in the form of graphs (instead of relational databases) is the flexibility of the schema, which can be defined at a later stage and adjusted over time. This allows more flexibility for data evolution and capturing incomplete knowledge.

Building on a storied tradition of graphs in the AI community, a KG can be defined as a directed, labeled, multi-relational graph with some form of semantics integrating diverse data into a common format. KGs provide graph-structured topologies to organize data and can present interlinked descriptions of its entities, including objects, events, situations, and abstract concepts.[4]

IN THIS ERA OF INFORMATION EXPLOSION, KGs HAVE TREMENDOUS POTENTIAL TO ELICIT, INTEGRATE, PROCESS, USE, AND POPULARIZE LARGE DATA SETS

A 2012 Google blog entry is often cited as having sparked KG development.[5] In truth, Google revived KG technology rather than inventing it — a great deal of KG research was done in the 1980s. For Google, KG technology was about enhancing search engine performance through information gathered from a variety of sources.

In 2016, researchers Lisa Ehrlinger and Wolfram Wöß proposed a more widely acknowledged definition: "A knowledge graph acquires and integrates information into an ontology and applies a reasoner to derive new knowledge."[6]

In this era of information explosion, KGs have tremendous potential to elicit, integrate, process, use, and popularize large data sets embedded in industrial products and services. KGs allow reasoning about the underlying data, provide significant increased precision with information retrieval, and facilitate complex decision making.[7]

KGs' underlying structure offers both humans and machines better knowledge comprehension and interpretation.[8] Today's KGs supplement manual knowledge-engineering techniques with crowdsourcing and use ML to significantly increase automation.

Although KGs can improve AI predictions by providing them with knowledge expressed and used by ML methods, most ML models require a set of feature vectors as input. As a result, considerable research has been done to generate "embeddings" from KGs. A KG embedding transforms the nodes and the edges of the graph topology to a numeric feature vector that can serve as a direct input to the ML model.[9]

AI experts are therefore manipulating structured KGs for deep learning with relational inductive techniques, transferring learning (inter-domain knowledge sharing), and seeking other methods of infusing KG into ML.[10] In some cases, KGs have been used to extend existing data models depicted by domain ontologies and establish a new form of advanced analytics that can capture large, semantically interconnected data sets.[11]

Even though directed labeled graphs represent a common thread linking today's KGs with early AI semantic networks, there are some important differences in research methodologies and technical challenges. In early AI semantic networks, the emphasis was on complex logical inferencing; modern KGs focus on supporting advanced analytics operations.[12]

Additionally, early semantic networks were created using top-down design methods and manual knowledge-engineering processes. They never reached the size and scale of today's KGs. Modern KGs are larger in scale and are constructed using both manual and automated strategies. The vast proliferation of available data and the data-driven nature of today's ML support a bottom-up methodology for creating KGs.

Unlike rigid relational database structures, KGs' flexible semantic data layer allows users to perpetually link and network the complex relationships contained within their data platform and external sources without changing the underlying data, thus enriching the data's semantic meaning.

ENHANCING ENTERPRISE KGs WITH BLOCKCHAIN

It began as a digital currency technology, but blockchain has rapidly entered virtually every aspect of our lives, from enhancing food safety and preventing medical errors to diamond provenance information disclosure and artwork ownership authentication. Blockchain provides decentralized trust management in chronological, encrypted, chained blocks to store verifiable, synchronized data across peer-to-peer networks.

Blockchains maintain their data integrity, while providing tamper-proof and secure data storage and immutable task execution. Blockchain's unique tracing ability lets it identify malicious activity on the network. If a malicious user tries to tamper with an enterprise KG, he or she has to access the pointer from the blockchain, so the user's account can be traced.

The KG/blockchain combination marries integrity with interoperability and interconnectivity. Blockchain's visibility into the decisions of all AI agents on a KG network makes it difficult for AI agents to modify or refute decisions. Blockchains also let AI agents collaborate to save new decisions on blocks that can be traced back and are therefore resistant to alteration.

A forward-looking article in *IEEE Access* recommends using crowdsourcing on a blockchain platform to update KG and AI systems with a "trustful" value.[13] The research proposes a cutting-edge, decentralized KG construction method using crowdsourcing, with the business logic of crowdsourcing implemented by blockchain-powered smart contracts to guarantee transparency, integrity, and auditability. This technique represents a beneficial tradeoff between the completeness and the correctness of KG, as it takes full advantage of the wisdom of crowds.

Another set of researchers created a visionary framework to enhance KGs with fundamental blockchain concepts, improving the reasoning algorithm with trustworthy and historical knowledge to produce more reliable results.[14] The framework includes a verified, trusted state provided by blockchain technology in KGs to help an AI system show why it made a specific decision. Fully provable explanations of AI decisions can be produced by going back in time via blockchain technology.

> **BLOCKCHAINS MAINTAIN THEIR DATA INTEGRITY, WHILE PROVIDING TAMPER-PROOF AND SECURE DATA STORAGE AND IMMUTABLE TASK EXECUTION**

Having such an integrated system could provide a path to real-time KGs, amalgamating the unmodifiable and accessible history concept and providing verified KGs by blending the concept of digital signatures, which would build a secure connection between KGs and blockchains.

Furthermore, since KGs are designed for complex data and knowledge integration tasks as well as reasoning tasks and do not require hard-coding knowledge into reasoning algorithms, they resolve the scalability challenges in blockchain implementation.

Semantic linking to a data source is one aspect of what is usually called "provenance information." Provenance tracks the origin of that data and is one form of metadata that is rarely captured in typical relational databases. However, it is relatively easy to capture in KGs.

The convergence of blockchains and KGs allows dynamic enrichment of logic that ends up in a decentralized graph for trustworthy decision making.

TRUSTWORTHY & EXPLAINABLE AI SYSTEMS BY BLOCKCHAIN AND KGs

Increasing computational power and big data proliferation are driving AI system adoption. Decision support algorithms are carried out by mathematical models (trained using ML techniques) on data collected from past experiences. However, the opaqueness of AI decisions is a major drawback in systems like industrial design, where precision and safe product development are required.

AI technologies that can provide human-understandable explanations for their output or actions are usually referred to as "explainable AI." End users wonder about the reasoning behind the decisions made by algorithms, and increasing complexity results in a lack of transparency that negatively affects user trust.

The Cambridge Analytica scandal and the 2016 US election disruptions clearly showed why modern ML methods need to be more transparent.[15] A number of initiatives have been launched since then, including the US Defense Advanced Research Projects Agency's Explainable AI program[16] and the EU's Ethical Guidelines for Trustworthy AI.[17] Both encourage the design of ethical systems that humans can understand, manage, and trust.

One way to build explainable AI systems is by using KGs. New research in *Semantic Web* shows how KGs represent a valuable form of domain-specific, machine-readable knowledge.[18] The resulting KG-connected or centralized data sets can serve as background knowledge for AI systems to better explain their decisions to users.

Explainable AI is also needed to combat adversarial attacks against ML and deep neural networks that may poison learning or inference processes. These attacks come in a variety of flavors, such as data set poisoning, internal network manipulation, and side-channel attacks. Malicious actors can cause random or targeted misclassifications by manipulating the environment around the AI system, the data acquisition block, or the input samples. The attack can be as simple as adding adversarial noise to input samples and as malicious as incrementally shifting the decision boundaries during the ML training process.

Blockchain technology can be used to produce trustworthy AI requirements to mitigate biases and guard against adversarial attacks.[19] With blockchain, explanation systems, including decision outcomes, can be audited in an immutable, tamper-proof, decentralized way that can be traced with high reliability. If any node fails or leaves the chain, the blockchain remains unaffected.

By merging blockchain technology with KGs, we can achieve next-generation industrial information systems for secure data sharing among stakeholders, maintaining data privacy and integrity through data authentication and robust data adaptation. This type of industrial platform would improve trust, elevate scalability, and increase efficiency through multi-party and multi-agent decision-making systems that follow various consensus protocols. It could be used to host a trusted trail of all records used by ML algorithms before, during, and after the learning and decision-making process.

ENHANCING PRODUCTIVITY THROUGH KG-ENABLED INDUSTRIAL PRODUCT DEVELOPMENT: RECENT EXAMPLES

DEMAND FORECASTING

Demand forecasting and requirement analysis are crucial topics in industrial product development, and they require massive information inputs and robust analytic models to make better predictions. Processing multi-source information and conducting logical knowledge reasoning are two major strengths of KG-enabled information systems. The explainable capability of knowledge reasoning and recommendations enabled by KG are valuable for demand forecasting and requirement analysis, given that stakeholders care more about the insights and logic behind the results than about ordinary point estimates.

Recently, an article in *International Journal of Production Research* demonstrated KGs' ability to collect extensive information from online technical forums and portal websites to capture market trends and other events impacting consumer demand.[20]

SMART SOLUTION DESIGN

Industrial product development requires a high degree of knowledge synthesis and precision specification. KGs' ability to gather data from multiple sources, usually in different formats, facilitates the creation of easily extendable, flexible data models ideal for made-to-order manufacturing. Several studies have shown how such automated knowledge extraction and fusion improve manufacturing design capacity.[21] Furthermore, KGs' real-time information exchange enables last-minute customer changes, even after production has begun.

KG-based design systems not only automatically save and store final solutions, but also earlier rejected ideas. The solutions and ideas are stored as knowledge in the KG, creating a more holistic knowledge base for the manufacturer that can enhance the product development lifecycle.[22]

RISK PREDICTION & SOLUTION PRESCRIPTION

Recently, researchers in the manufacturing field have attempted to drive industrial services with KG to optimize process safety and product quality. A paper in *Systems Research and Behavioral Science* proposed an advanced paradigm to apply KGs in smart factories to support safety management in the manufacturing process.[23] Researchers proposed KGs as a way to: (1) improve decision making based on problem diagnoses and (2) predict potential risks based on information (e.g., worker location or machine status) and suggest preventative measures.

Similarly, a recent article in *Computers in Industry* showed how design rules and context information could be combined to build a computable KG, improving computer-aided design and allowing designers to spend time on design rather than looking for design rules.[24]

Through a better understanding of the relationship between function-behavior-structure and knowledge representation, a KG-based risk prediction and prescription system could prompt smart components to adjust themselves to solve problems.[25]

INFORMATION DISTILLATION

In the industrial product lifecycle efforts, there is typically a huge gap between massive heterogeneous knowledge resources in information systems and system users' limited cognitive ability. Holistic, nonspecific information is either useless or confusing to users. What's needed is an information system that can dispense the right information at the right moment to those working on specific designs.

With their ability to distill information, KGs can support those designers to better complete creative manufacturing tasks. For example, a KG-based system was able to deliver helpful information in a multi-language environment to employees without a technical background in fashion design manufacturing.[26]

CONCLUSION

KG-enabled multi-disciplinary information systems integrated with blockchain technology can facilitate industrial data mining with trustworthy principles. Such systems are capable of producing originative ideas to help users productively and safely complete product development process tasks with increased precision.

Demand forecasting and requirement analysis, smart engineering solution design, automatic risk prediction and prescription, operational maintenance, and information distillation all lead to time and manpower savings.

By leveraging KGs and blockchains, manufacturing enterprises can tap into innovations like explainable AI; reusable semantic data modeling; and scalable, trustworthy, complex-query performance to help accelerate advanced analytics insights and reduce data operations cost.

REFERENCES

[1] Gordana, Zeba, et al. "Technology Mining: Artificial Intelligence in Manufacturing." *Technological Forecasting and Social Change*, Vol. 171, October 2021.

[2] Zeba et al. (see 1).

[3] Kejriwal, Mayank, and Pedro Szekely. "Knowledge Graphs for Social Good: An Entity-Centric Search Engine for the Human Trafficking Domain." *IEEE Transactions on Big Data*, Vol. 8, No. 3, 1 June 2022.

[4] Pan, Jeff Z., et al. *Exploiting Linked Data and Knowledge Graphs in Large Organisations*. Springer, 2017.

[5] Singhal, Amit. "Introducing the Knowledge Graph: Things, Not Strings." *The Keyword*, Google, 16 May 2012.

[6] Ehrlinger, Lisa, and Wolfram Wöß. "Towards a Definition of Knowledge Graphs." *Joint Proceedings of the Posters and Demos Track of the 12th International Conference on Semantic Systems (SEMANTiCS 2016) and the 1st International Workshop on Semantic Change & Evolving Semantics (SuCCESS'16) Co-located with the 12th International Conference on Semantic Systems (SEMANTiCS 2016)*. CEUR Workshop Proceedings, 2016.

[7] Dalle Lucca Tosi, Mauro, and Julio Cesar dos Reis. "Understanding the Evolution of a Scientific Field by Clustering and Visualizing Knowledge Graphs." *Journal of Information Science*, Vol. 48, No. 1, 1 February 2022.

[8] Sheth, Amit, Swathi Padhee, and Amelie Gyrard. "Knowledge Graphs and Knowledge Networks: The Story in Brief." *IEEE Internet Computing*, Vol. 23, No. 4, 17 October 2019.

[9] Wang, Quan, et al. "Knowledge Graph Embedding: A Survey of Approaches and Applications." *IEEE Transactions on Knowledge and Data Engineering*, Vol. 29, No. 12, 1 December 2017.

[10] Hogan, Aidan, et al. "Knowledge Graphs." *ACM Computing Surveys*, Vol. 54, No. 4, May 2022.

[11] Nanduri, Jay, et al. "Microsoft Uses Machine Learning and Optimization to Reduce E-Commerce Fraud." *INFORMS Journal on Applied Analytics*, Vol. 50, No. 1, 24 January 2020.

[12] Chaudhri, Vinay K., et al. "Knowledge Graphs: Introduction, History, and Perspectives." *AI Magazine*, Vol. 43, No. 1, 31 March 2022.

[13] Wang, Shuai, et al. "Decentralized Construction of Knowledge Graphs for Deep Recommender Systems Based on Blockchain-Powered Smart Contracts." *IEEE Access*, Vol. 7, 19 September 2019.

[14] Bellomarini, Luigi, et al. "Blockchains as Knowledge Graphs — Blockchains for Knowledge Graphs (Vision Paper)." *Proceedings of the International Workshop on Knowledge Representation and Representation Learning (K4RL) Co-located with the 24th European Conference on Artificial Intelligence (ECAI 2020)*. CEUR Workshop Proceedings, 2020.

[15] Hern, Alex. "How Social Media Filter Bubbles and Algorithms Influence the Election." *The Guardian*, 22 May 2017.

[16] Gunning, David, et al. "DARPA's Explainable AI (XAI) Program: A Retrospective." *Applied AI Letters*, Vol. 2, No. 4, 4 December 2021.

[17] "Ethics Guidelines for Trustworthy AI." European Commission, 8 April 2019.

[18] Lecue, Freddy. "On the Role of Knowledge Graphs in Explainable AI." *Semantic Web*, Vol. 11, 2019.

[19] Salah, Khaled, et al. "Blockchain for AI: Review and Open Research Challenges." *IEEE Access*, Vol. 7, 1 January 2019.

[20] Wang, Zuoxu, et al. "A Graph-Based Context-Aware Requirement Elicitation Approach in Smart Product-Service Systems." *International Journal of Production Research*, Vol. 59, No. 2, 18 December 2021.

[21] Yuan, Jianbo, et al. "Constructing Biomedical Domain-Specific Knowledge Graph with Minimum Supervision." *Knowledge and Information Systems*, Vol. 62, 23 March 2019.

[22] Wang, Ru, et al. "A Process Knowledge Representation Approach for Decision Support in Design of Complex Engineered Systems." *Advanced Engineering Informatics*, Vol. 48, April 2021.

[23] Liu, Zimei, et al. "A Paradigm of Safety Management in Industry 4.0." *Systems Research and Behavioral Science*, Vol. 37, No. 4, 25 June 2020.

[24] Huet, Armand, et al. "CACDA: A Knowledge Graph for a Context-Aware Cognitive Design Assistant." *Computers in Industry*, Vol. 125, February 2021.

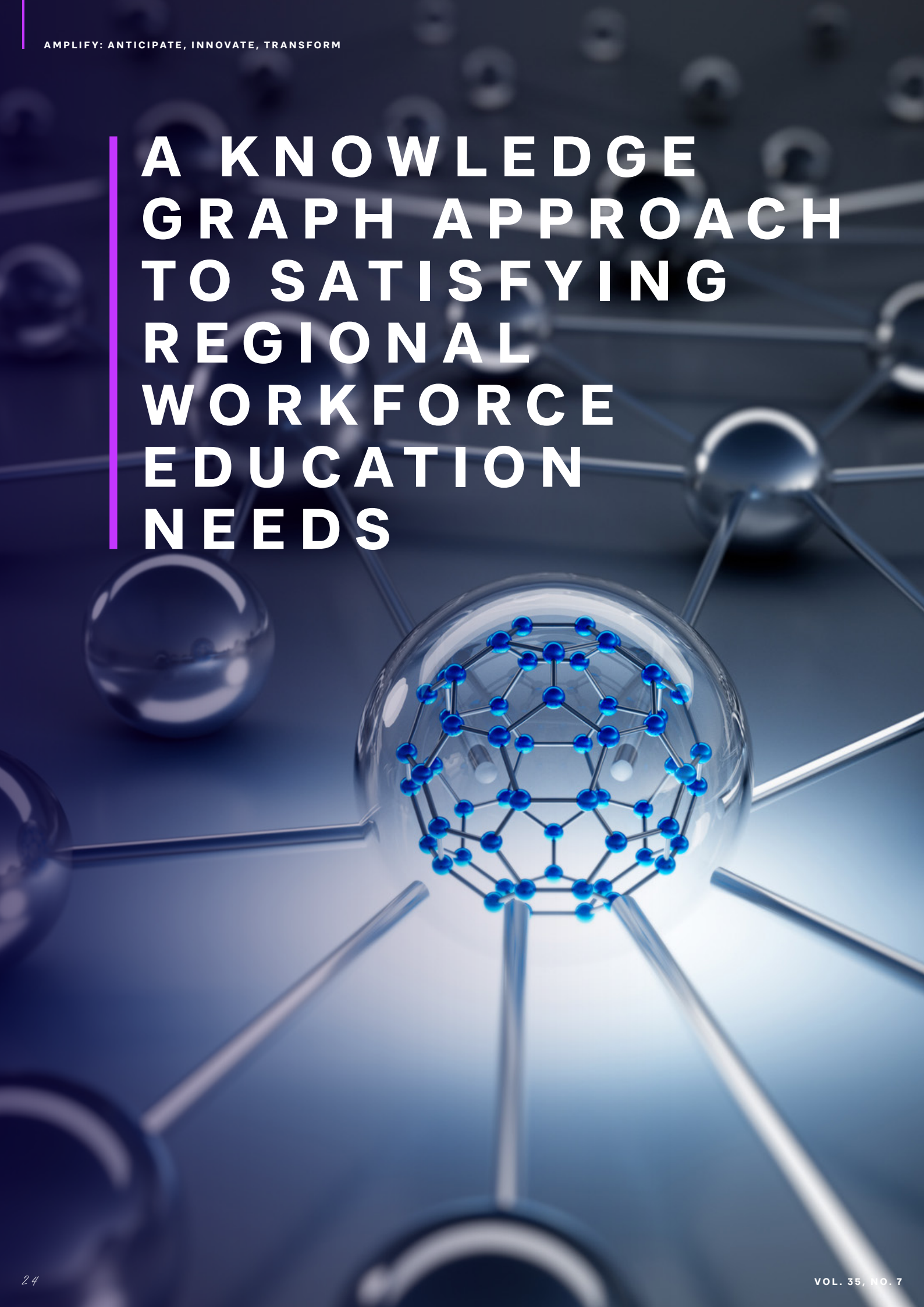
[25] Shi, Hui-Bin, et al. "An Information Integration Approach to Spacecraft Fault Diagnosis." *Enterprise Information Systems*, Vol. 15, No. 8, 2021.

[26] Peroni, Silvio, and Fabio Vitali. "Interfacing Fast-Fashion Design Industries with Semantic Web Technologies: The Case of Imperial Fashion." *Journal of Web Semantics*, Vol. 44, May 2017.

About the author

Cigdem Z. Gurgur is Associate Professor of Decision and System Sciences at Purdue University. She is a data and management science expert with experience in optimization models under uncertainty and decision support systems development with algorithmic theory design. Dr. Gurgur's work utilizes meta-analytics, computational models, and artificial intelligence techniques for resource allocation and applies mathematical programming integrating financial and operational risk assessment. She conducts interdisciplinary research geared toward solving complex technical and societal challenges within sustainable supply chain management, healthcare operations, medical wire and device manufacturing, and technology and innovation. Dr. Gurgur's most recent work encompasses blockchain technologies in advancing applied research for data science and the United Nations (UN) Sustainable Development Goals. She has consulted with companies such as Lockheed Martin Space Systems, Fort Wayne Metals Research Products Corporation, and S&P Global. Dr. Gurgur is Area Editor for Data Technologies and Analytics in *Data & Policy*. Her research has been published in major journals such as *Naval Research Logistics*, *The Engineering Economist*, *Journal on Applied Analytics*, *Journal of the Operational Research Society*, *Renewable Energy: An International Journal*, and *International Journal of Energy Sector Management*. Previously, she was on the faculty at Colorado School of Mines and held an NSF fellowship in environmentally benign manufacturing. Dr. Gurgur is member of international academic communities, including European Operational Research and System Dynamics Societies. She earned a master's of science degree in management science from the University of Warwick, UK, through a British Council scholarship; a master's of science degree in applied and mathematical statistics from Rutgers University; and a PhD in industrial and systems engineering from Rutgers University. She can be reached at cgurgur@purdue.edu.

A KNOWLEDGE GRAPH APPROACH TO SATISFYING REGIONAL WORKFORCE EDUCATION NEEDS



Author

George F. Hurlburt

Regional centers (RCs) of major university systems typically lack the necessary accreditation to create courses to satisfy local workforce trends. Instead, they must rely on courses developed elsewhere under proper academic oversight, and it's difficult to attract such courses for myriad reasons. To help solve this dilemma, a knowledge graph (KG) was created to assist a new RC in the course-selection process.

Rather than concentrating on employers and their known unmet needs, this KG examined already-established educational pathways in the region. The results highlight numerous educational drivers that clearly relate to regional workforce needs while showing the subtle differences among the strategies employed and revealing workforce needs among differing counties in the region.

The KG vividly shows numerous richly attributed educational pathways from K-12 to community college (CC) coursework. Each well-defined pathway, complete with collegiate credit offerings, certifications, and valuable career linkages, demonstrates numerous established articulation agreements between universities and CCs.

These agreements include a broad range of upper-level undergraduate offerings with real potential that correspond well with the known workforce needs in the region. The KG is now entering a new phase with further emphasis on extracurricular programs, internships, and apprenticeships to further reinforce the region's main education drivers.

RCs ARE VALUABLE WORKFORCE INTERMEDIARIES

An RC's nature is largely defined by the community in which it's placed. Most often, a university system's RC supports a regional CC focused on responding to local economic development needs. These needs often generate requirements for advanced technical education and frequently lead to full-degree programs. Moreover, the quantity and quality of students feed both workforce expansion and program development. This, in concert with strong industrial support, position RCs as valuable workforce intermediaries.[1]

Higher education RCs don't have the authority to act as a college or university, instead relying on fully accredited courses from their respective state university systems or other universities. RCs typically work well with established CCs. This means the RC undergraduate focus typically involves multiple upper-division course partnerships. The development and cultivation of these partnerships is the predominant challenge, not the lack of accreditation.

This article reports on a KG created for a new RC that is uniquely placed in a thriving US technological corridor dedicated to excellence. The notion of autonomous systems, including unmanned vehicles, has a keen economic development interest in this area. This is reinforced by a new RC building dedicated to community cohesion, education, and state-of-the-art research in autonomous vehicles, setting the RC apart from its more traditional counterparts. This emphasis on research is important, as it focuses on science, technology, engineering, and mathematics (STEM), an important aspect of the region's existing educational framework.

Proximity to a robust CC spanning the three counties that define the region rounds out the equation for a successful RC. The diversity of offerings at the regional CC ensures that regional educational needs can be met, extending well beyond those of the STEM community.

The new RC faces the dilemma of defining manageable educational pathways that: (1) support regional needs with a qualified locally grown and diversified workforce, and (2) provide real opportunity for regional industrial expansion.

An initial survey proved useful from an academic's professional standpoint but was unwieldy for use in a more broadly based population. Hiring statistics and future job projections served to satisfy immediate needs but lacked sufficient reliability in the longer empirical view when viewed by themselves. This is largely due to a volatile economy further beset by pandemic-induced stress. Other indicators needed to be brought to bear to reinforce the more generalized projections with tangible regional roots. For that, the new RC turned to harvesting reliable regional educational data as reinforced via a KG.

CREATING KG PATHWAY REPRESENTATION

KGs are valuable instruments when studying complex adaptive systems. Unlike relational database management systems, which require elaborate inflexible schemas, KGs rely on fluid relationships that can come and go over time without major analytic sustainability overhauls.[2]

A KG relies on the notion of triples in a subject-predicate-object relationship structure (e.g., "regional center offers electrical engineering"). Spanning these triples allows the development of logic chains that provide traceable pathways, often indirectly linking cause to effect across many nodes and frequently offering multiple alternative paths (predicates) between nodes (subjects and objects).

The RC KG shown in Figures 1-5 drew on readily available public data from the three regional high school (HS) systems, the tri-county CC, the RC itself, and other related national higher education resources. It was built in Neo4j graph database version 1.4.15 using the Cypher query language in the 4.4.6 browser. By design, it permits the construction and visualization of useful graph pathways for existing educational programs of study within the tri-county region.

KG research is emerging from relative infancy. It extends to graph algorithms of all kinds as well as applied graph theoretical mathematics. Advanced KG research deals with the KG for knowledge-aware applications, knowledge acquisition and temporal KGs, and knowledge representation and learning (KRL). The current focus on KRL involves building KG frameworks conducive to applied artificial intelligence and, more specifically, machine learning.[3]

KGs provide vivid, meaningful visualizations. Although many visualization formalisms and frameworks exist, the age-old art versus science argument is likely applicable.[4] Central to both art and science, however, is context, which adds the spice of domain-driven diversification and semantic relevance to the argumentative mix.

Context is often cited as an essential element to establish KG credibility. Some researchers go so far as to perform extensive Web crawls to develop added triples that lend temporal, spatial, and other contextual value to established triples of interest.[5] Others rely on direct representation, reification, higher arity[6] representation, and annotations. Ontological-based schemas, of course, provide more precise semantic relationships.[7]

The RC KG was built using direct representation. Several intentional design steps were taken to ensure context was appropriately addressed. Arcs, the predicates, were limited in number and carefully controlled to ensure relevance. Nodes, the subjects and object entities, were also intentionally kept unambiguous. Where appropriate, triples were added to establish spatial location to the county level within the region. Aggregations were then managed by query.

Each program of study was attributed with several specific properties. Among these properties, enrollment numbers provided valuable quantitative information as to the true viability of potential pathways. When available, feeder courses were also listed.

Educational pathways were defined by progressive programs of study. Program-of-study properties were reinforced by additional related triples showing who is engaged in what and the outcomes of these engagements. The resulting RC KG is, it is hoped, accurate and reflective of existing cultural values of the region served. It is intended to offer

a robust backdrop from which future upper-level coursework may be defined based on established regional educational patterns, adding relevance to otherwise speculative future job predictions. The resulting schema appears in Figure 1.

The ideal pathway extends from public schools to a specific higher-education degree. For example, the pathway for electrical engineering (EE) has its roots in the public schools. Some regional schools adopt nationally accepted programs of study for their curricula. The RC KG defines specific regional HS programs of study, many of which lead to college credit. This credit includes the regional CC, which offers an engineering program leading to an associate of science (AS) degree. The combination of college credits at the HS level and accreditation programs between the CC and four-year higher educational institutions were taken into consideration in constructing what was considered a viable upper-division pathway.

The CC pre-EE offerings are accompanied by internships at the leading technology employer in the region. Working through the school of engineering at the state level, with whom the CC has an articulation agreement, the RC offers the necessary upper-level coursework leading to a bachelor's degree in EE. In addition to ongoing internships, EE graduates are guaranteed entry-level positions with that employer. The pathway graph in Figure 2 captures all these STEM-based elements and their key relationships.

The new RC can currently demonstrate 10 such operational pathways in five regionally relevant programs of study. Some involve state institutions; others engage universities and colleges outside the state's educational ecosystem. Because the RC is unable to develop or accredit its own coursework, these external relationships are crucial.

One key workforce driver involves strategic academic partnerships designed to bring strategic economic development to bear in the region. The initial RC KG focuses on the academic side of this equation, which is intended to extend to industrial and governmental program partnerships as well. The academic KG, by itself, has already affirmed some significant insights, not easily recognized without the vivid visualizations made possible by the KG design.

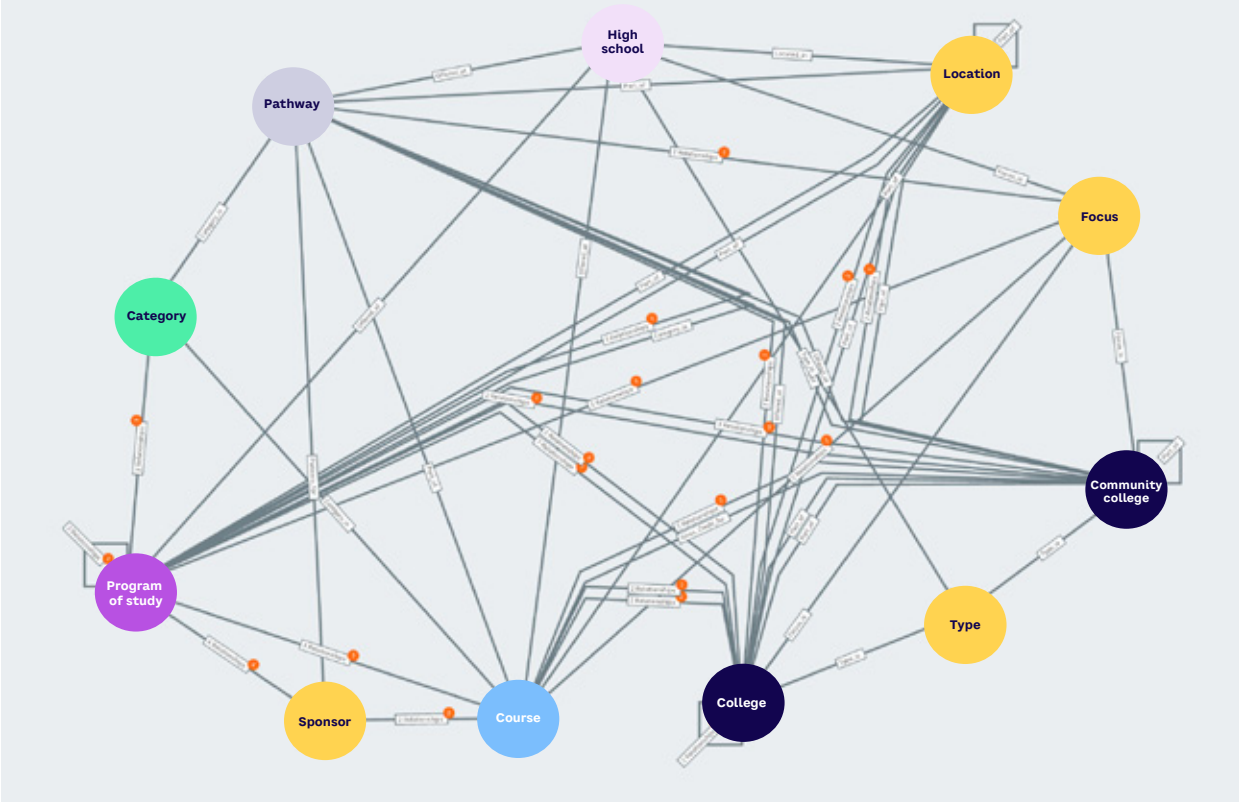


Figure 1. RC KG schema resulting from direct representation

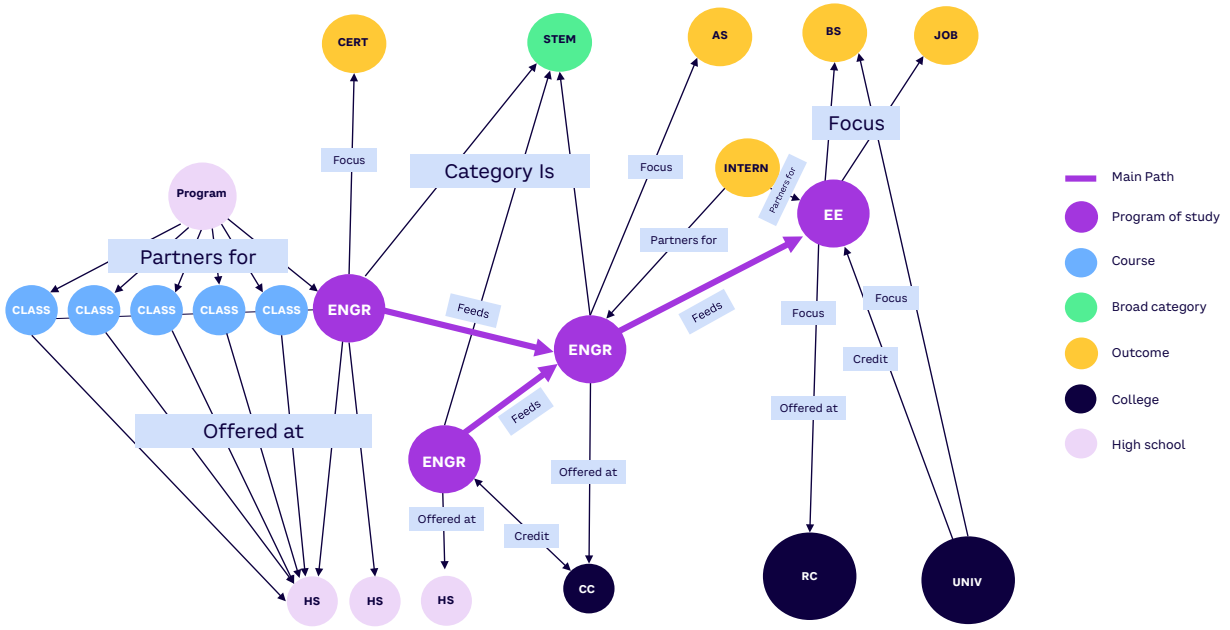


Figure 2. Graph representation of the full regional pathway to a BS in EE at the RC

ACADEMIC KG INSIGHTS

The RC KG clearly shows five key academic relationships and trends.

1. EDUCATIONAL DISCIPLINES

The educational focus of a given region speaks directly to the region's implicit values. Figure 3 depicts comprehensive regional HS and college concentrations, which are often completed with courses supporting prescribed programs of study and major concentrations of study. These include STEM, business, healthcare, and education, each of which houses programs of study.

We can see that the region places a high value on STEM education. Given the regional focus on autonomous systems and information technology, this is hardly shocking. The emphasis on business is also not a surprise, as both the military and the federal government are major international buyers in the region. In light of national and regional healthcare demands, the strong emphasis on healthcare doesn't seem out of place, either. The large concentration of trades, construction, communications, justice, and services reflect levels of workforce diversification, especially for pathways in which specific technical knowledge and skills prevail.

However, the reduced concentration on education programs of study throughout the region is disappointing. Fortunately, the region has a relatively high concentration of trained teachers, which leads to greater quality of instruction. Unfortunately, the pathways to teaching are not as strong as other pathways in the region. This is reflective of the national shortfall of qualified teachers, further fueled by COVID-19 burnout.[8]

In each case, the graph database underlying the RC KG captures actual enrollment data, giving a basis for sound quantitative trend analysis. This is most useful in constructing potential pathways that can yield reasonably sized cadres.

2. CERTIFICATIONS

The KG revealed some 133 named certifications related to the concentrations within each field of study. Of these, 38 were unique, career-enhancing certifications awarded by the CC and other educational institutions. The remaining 96 were awarded by recognized professional organizations. Should the notional trend toward stackable certificates underpinning a degree come to full fruition, this data, quantitatively related to the concentrations within the more general fields of study, will prove most useful. These certifications also represent an opportunity for analysis of certification opportunities by concentration to reinforce desirable workforce competencies where workforce training is applicable.

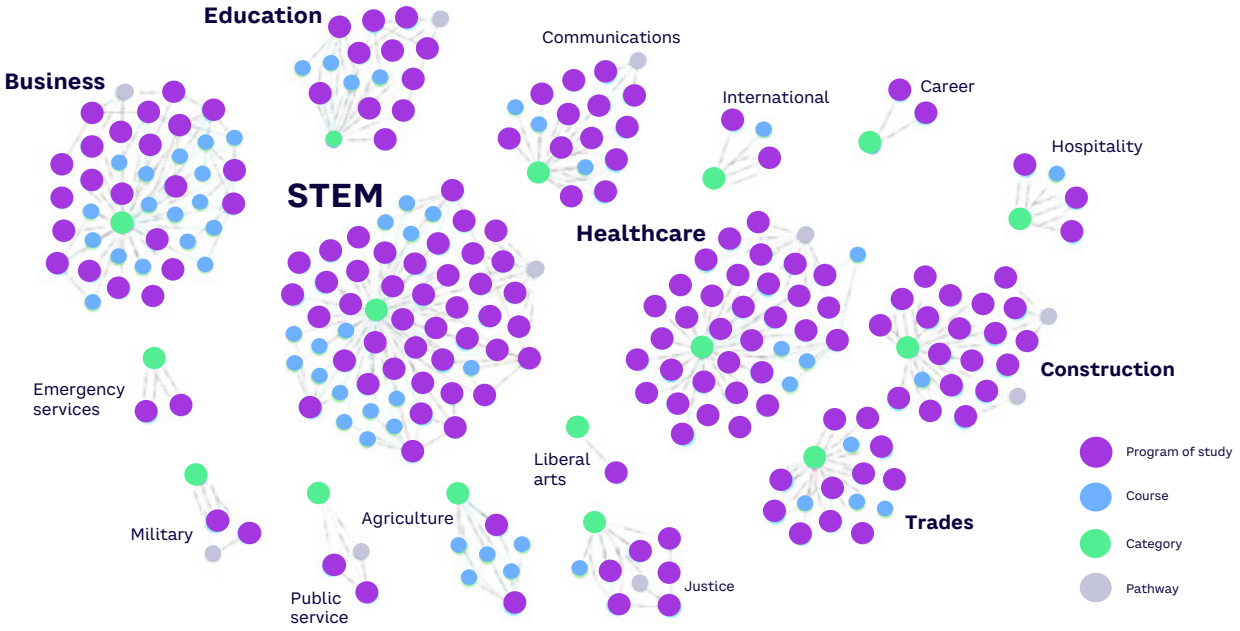


Figure 3. Aggregation of educational concentration areas in the region

3. CREDITS

The graph database can capture and visually depict both individual HS courses and the broader concentrations for which external credit may be earned. This includes work-study programs, earned credit programs, and concurrent education offerings. It is significant in that it relates directly to the viability of potential pipelines.

Where offered, such credits provide the incentive to pursue a given program of study. Likewise, where offered, internships may also be represented and depicted. Both appear as valuable components of Figure 2. The left side of Figure 4 shows the available HS credit offerings from which pathways may be derived. These credits represent specific and academically valuable pathway building blocks.

4. ARTICULATION AGREEMENTS

The right side of Figure 4 depicts the articulation agreements that exist between the CC and its partner four-year institutions, including those within the state's educational ecosystem. The concentrations and their links represent the 10 existing pathways mentioned earlier that are currently available at the new RC. The remaining institutional affiliations show articulation agreements that represent potential pathways available via the regional college and the CC juggernaut. Indeed, enrollment numbers will vary and truly represent regional appetite for a given program of study. Thus, actual enrollments will dictate ultimate pathway viability. Visual representation of the potential pathways, combined with enrollment data, represents a significant point of departure for further academic partnership exploration.

5. STRATEGIC VARIANCE

The RC exists to support the entire region, but it is essential to appreciate that the three county school districts each is responsive to its county constituencies and localized values. This shows up in how credits are managed and what concentrations are emphasized among the three jurisdictions.

Figure 5 shows the differences between the three regional county school districts, with increasing density from right to left. There is no vertical correlation to individual county school systems. Appreciation of these salient differences is crucial to building a well-balanced set of pathways from which useful cadres may be recruited. Here, the KG shines in defining realities without imposing value judgment.

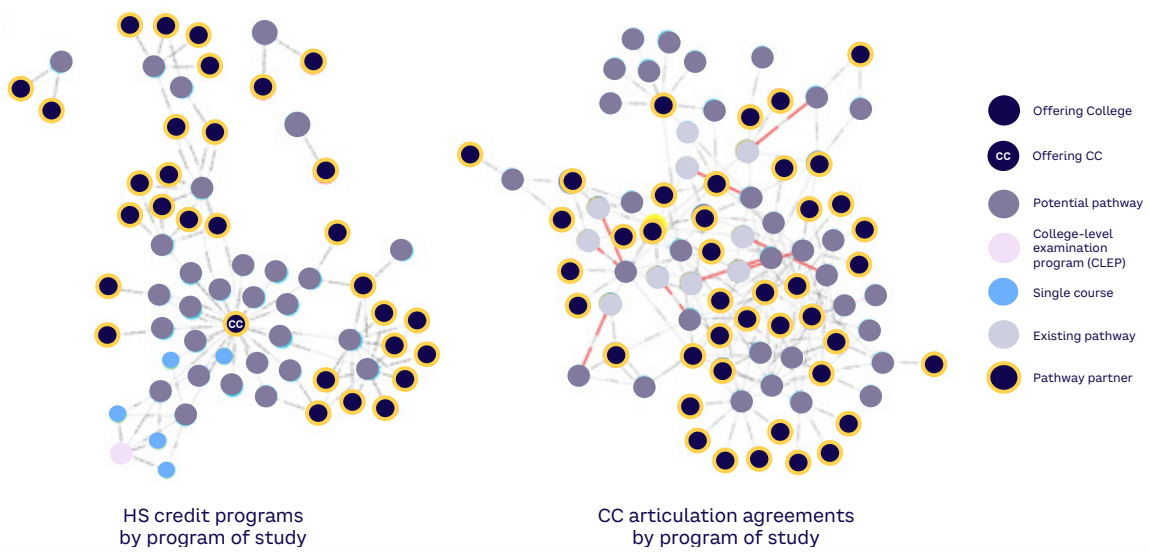


Figure 4. HS college credits and CC articulation agreements as pathway building blocks

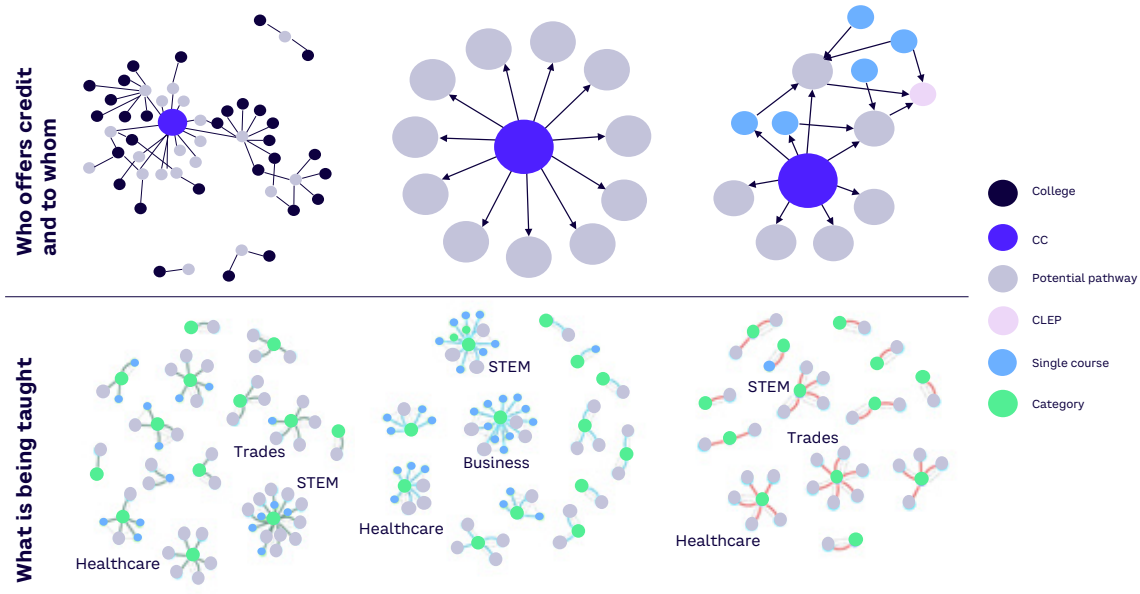


Figure 5. Volumetric variation among strategic approaches within three regional county school districts

CONCLUSION & FUTURE DIRECTION

The current RC KG provides a useful analytical framework for analysis in pursuit of viable regional pathways. It will reinforce job predictions with a real-world view of potential quantitative readiness to meet demands. It also holds some promise in increasing the probability of building strong cadres by selected programs of study.

However, the KG is but one tool in the academic toolkit, as it is currently far more descriptive than prescriptive. As a work in progress, the KG met with some initial acceptance, but time and further development will determine its ultimate utility in the decision-making process.

Although the academic aspect of the workforce equation is essential to future academic partnerships, this KG cannot stand alone. Industrial and governmental partnerships are equally important in building a balanced regional workforce attuned to its own best interests and needs.

To that end, the next step is to grow the KG to reflect regional industrial and governmental programs extending well beyond internships and

apprenticeships that serve to reinforce regional workforce development. This partnership data already exists and, when incorporated, should be instrumental in further identifying both gaps and opportunities to help build industrial and governmental partnerships. Such vital partnerships will serve to grow a vibrant and diversified workforce that truly supports the regional culture.

REFERENCES

[1] Kerrigan, Monica Reid, et al. "ATE Regional Centers: CCRC Final Report." Community College Research Center, Columbia University, May 2007.

[2] Kejriwal, Mayank, Craig A. Knoblock, and Pedro Szekely. *Knowledge Graphs: Fundamentals, Techniques, and Applications*. MIT Press, 2021.

[3] Ji, Shaoxiong, et al. "A Survey on Knowledge Graphs: Representation, Acquisition, and Applications." *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 33, No. 2, 26 April 2021.

[4] Antoniazzi, Francesco, and Fabio Viola. "RDF Graph Visualization Tools: A Survey." *Proceedings of the 23rd Conference of Open Innovations Association (FRUCT)*. IEEE, 27 December 2018.

[5] Dörpinghaus, Jens, and Andreas Stefan. "Knowledge Extraction and Applications Utilizing Context Data in Knowledge Graphs." *Proceedings of the 2019 Federated Conference on Computer Science and Information Systems (FedCSIS)*. IEEE, 7 October 2019.

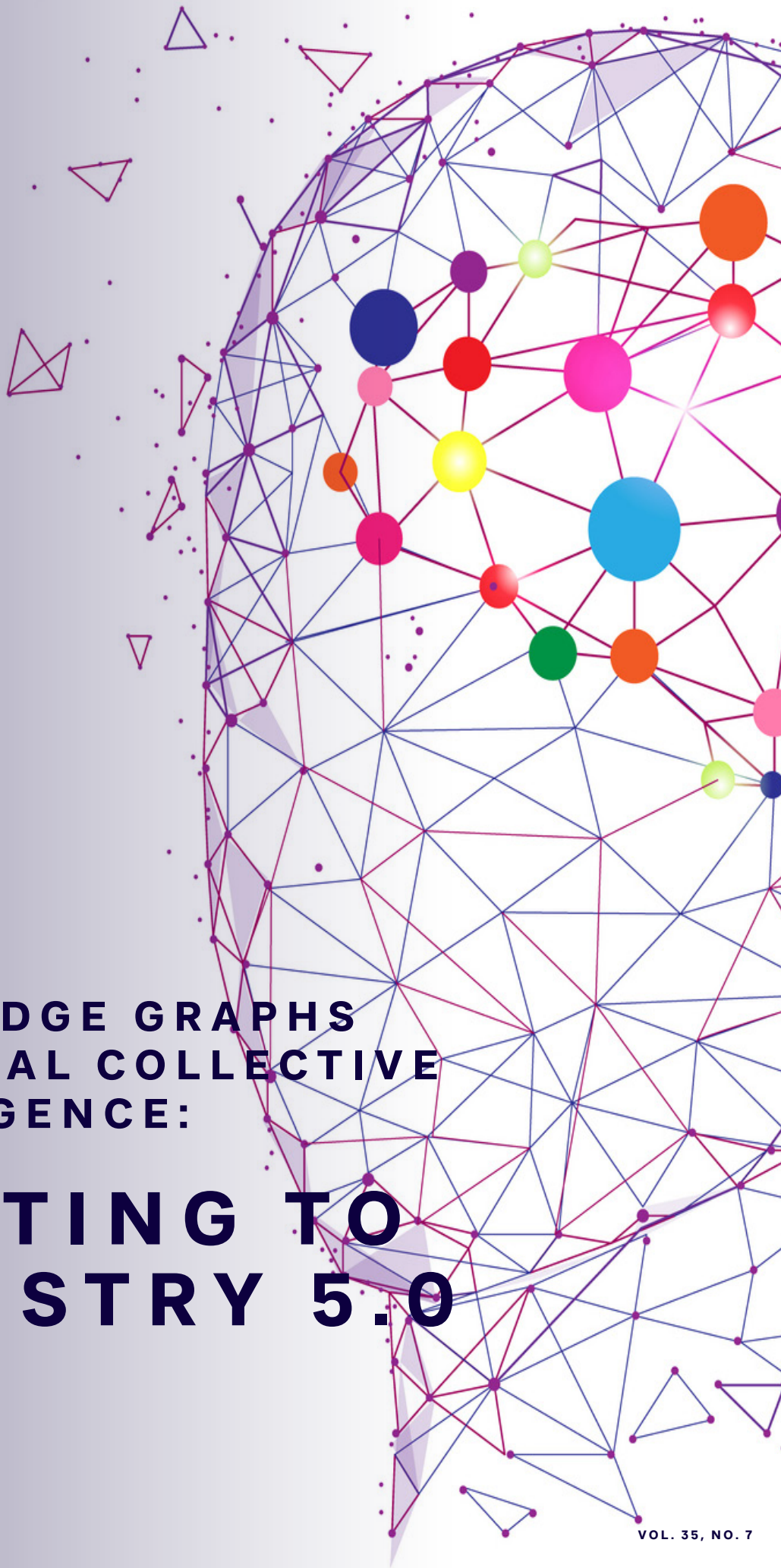
[6] Zapata-Carratalá, Carlos, and Xerxes D. Arsiwalla. "An Invitation to Higher Arity Science." Cornell University, 21 January 2022.

[7] Hogan, Aidan, et al. *Knowledge Graphs*. Springer, 2021.

[8] Barnes, Adam. "Here's What's Driving the Nationwide Teacher Shortage." *The Hill*, 21 April 2022.

About the author

George F. Hurlburt is Chief Scientist at STEMCorp. He serves on the board of advisors for a state regional center and sits on the *IEEE IT Professional* editorial board as Associate Editor for Departments and Columns. Mr. Hurlburt retired with a Meritorious Civilian Service Award after 38 years as a Navy Senior Systems Analyst, where he pioneered collaborative network architectures for the US Department of Defense and ran a network for its test and evaluation community. Mr. Hurlburt developed dynamic system architecture designs that took real-world state changes into serious consideration. Active in his local community, he also remains engaged in architecture work through his affiliation with STEMCorp. Mr. Hurlburt's work focuses on applied network science and complexity engineering. His aim is to create computational solutions that better illustrate nonlinear system behavior. Mr. Hurlburt specializes in operational architecture. He can be reached at ghurlburt@change-index.com.



KNOWLEDGE GRAPHS & GENERAL COLLECTIVE INTELLIGENCE:

SHIFTING TO INDUSTRY 5.0

Author

Andy E. Williams

The digital revolution has likely driven the single greatest transformation in the history of human civilization, but it might pale in comparison to the next great transformation: human-centric knowledge graphs (KGs) and functional computing approaches like general collective intelligence (GCI) that leverage such graphs.[1]

The digital revolution arose from the discovery that problems and solutions could be modeled digitally (i.e., in terms of ones and zeros) and from mass production of the transistors required to compute solutions to digital problems at exponentially greater speed and scale. But computer software or hardware created to solve one problem must be reengineered by people to solve another problem. Unlike the human brain, which has general problem-solving ability, computer hardware has only narrow problem-solving ability.

Human-centric functional modeling (HCFM) is a way to allow computers to solve general problems.[2] HCFM represents problems via constructs called "functional state spaces." These hypothetical functional state spaces are special types of KGs used to model systems. Functional state spaces are required for GCI and are of unprecedented importance if, as predicted, GCI can exponentially increase our capacity to understand systems.[3]

For example, GCI has the potential to automatically reapply existing solutions to an exponentially greater number of different problems without any additional programming.[4] Furthermore, the exponential increase in general problem-solving ability predicted to be possible through GCI applies to every process from design to recycling for every product or service that can be modeled with functional state spaces.

Taken a step further, GCI might be required to solve several key societal challenges corporations face. For example, humans can't reliably discern social good — we tend to discern whatever matches the ideology to which our cognitive bias is predisposed.[5] This may be why our current, non-GCI corporate environmental, social, and governance (ESG) programs have had limited success.[6]

Additionally, as technology advances, a phenomenon called the "technology gravity well" is expected to cause decision making to prioritize the interests of an ever-decreasing minority of individuals and businesses at the expense of achieving collective social good.[7]

TAKEN A STEP FURTHER, GCI MIGHT BE REQUIRED TO SOLVE SEVERAL KEY SOCIETAL CHALLENGES CORPORATIONS FACE

It can take a long time for groups to understand when an individual is about to make (or has made) a decision that serves the group poorly (because of the sheer volume of decisions made by individuals within any large group and the often intractable number of potential interactions between those decisions). Looking out only for oneself is much easier, and those doing so might see benefit quickly.

Therefore, any force that acts to continually centralize decision making to an ever-decreasing minority of individuals and businesses would be expected to work to serve the interests of that minority far faster than collectively optimal choices can be understood and made by any group: a technology gravity well.

Some potential impacts of GCI (and therefore impacts of using HCFM to define KGs that are functional state spaces as required by GCI) relevant to this issue of *Amplify* on knowledge graphs are summarized in Table 1.

MOVING TO INDUSTRY 5.0

Industry 4.0 refers to the transition to a world in which there is pervasive integration of manufacturing equipment and other physical systems with digital computing (cyber systems).

Industry 5.0 is defined here as the transition to a world in which far greater integration is possible through the use of HCFM to define KGs (functional state spaces) capable of providing complete semantic models of systems.

Assuming that human internal representations of anything that can be perceived with the physical senses as well as any emotion, thought, and conscious awareness can be defined in terms of functional state spaces through HCFM, then those spaces can be used to represent every physical object that can be seen and every idea that can be conceived. It follows, then, that HCFM potentially applies to every discipline from physics and mathematics to biology, psychology, computer science, and perhaps all others.[8]

In the design of products and services, the use of GCI implies the ability to explore a vastly larger region in the space of possible design configurations. This means exploring all possible permutations of all possible components and combinations of components. Rather than advancing through known research innovations, such design processes would mimic nature's process of designing living things: incorporating any change that increases the likelihood of achieving an objective, even when the mechanisms by which that increase was achieved remain unknown.

In manufacturing, modeling physical products in terms of functional state spaces implies the ability to accommodate manufacturing processes too complex to be understood by any individual process designer in an effort to achieve competitive advantage.

In recycling, modeling products and services in terms of functional state spaces and the use of GCI might enable sustainability solutions that are impossible otherwise, such as radically reducing greenhouse gases through an exponential reduction in consumption. This could result from an ecosystem of GCI-based products that cooperate to become more durable, reusable, and recyclable than could be accommodated by any business model today.

A PEEK INTO THE FUTURE

It might seem like HCFM and GCI could be used to improve any product or service. However, research suggests that managers can't lead teams effectively when those managers are too smart (i.e., an IQ that is 1.2 standard deviations or higher than that of the group).[9]

So what might be the consequence of deploying a GCI with an IQ billions of points greater than that of the smartest human who has ever lived?[10] One possible outcome is a technology design process in which human contributions come together too quickly and in ways too complex for any human to understand, resulting in technology so complex it can't be reliably distinguishable from magic.

KG-RELATED TOPIC	PREDICTED IMPACT OF HCFM AND/OR GCI ON TOPIC
What are some real-world/novel applications of KGs?	The most novel potential applications of KGs are as functional state spaces that HCFM predicts might be used to create an artificial general intelligence (AGI) or a GCI.
How are KG applications being used across industries?	GCI applications that leverage functional state spaces as KGs have been conceptualized in a wide variety of industries from healthcare to sustainable housing development, but since GCI requires large-scale collaboration, which in turn requires educating more broadly about this unknown concept, these applications haven't yet been deployed.
What are the business benefits of KGs? What are the challenges/limitations?	Together with GCI, KGs can potentially be used by groups of cooperating businesses to gain unbeatable competitive advantage over any business that competes as an individual entity.
How can business leaders benefit from KGs?	Business leaders are predicted to benefit from functional state spaces as KGs when a critical mass of mindshare has been created about these concepts for sufficient participation in a project to implement GCI.
What is an example KG use case?	One example is about achieving a radical increase in sustainable economic development through reliable patterns for doing so.[1] One such pattern is combining sustainable economic development projects into networks of cooperation that increase their value to the point that such development becomes sustainably self-funding. This would allow social impact to be achieved at the scale at which it is needed globally, rather than at the limited scale at which funding is available, creating the possibility of an exponential increase in capacity to fund social impact.
What are some novel research findings pertaining to KGs?	The most novel research finding pertaining to KGs as functional state spaces used to represent the behavior of systems within HCFM is that GCI might exponentially increase our ability to solve any problem related to understanding or applying a system.
How do KGs differ from current data management technologies?	The technology gravity well effect is predicted to cause the centralization of data management and other processes into the control of the owners of that technology, wherever that centralization is in their interests, resulting in the inability to improve collective outcomes of data management wherever doing so is inconsistent with the interests of those owners. Combined with GCI, KGs may be able to prevent this centralization.
How can KGs be successfully integrated with artificial intelligence (AI)/machine learning (ML)?	KGs can potentially be successfully integrated with AI/ML through modeling AI/ML solutions as paths through a KG that has been defined through HCFM and which is therefore also a functional state space.[2]
What are some examples of state-of-the-art KG-enabled AI?	Every existing AI algorithm and every possible AI algorithm can potentially be represented as a path through a single, universal KG. Modeling problems in terms of the lack of a path from one point in this KG to another creates the possibility of enabling any AI algorithm to reuse any other AI algorithm. Until the implementation of a KG with the complete properties of a functional state space, this can only be approximated.
What are the implications of KGs on AI systems?	Modeling AI/ML solutions as paths through a KG, together with GCI, can potentially drive an exponential increase in capacity to reuse AI/ML solutions without reprogramming or retraining.[3]
How can KGs address the issue of explainable AI?	As KGs, functional state spaces constitute a revolution in explainable AI because they can represent all possible models of type 1 (intuitive) reasoning that might be implemented by AI/ML to solve uncomputable problems through pattern detection while also representing all possible models of type 2 (rational methodical) reasoning that might be implemented by procedural programs to solve computable problems.
What are some KG applications for social good?	As KGs, functional state spaces might drive an exponential increase in impact on social good (see examples KG use case above in this table), which would be revolutionary.[4]
What role do KGs play in gaining deeper insight into the COVID-19 crisis?	As KGs, functional state spaces potentially make it possible to understand that there are deep-seated cognitive biases that are a fundamental part of virtually all humans,[5] and that without GCI, these biases make it impossible for groups to gain deeper insight into the COVID-19 crisis. With GCI, this insight might reliably be gained.
What is the role of KGs in complexity science?	Functional state spaces suggest an objective definition of complexity.
How can the quality of KGs be assessed and validated?	All implementations of open functional state spaces have common properties. Therefore, any tools capable of assessing and validating the quality of one functional state space might be reused to assess and validate the quality of other functional state spaces. However, such tools have only been explored at a conceptual level and remain to be implemented.
What are the trust, privacy, and security considerations for KGs?	KGs must be accessed through a decentralized system of decision making to avoid issues that result from the technology gravity well naturally removing trust, privacy, and security.

[1] Williams, Andy E. "Breaking Through the Barriers Between Centralized Collective Intelligence and Decentralized General Collective Intelligence to Achieve Transformative Social Impact." *International Journal of Society Systems Science*, forthcoming 2022.
[2] Williams, Andy E. "Defining and Quantifying an Exponential Increase in General Problem-Solving Ability Within Groups." AfricArXiv, 22 February 2022.
[3] Williams (see 2).
[4] Williams, Andy E. "Increasing the Societal Impact of Science, Technology, Engineering, and Math with General Collective Intelligence." AfricArXiv, 2 March 2022.
[5] Williams, Andy E. "Innate Collective Intelligence and the Collective Social Brain Hypothesis." PsyArXiv, 26 May 2022.

Table 1. The predicted impacts of using HCFM to define KGs that are functional state spaces as required by GCI

GCI-based technology will likely also be different in how it's used. Rather than having to learn to operate such technology, the technology might learn what we are trying to do and self-assemble from available components to accomplish our goals to the optimal degree, removing the need for any specialized tools or expertise.

Although hard to envisage, this implies GCI might make individuals who are novices much more effective at specialist tasks like product design than the most gifted designers today and allow people with no medical training to perform surgeries and other medical interventions that the most gifted of today's doctors would consider miraculous.

Just like the revolutionary digital technology that came before it, through simple geometric arguments in conceptual space, the case can be made that GCI is the most important technological development in the history of human civilization with regard to problems that can be modeled in terms of functional state spaces, which potentially includes all problems.

> **BECAUSE GCI CREATES POTENTIALLY UNBEATABLE COMPETITIVE ADVANTAGE, COMPANIES THAT FIGHT GCI WOULD MOST LIKELY GO EXTINCT**

GCI might be as profoundly important as this (seemingly preposterous) claim, or it might prove impossible. However, observation of natural systems such as our own human organism suggests that adaptive problem-solving systems based on functional state spaces such as GCI are a real possibility. Nature has already created such solutions, and they have proven successful for hundreds of millions of years.

ESCAPING THE TECHNOLOGY GRAVITY WELL

The emergence of GCI isn't a certainty. An analysis based on HCFM predicts that any civilization might go one of two ways. The first is to develop a mechanism for *individual* optimization, a necessarily centralized process that eventually might exponentially increase our ability to solve problems for companies or other individual entities. The second is to develop a mechanism for *collective* optimization, a necessarily decentralized and distributed process that eventually might exponentially increase the ability to solve problems for all.

The first option implies a civilization that will fall deeper into the technology gravity well toward the emergence of AGI, which is predicted to act as an exponentially more powerful system of individual optimization that makes a system of collective optimization like GCI impossible.[11]

Since this fall into the technology gravity well is likely to be accompanied by the removal of protections against abuse while radically increasing the ability for corporations, governments, and other entities to be abusive, this suggests unprecedented levels of abuse and control on the part of the company that falls to the bottom of the well first.

This would mean a negative outcome for every business except the one at the top of the hierarchy, which would be expected to gain all possible technological advantages to control more revenue than any company that has ever existed.

The other option would be using GCI to *escape* the technology gravity well, resulting in a positive outcome for the majority of businesses (except those that decide to fight this transition rather than embrace the far larger opportunities expected to come with it). Because GCI creates potentially unbeatable competitive advantage, companies that fight GCI would most likely go extinct.

Functional state space is involved in both transitions (to AGI or GCI). Even though functional state spaces and GCI have not yet been fully implemented, if it's true that they have the greatest potential for impact on all technologies known today, then it's important they are on every business leader's radar.

Unlike today's databases, which can only store a limited subset of information, functional state space has the potential to model any system and all possible behaviors of that system, potentially storing all possible information about a given system. Thus, a functional state space is a complete semantic model that enables meaning (understanding) rather than just information to be communicated at exponentially greater speed and scale.

But even without any understanding of functional state space, it is possible to use patterns of collectively intelligent cooperation to reliably achieve a radical increase in the ability to solve any problem.[12] These patterns leverage a set of well-defined and generalizable relationships between businesses, their products or services, and other entities, without the need to recognize these relationships as existing in functional state space at all (though representing those relationships this way might allow them to be further generalized to achieve more impact).

The most stunning claim of GCI is that for certain categories of "wicked" problems (like achieving social good in difficult cases), the more we're fixated on solving these problems, the less we are able to do so. Problems too complex to be solved directly through any choices that can be deduced by any individual must be solved indirectly through development of a more powerful distributed problem-solving system (such as GCI) that is capable of discovering far more complex choices that might be capable of radically better outcomes. All problem-solving methods that are not orchestrated by GCI can be considered direct (in that choices are generated by individuals) and centralized (in that these individuals can't be prevented from prioritizing their own interests). This is problematic because no direct approach can reliably solve wicked problems (those currently considered not solvable).[13] This is supported by the fact that no approach has reliably created durable solutions to these problems at any time in the history of human civilization. That means focusing energy toward any efforts other than GCI is the best way to *not* solve the world's most challenging problems.

This is counterintuitive, since it would mean people who want to solve complex problems of social good might be the ones ensuring that the most pressing problems of social good cannot be solved. That is, when these individuals believe they know the solution, they don't ensure that they or someone else diligently explores whether or not it is feasible to achieve an exponential increase in impact on social good through modeling problems and solutions in terms of functional state spaces/knowledge graphs together with the use of GCI. In other words, their caring ensures those problems can't be solved.

Opposition from such well-meaning individuals is thus an important consideration when attempting to launch any GCI-based initiative if it is true that people interested in social good are predisposed to having cognitive biases toward type 1 reasoning (intuitive), preventing them from asking whether or not disruptive new solutions like GCI are needed.

Similarly, the institutions we rely on for coordinating social good globally use intuitive reasoning, making it impossible for them to choose interventions like GCI that are not similar to patterns of interventions in the past. Thus, any plan to achieve a radical impact on social good must consider working outside such institutions.

These innate cognitive biases are also an important factor to consider if, in addition, the type 2 reasoning (rational) that allows individuals to assess radically different solutions is typically not effective at building the mindshare required to build the consortia and attract the resources necessary to implement such an idea.

Any implementation of GCI might bridge these two reasoning types when it becomes available, but implementing GCI is the precise problem we're trying to solve. The only solution might lie in understanding how nature has evolved complex adaptive systems in an iterative way, so that GCI could be implemented incrementally to bridge these reasoning types while enabling the implementation of a larger subset of GCI functionality.

GCI in turn requires implementing KGs that meet the requirements of functional state spaces, a problem that hasn't yet been solved. By informing a variety of stakeholders (especially mathematicians, physicists, computer scientists, and others who study constructs with similar features) about how the combination of GCI and functional state spaces might radically increase our ability to solve every problem in general, it might be possible to inspire a collaborative effort to solve this problem as well.

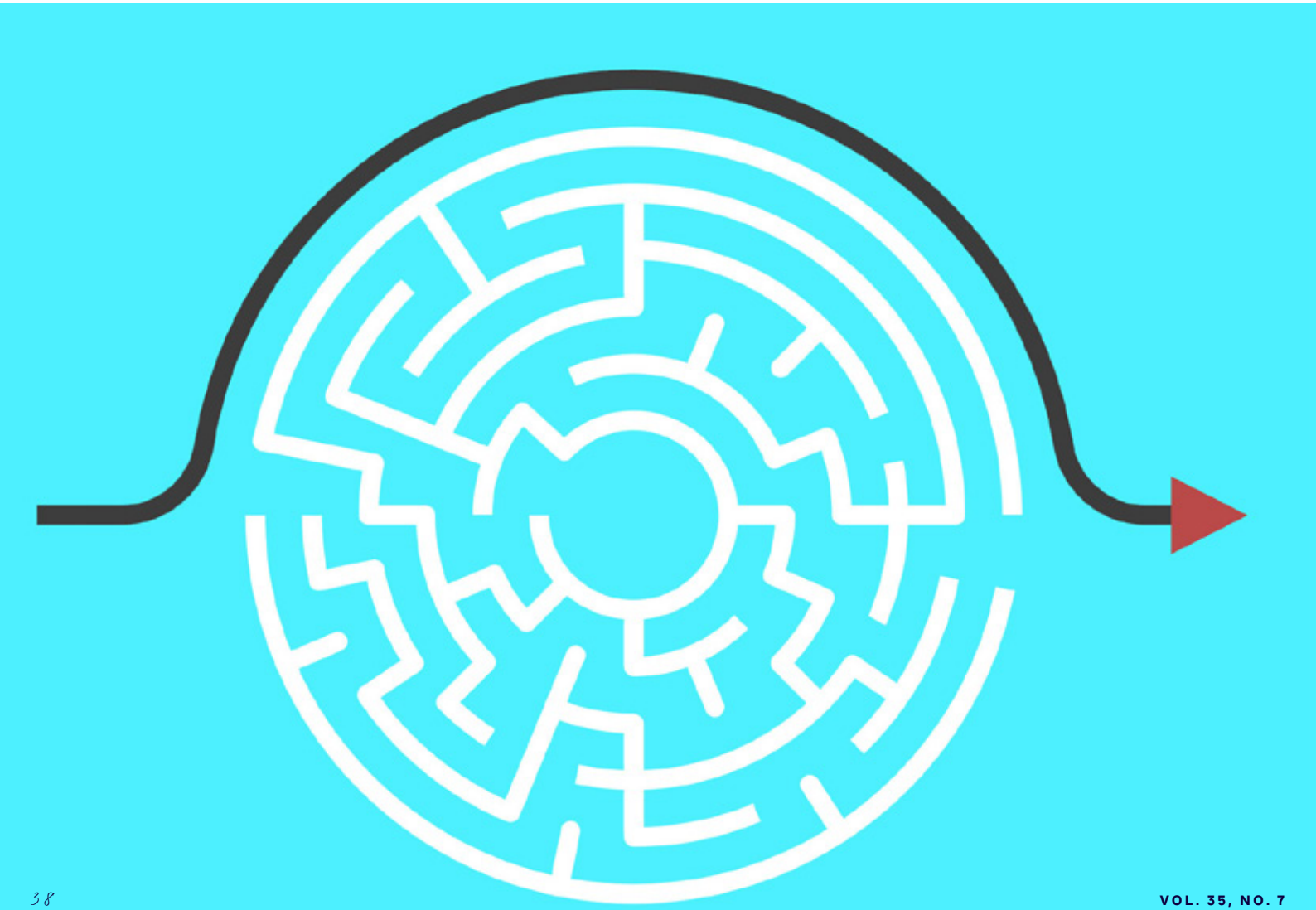
REFERENCES

1 Williams, Andy E. "Defining a Continuum From Individual, to Swarm, to Collective Intelligence, and to General Collective Intelligence." *International Journal of Collaborative Intelligence*, Vol. 2, No. 3, 2 May 2022.

2 Williams, Andy E. "Automating the Process of Generalization." AfricArXiv, 12 March 2022.

3 Williams, Andy E. "Human-Centric Functional Modeling and the Unification of Systems Thinking Approaches: A Short Communication." *Journal of Systems Thinking*, 20 August 2021.

4 Williams, Andy E. "Cognitive Computing and Its Relationship to Computing Methods and Advanced Computing from a Human-Centric Functional Modeling Perspective." *SCRS Conference Proceedings on Intelligent Systems*. SCRS Publications, 21 September 2021.



[5] Williams, Andy E. "Innate Collective Intelligence and the Collective Social Brain Hypothesis." PsyArXiv, 26 May 2022.

[6] Williams, Andy E. "General Collective Intelligence as a Platform for Computational Social Systems." AfricArXiv, 12 March 2022.

[7] Williams, Andy E. "Are Wicked Problems a Lack of General Collective Intelligence?" *AI & Society: Journal of Knowledge, Culture, and Communication*, 4 October 2021.

[8] Williams, Andy E. "Human-Centric Functional Modeling and the Metaverse." *Journal of Metaverse*, Vol. 2, No. 1, 29 April 2022.

[9] Antonakis, J., R.J. House, and D.K. Simonton. "Can Super Smart Leaders Suffer from Too Much of a Good Thing? The Curvilinear Effect of Intelligence on Perceived Leadership Behavior." *Journal of Applied Psychology*, Vol. 102, No. 7, 2017.

[10] Williams, Andy E. "Defining and Quantifying an Exponential Increase in General Problem-Solving Ability Within Groups." AfricArXiv, 22 February 2022.

[11] Williams, Andy E. "Breaking Through the Barriers Between Centralized Collective Intelligence and Decentralized General Collective Intelligence to Achieve Transformative Social Impact." *International Journal of Society Systems Science*, forthcoming 2022.

[12] Williams, Andy E. "Increasing the Societal Impact of Science, Technology, Engineering, and Math with General Collective Intelligence." AfricArXiv, 2 March 2022.

[13] Williams (see 7).

About the author

Andy E. Williams is chairman and CTO of Nobeah Technologies and founder and Executive Director of Nobeah Technologies Foundation. He is an expert in general collective intelligence (GCI) and human-centric functional modeling (HCFM). As a social entrepreneur, Mr. Williams focuses on understanding the equations underlying human challenges, so that by changing those equations, the problems solve themselves. His research on how to overcome the problem of decision-making systems that don't reliably select the best solutions led to the development of a GCI model, a system that organizes individuals or intelligent agents to create the potential for exponentially greater general problem-solving ability. The theory behind this model suggests that entire classes of problems can't reliably be solved without a GCI or equivalent system due to inherent properties of "collective optimization" problems in which a collective outcome is optimized by optimizing outcomes for each individual agent. Without a decentralized mechanism for self-assembling participating agents to execute a self-organized process, processes and participants are free to become aligned with the goals of a single agent or subset of agents that act as a centralized process owner. Mr. Williams earned a bachelor's of science degree in physics from the University of Toronto. He can be reached at awilliams@nobeahfoundation.org.

KNOWLEDGE GRAPHS IN ENGINEERING: A NEW PERSPECTIVE

Authors

Michael Eiden, Philippe Monnot,
and Armand Rotaru

In recent years, closely related terms "artificial intelligence" (AI) and "machine learning" (ML) have become staples of corporate jargon. As management consultants, we have noticed that many customers have an incorrect or incomplete understanding of these buzzwords, including when and how to apply the concepts and recognizing their inherent limitations. Such behavior is an expected consequence of the accelerating adoption and integration of data-driven approaches to business processes.

As discussed in our *Amplify* article last year, several key factors drive the success or failure of an ML project.[1] Having access to quality data in sufficient quantity is critical, but this aspect is commonly overlooked/underestimated by the decision makers leading these endeavors. Such misfocus is due to the significant hype around ML algorithms. The press (and technical literature) invariably present notable achievements of new, sophisticated algorithms without describing the data that powers the algorithms to allow them to reach unprecedented levels of performance. Consequently, for most companies wanting to venture into the world of AI, this misplaced focus means that building a team of talented individuals to work on developing fancy new algorithms for solving unique business problems will be costly and unlikely to render a positive ROI.

A more effective and productive option is to focus on formulating the problem correctly, building the appropriate infrastructure that allows you to gather informative and unbiased data, and using state-of-the-art algorithms.

Historically, the default option for storing and retrieving data has been relational databases, which represent data in a tabular format. Recently, however, many companies have begun migrating to knowledge graphs (KGs), an alternative solution for representing and querying data.

The good news in this shift? Building a graph is as easy as connecting dots with lines.

GRAPHS IN A NUTSHELL

The geometric nature of graphs makes them intuitively accessible. In their simplest form, graph theory describes them as "networks of dots and lines"[2] — meaning they can be intuitively represented with drawings. Most of us have drawn a graph at least once in our lives. Who among us has never written ideas on a whiteboard or piece of paper and then connected them together (if you haven't, you've probably at least watched TV detectives do that to catch a criminal)?

BUILDING A GRAPH IS AS EASY AS CONNECTING DOTS WITH LINES

Leaving behind the formalities and strict language of graph theory, a graph is composed of nodes (dots) and relationships (lines) that connect them. Practically speaking, nodes represent entities: things or concepts that can be described by a set of properties.

Let's jump right in with a simple example of a labeled graph, the most common type. Figure 1 shows various entities in the context of an organizational diagram. Employees, managers, and the company are shown as circles (nodes).

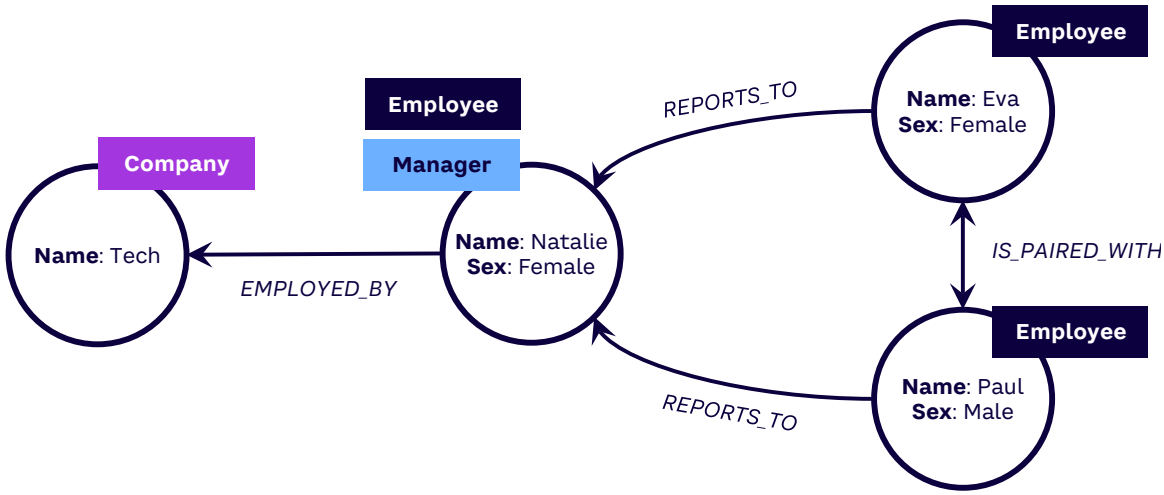


Figure 1. Information about company employees and departments, represented in KG format (source: Arthur D. Little)

Each node contains one or more properties. For example, the bottom-right node has an Employee label, with Paul and Male as properties that describe it. Natalie is also an Employee, as well as a Manager and a Female. Note that nodes can have multiple labels. Relationships between nodes are named and directed, meaning that they have a start node and an end node. Eva, an Employee, has Natalie as her Manager. Therefore, a REPORTS_TO relationship links them to each other. Paul and Eva work together as a binome on projects. Thus, two IS_PAIRED_WITH relationships connect them (indicated by a double arrow). Finally, relationships can also have properties.

Figure 2 shows how the organization graph might be structured if defined in a relational database. One can appreciate how information is duplicated and more difficult to visualize when compared to a graph representation.

Name (PK)	Sex
Natalie	Female
Paul	Male
Eva	Female
Tech	-

Name (FK)	Role
Tech	Company
Natalie	Manager
Eva	Employee
Paul	Employee

Name (FK)	Relationship	Object
Natalie	PART_OF	Tech
Paul	REPORTS_TO	Natalie
Eva	REPORTS_TO	Natalie
Paul	IS_PAIRED_WITH	Eva
Eva	IS_PAIRED_WITH	Paul

Figure 2. Organizational diagram (equivalent to Figure 1) represented through a relational database containing three tables — top-left table contains the name of entities and their properties; top-right table contains the role of each entity; bottom-right table contains how employees are related to each other (source: Arthur D. Little)

GRAPHS IN DAILY LIFE

Given their generality, graphs (or networks) have a variety of applications, ranging from modeling the progression of Alzheimer's disease to finding the optimal route between two cities. Table 1 contains a short selection of some of the most well-known applications.[3]

Recommendation systems, an example in Table 1, are a frequently encountered application of graphs. Such systems track user behavior (e.g., products the user bought or films the user watched) to predict new content the user might be interested in. From a graph perspective, this means inferring relationships of the type "X might be interested in Y," where X is a user (e.g., Jane Doe), and Y is a piece of content (e.g., *Star Wars*). To do so, the recommendation system looks at similarities between users and between pieces of content, measured in terms of shared neighbors (i.e., nodes connected to a given node) and relationship types. For instance, if users Alice and Bob have similar behaviors (e.g., they both watched *The Matrix*, *Dark City*, and *The Crow*), and Alice watched *Equilibrium* (but Bob did not), then the engine should infer that Bob might want to watch *Equilibrium*. Following the same logic, if *Equilibrium* is in the same genre/has similar aesthetics to *The Matrix/Dark City/The Crow*, and Alice watched *The Matrix/Dark City/The Crow* (but not *Equilibrium*), the system should infer that Alice might want to watch *Equilibrium*.

NETWORK TYPE	NODES	RELATIONSHIPS	APPLICATIONS
Social networks (e.g., Facebook)	People	*X is a friend of Y*	Finding the most efficient way of propagating information through the network via the people who are most strongly connected to the rest of the network (e.g., celebrities, experts, community leaders, politicians)
Transportation networks (e.g., street networks)	Locations	*X and Y are directly connected, via a road, airway, waterway, etc.*	Finding the shortest/quickest/least expensive route between two or more locations; optimizing the placement of a given resource, in terms of accessibility
World Wide Web	Web pages	*X includes a link to Y*	Improving Web search results by promoting results/sites that have the most incoming hyperlinks (e.g., the PageRank algorithm)
Recommendation systems	People, products	*X bought/viewed Y*	Generating product recommendations, based on the users' purchase/viewing history (e.g., "Customers who viewed this item also viewed" feature on Amazon, or the "More like this" feature on Netflix)

Table 1. Prominent, real-world applications of graphs (source: Arthur D. Little)

ASSESSING SAFETY THROUGH GRAPHS

As graphs gain in popularity, novel applications in traditional business settings are more likely to come up. Thus, it is crucial to make graphs accessible and part of the standard toolset for AI practitioners. In the remainder of this article, we show how we've used graphs to power a recommendation system that supports independent safety assessments (ISAs) of safety-critical systems.

In highly regulated industries, safety is always at the top of the agenda. Industries like aerospace, railway, and nuclear have had dark track records when it comes to in-service failures, and these failures often lead to significant injuries and/or loss of life. The complexity and cost of the systems, combined with their notoriously long development cycles, make them error-prone, so even small errors can have big consequences. To help mitigate these risks, NASA developed a methodology called "systems engineering," which has been widely adopted:

> Systems engineering … focuses on defining customer needs and required functionality early in the development cycle, documenting requirements, and then proceeding with design synthesis and system validation …. Systems engineering considers both the business and the technical needs of all customers with the goal of providing a quality product that meets the user needs.[4]

When applied early on and at the right level, systems engineering can significantly limit cost overruns by reducing the odds of making ill-formed decisions throughout development. At each stage of this iterative methodology, commonly represented as a V-shaped lifecycle (see Figure 3), key artifacts are systematically generated to ensure traceability, design rationale, documentation, and verification.

Regulatory bodies make ISAs mandatory, with the intention to inspect and review internal processes (e.g., variations of the systems engineering methodology) and the outputs of those processes (e.g., system and software specification, safety analysis, verification activities, and testing evidence).

In official terms, an ISA is defined as "the formation of a judgment, separate and independent

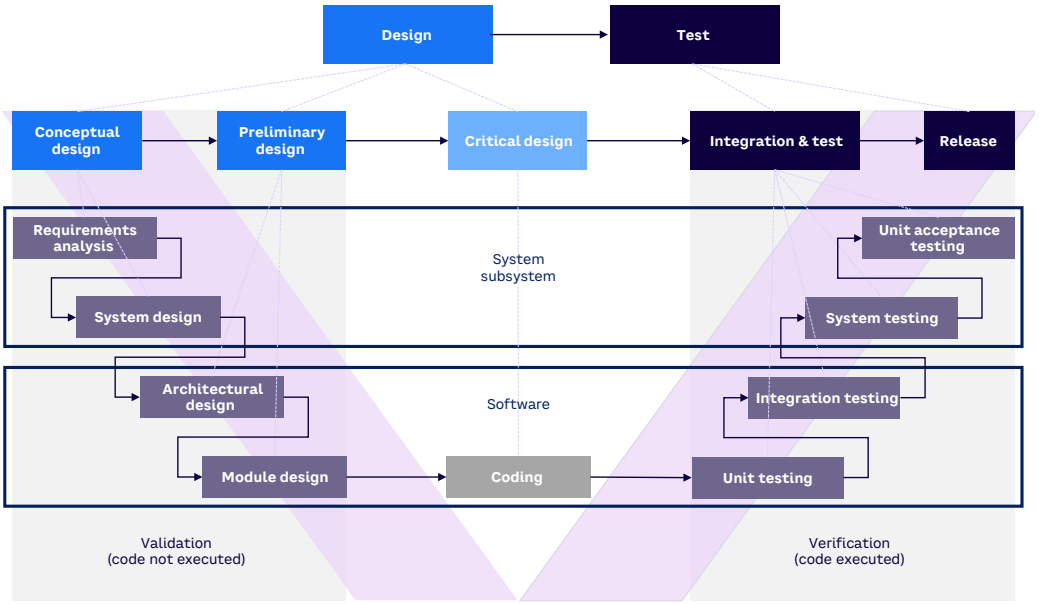


Figure 3. A systems engineering V-model that represents a systems development lifecycle — on the left side and starting at the top, customer requirements are captured and the design is defined with more and more granularity as we progress down the V; on the right side, going up the V, the system is tested at a component, subsystem, and system level to ensure the as-built system is compliant with the as-designed system while meeting the initial customer requirements (source: Arthur D. Little)

from any system design, and development, that the safety requirements for the system are appropriate ... and that the system satisfies those safety requirements."[5] The ISA therefore targets safety-critical systems (software and/or hardware) by auditing the documentation (i.e., artifacts), with the aim of assessing safety, robustness, and completeness.

LIMITATIONS

The inherent complexity of the systems targeted by ISAs means the development and safety demonstration usually relies on a large amount of documentation. The entirety of the documentation supplied can rarely be reviewed in the context of an ISA audit. Therefore, auditors — usually domain experts — manually perform their assessment by randomly sampling the artifacts to gain sufficient confidence in their quality.

Depending on the initial outcome, the auditor might continue to sample the artifacts or follow his or her experience/intuition and target some specific ones. The inherent nature of ISAs, and the context in which an ISA is performed, mean that total confidence cannot be realistically expected as an outcome. A residual safety risk always remains present. In Arthur D. Little's (ADL's) Digital Problem Solving (DPS) practice, we have used KGs and AI to reduce this residual risk and demonstrate how the technology successfully augments traditional ISA approaches.

USE CASE: VERTICAL TRACEABILITY ANALYSIS

DPS partnered with ADL's Risk practice to run a proof-of-concept in parallel to a live ISA audit, aimed at a railway signaling system undergoing a major overhaul. The use case was limited to a single aspect of the auditing process: ensuring the vertical traceability between software requirement specifications (SRSs) and software component specifications (SCSs).

Vertical traceability analysis aims to analyze the various levels of specification of a system. Specifications, depending at which level they sit, can be vague and general (e.g., a customer requirement) or specific and detailed (e.g., a specific behavior that a component must follow). Figure 4 provides a simple specification tree for a generic software system.

The vertical traceability between two layers of specifications is ensured if all three key criteria are met: correctness, completeness, and acceptable refinement (see Table 2). Note that these criteria must be validated both ways — down and up the specification tree.

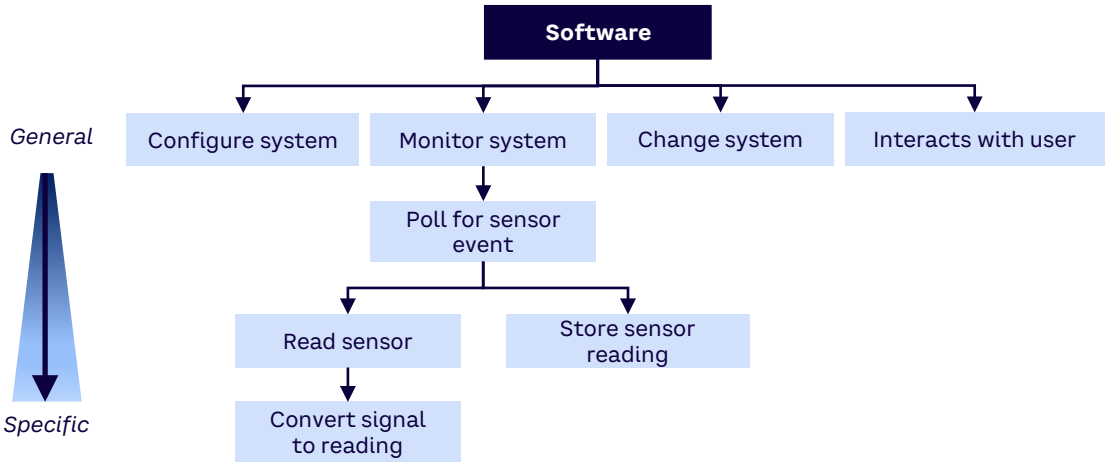


Figure 4. Generic specification tree for a software system (source: Arthur D. Little)

CRITERION	DEFINITION
Correctness	Not contradicting the functional specification
Completeness	Fully implementing the functional behavior of the upper-level specification
Acceptable refinement	(1) No additional behavior at the lower-level specification that cannot be justified as a refinement of the higher-level specification; and (2) an allowed (noncontradictory) elaboration, aligned with the level of abstraction expected for the particular specification level

Table 2. Vertical traceability criteria (source: Arthur D. Little)

The objective was to accurately predict whether or not the vertical traceability analysis of a given SRS would be flagged as a PASS or FAIL by a human ISA auditor. A raised FAIL would mean the ISA auditor believes there's a potential safety issue with a given SRS. The artifacts specific to the live case were provided by the team that had recently completed the ISA. They also provided the outcome of their vertical traceability analysis: whether each SRS was a PASS or FAIL. Out of the 199 SRSs provided, the ISA team flagged 46 as FAIL and 153 as PASS.

METHODOLOGY

Predicting a binary outcome (PASS or FAIL) for each specification was the key objective of this use case. Framing the problem in this manner made it a great candidate for supervised ML. In ML jargon, this would be referred to as a "classification-type" problem. Readers with some exposure to ML will recognize the approach shown in Figure 5, used to develop the model for the task at hand, going from raw data to ISA-specific insights. It shouldn't come as a surprise that Step 2, graphical representation, was added to the typical data engineering and modeling pipeline. Let's now dive into each step of the methodology.

The sections below present an overview of each step while expanding on steps where the graph plays a differentiating part.

ARTIFACT INGESTION & EXTRACTION

As the reader might expect, the documentation received was not stored in a well-structured, queryable database. Instead, it comprised a blend of PDFs, Word documents, spreadsheets, embedded images, and embedded formulas. There were 20,000 individual files. Although this is not uncommon, additional effort was required before any modeling could begin: data had to be hosted, staged, and processed. With the help of natural language processing (NLP) and domain expertise, the processing was done programmatically by only extracting data relevant to the use case. This included all SRSs and SCSs in addition to related specifications, their descriptions, their context, and how they related to each other.

GRAPHICAL REPRESENTATION

The next step is to define an ontology specific to the use case. Ontologies are data models that define what type of entities exist in the domain of interest, the set of properties that describe them, and the relationships that link them.[6]

Creating an ontology is usually time-consuming and requires in-depth domain expertise. In this case, the ontology was implicitly defined and documented through the supplier's artifacts and development process, which closely follows the systems engineering standard approach/terminology. Figure 6 shows a snapshot of the ontology employed in the use case.

Specifications such as SRS and SCS represent entities. The properties attached include their unique identifier (UID) description and UID description label. The label refers to the vertical traceability analysis outcome provided by the ISA team (PASS or FAIL). Other entities are also defined to provide the wider context in which the specifications sit. Using the ontology as a template, data previously ingested and extracted was used to construct a graph representing the use case domain.

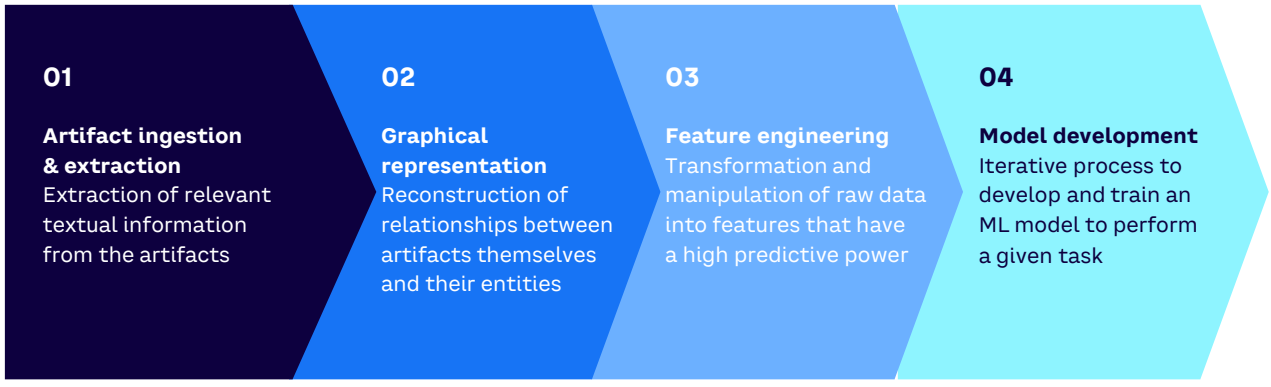


Figure 5. Methodology followed during the use case: from the original artifacts to predicting the outcome of a vertical traceability analysis (source: Arthur D. Little)

Figure 7 shows a portion of that graph, with only nodes and relationships related to SRSs (blue circles) and SCSs (orange circles) displayed.

By representing the problem as a graph, patterns and clusters quickly appear. For example, there seems to be an agglomeration of interconnected SRSs and SCSs in the middle of Figure 7. However, a high number of satellite groups disconnected from the central aggregate are also present around the edge of the graph. Adding more entities and relationships to the graph helped reveal insights into the interdependencies between entities, essentially revealing the inner workings of the audited system.

FEATURE ENGINEERING & MODEL DEVELOPMENT

Feature engineering is one of the most critical steps of the process because it's responsible for providing the model with informative features that help it accurately and precisely perform the task. The traditional way to approach the problem

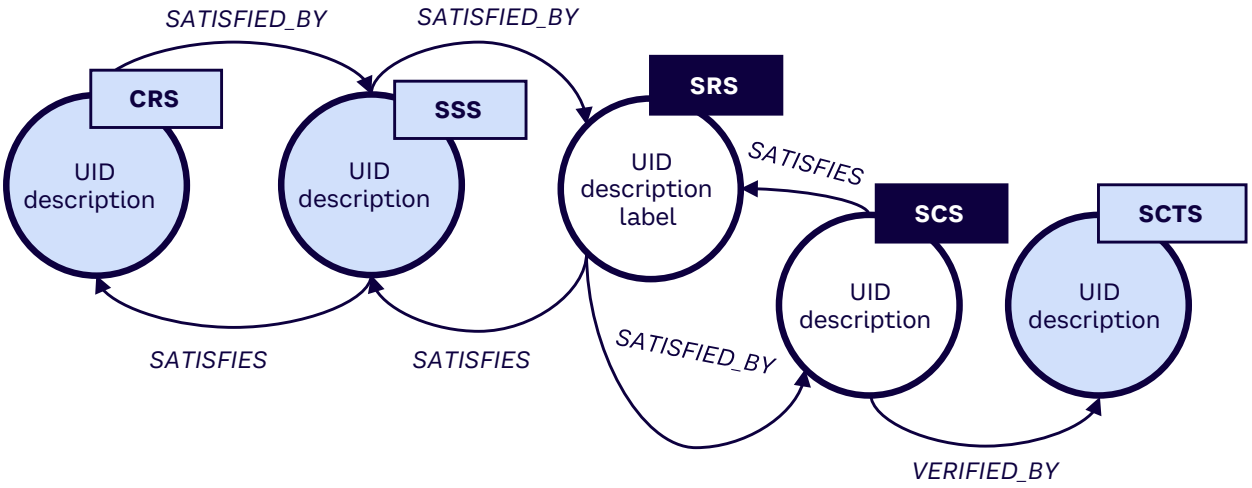


Figure 6. Partial ontology used for the use case, where specifications are entities, and relationships are based on where along the specification tree they sit; CRS, SSS, and SCTS are other entities linked to the specifications, namely SRS and SCS; they provide information on the wider context around which both entities sit (source: Arthur D. Little)

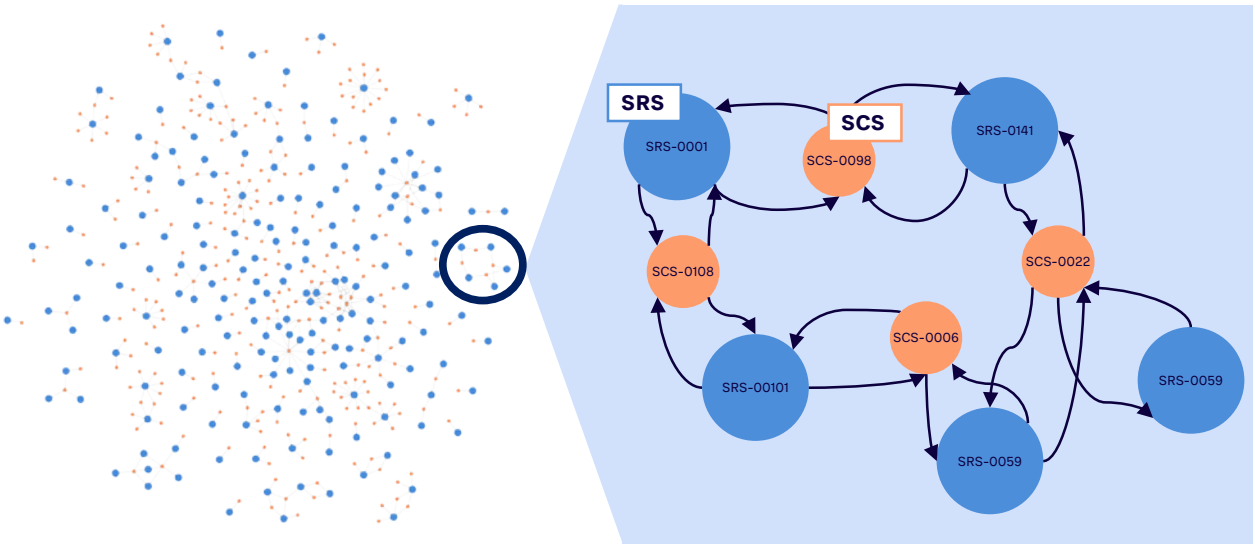


Figure 7. Sub-portion of the complete graph build to model the use case; only SRSs (blue nodes) and SCSs (orange nodes) are represented (source: Arthur D. Little)

FEATURE	TEXT-BASED FEATURES (NLP)	GRAPH-BASED FEATURES
Description	Use-case-specific and driven by how the ISA is typically performed: • Entity and expression similarity • Complexity measures • Keyword presence	• Community measure algorithms, used to evaluate how groups of nodes are clustered or partitioned • Centrality measure algorithms, used to determine the importance of distinct nodes in a network

Table 3. Main features derived from the data using NLP and graph algorithms (source: Arthur D. Little)

would be to try to generate semantic features by assessing whether Table 2 criteria are respected, as an ISA auditor would do. These features are referred to as text-based features in Table 3.

Representing the problem as a graph lets us instead extract features that describe the intrinsic architecture and interdependency of the data. As shown in Table 3, such features were generated using standard graph algorithms, namely community and centrality measures.[7] A well-known community measurement algorithm is called PageRank, named after Larry Page, cofounder of Google. This algorithm lets the Google search engine rank Web pages that are returned to the user by the search engine.[8]

REPRESENTING THE PROBLEM AS A GRAPH LETS US INSTEAD EXTRACT FEATURES THAT DESCRIBE THE INTRINSIC ARCHITECTURE AND INTERDEPENDENCY OF THE DATA

The model development process, which also incorporates feature engineering, follows the iterative process shown in Figure 8. Most of the time, features are removed, tweaked, or created based on the performance achieved and desired. The outcome is a trained, tested ML model that classifies the vertical traceablity analysis outcome of SRSs, as an ISA auditor would (PASS or FAIL).

RESULTS

The model's performance was evaluated using standard metrics: precision and recall. Precision is the model's accuracy at flagging FAILED SRSs; recall is the model's accuracy at recognizing true FAILED SRSs. Combining the graph-based features with the text-based features (i.e., features solely inspired by the vertical traceability criteria from Table 2) gave the best performing model, boosting both precision and recall by approximately 10% in comparison to using only text-based predictors.

Six of 14 additional SRSs were found to be incorrectly flagged by the ISA team: four being wrongly flagged as FAIL and two being wrongly flagged as PASS. Such findings demonstrate the model's ability to uncover safety failures not immediately obvious to the ISA team. These results show how powerful graphs can be, when used to represent highly interconnected data, and how to extract informative features from them. Additional methods, such as graph embeddings, can also be used to derive features from the graph's architecture.[9]

During a live ISA, both the graph and the model can be directly employed by the auditor to help him or her effectively perform the audit. First, the ML model would provide an ISA auditor with a prioritized list of potential safety issues. The highest-ranked issues would have the highest probability (as assessed by the model) of being actual FAILS and should be quickly investigated by the auditor. The model is not replacing the auditor; it provides a non-random, principled way to sample artifacts for analysis, making the best use of the auditor's time on potentially safety-critical issues.

The auditor could also interact directly with the graph through a user-friendly interface to explore the audited artifact and get additional insights. The visual, accessible nature of graphs makes

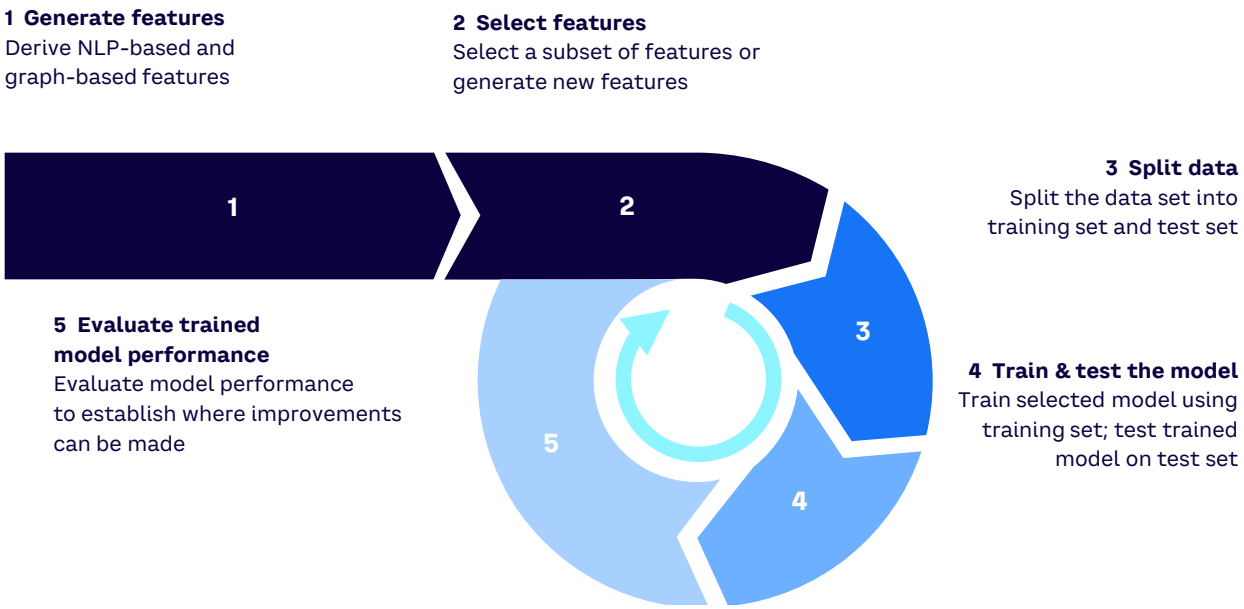


Figure 8. Iterative model development methodology followed to train and test an ML model that classifies the vertical traceability outcome of SRSs, as an ISA auditor would (PASS or FAIL) (source: Arthur D. Little)

them great mediums for exploration. One could also imagine the ML model being updated on the fly as the auditor progresses his or her audits, or by leveraging patterns uncovered by the auditor.

Finally, putting in place a good ontology meant that expanding the use case was easy, and the graph could easily be extended to accommodate new nodes/relationships. This also meant the data feeding the ML developed for the use case would not be affected by the graph scaling. Indeed, sub-portions of the graph can easily be isolated through simple queries, making it virtually isolated from the complete graph.

WHAT TO EXPECT FROM GRAPHS

The world's giant tech companies jumped on the "graph train" a while ago and now power some of the best-known tools, platforms, and services through graphs: the World Wide Web, social media, Web stores, and search engines. However, this does not mean companies should immediately start replacing all relational databases with graphs. New AI technologies tend to be initially seen as miracle solutions that will solve most problems (e.g., deep learning).

When deciding whether to use only graphs, only relational databases, or a combination, make sure to ask some key questions. For example, how important is rapid data exploration? How crucial is the speed at which data can be added to, and/or retrieved from, the data store? If graphs still come out as highly viable candidates, here are a few advantages you can expect from using them in your next data-driven project.

VISUALIZE YOUR DATA TO UNLOCK NEW INSIGHTS

Because graphs are so easy to visualize, it takes little effort to find all the information associated with a node and the direct/indirect relations that link two nodes. This property of KGs both simplifies data exploration and provides richer insights into it. The multiple dimensions of the KG can easily be explored by slicing it across one or more dimensions. In the use case, visualizing the SRS-SCS architecture (see Figure 6) led to a key hypothesis of the problem: the way specifications are linked and clustered together is closely related to the vertical traceability analysis outcome. An SRS connected to a failed SRS through nearby elements is more likely to be flagged as a FAIL.

EXTRACT MORE FROM YOUR DATA

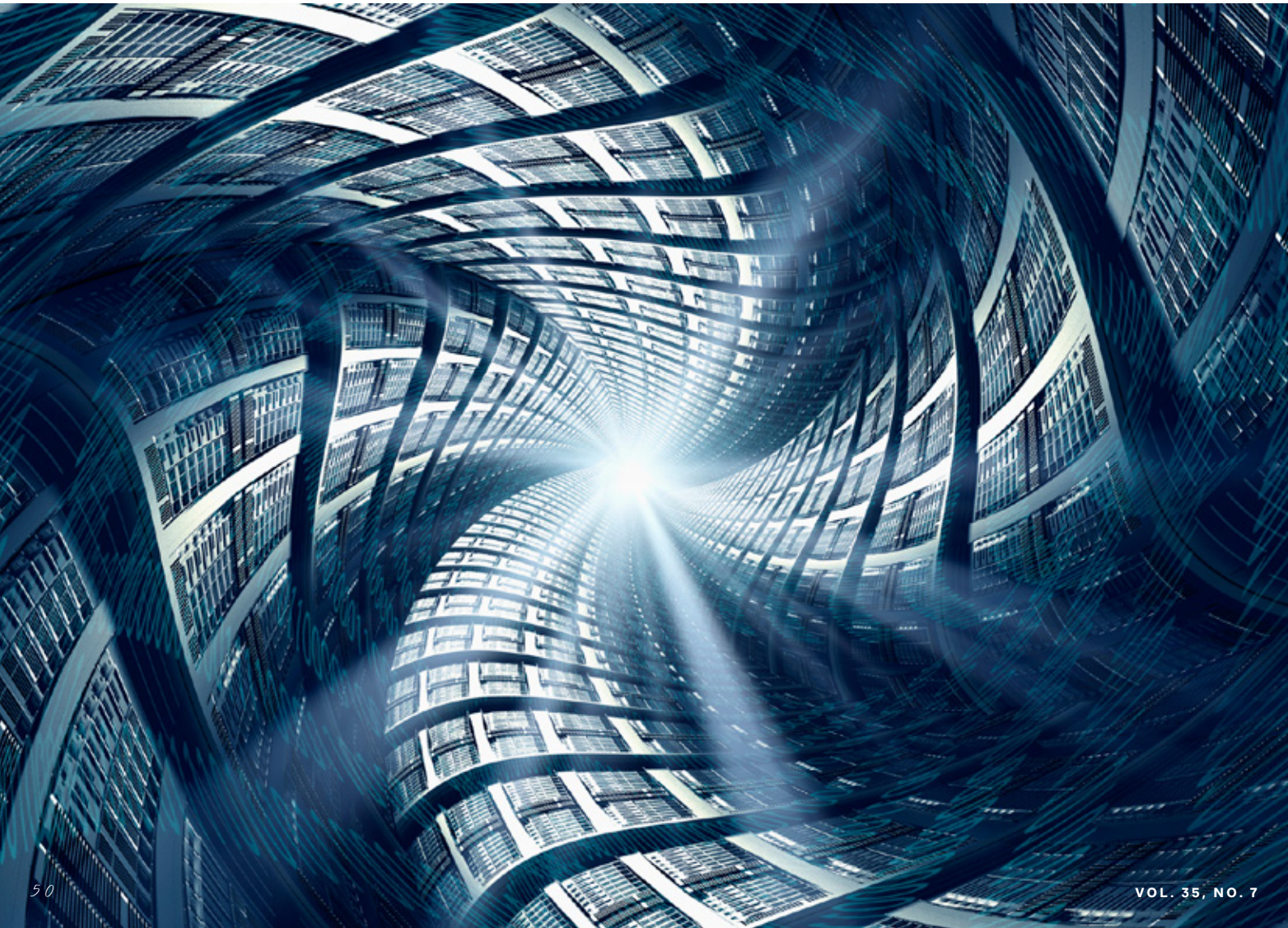
The inherent interdependencies hidden in your data can be brought to light and leveraged by running algorithms on your graph. As seen in the use case, they can be used to compute several metrics on the whole graph or a sub-portion that can help make sense of your connected data and their inner workings. These metrics can then be used as features to power an ML model.

START SMALL, SCALE FAST

You might initially decide to build a graph that models a small portion of your domain space. That's fine. Nothing prevents you from later expanding it to answer new questions or because more data becomes available.

Within graphs, it is easy to add a new type of node property or relationship. That is, the new property/relationship can be applied to a (potentially small) subset of nodes. If you have many node properties and/or relationships that apply only locally, KGs will be both much smaller and faster to process than their corresponding relational databases. Multiple graphs can also be combined if they share or have related entities, limiting high rearchitecting costs and enabling you to quickly grow your solution.

Our parting thought: if the world's big data is a mountain of dots, knowledge graphs will help you connect them all.



REFERENCES

[1] Papadopoulos, Michael, and Philippe Monnot. "Why Do Machine Leaning Analytics Projects Fail?" *Cutter Business Technology Journal* (renamed *Amplify*), Vol. 34, No. 6, 2021.

[2] Trudeau, Richard J. *Introduction to Graph Theory*. Dover Publications, 1993.

[3] Barabási, Albert-László. *Network Science*. Cambridge University Press, 2016.

[4] Walden, David D., et al. *INCOSE Systems Engineering Handbook: A Guide for System Life Cycle Processes and Activities*. 4th edition. Wiley, 2015.

[5] "What Is Independent Safety Assessment (ISA)?" ISA Working Group, 2011.

[6] Robinson, Ian, Jim Webber, and Emil Eifrem. *Graph Databases: New Opportunities for Connected Data*. O'Reilly Media, 2015.

[7] Needham, Mark, and Amy E. Hodler. *Graph Algorithms: Practical Examples in Apache Spark & Neo4j*. O'Reilly Media, 2019.

[8] Needham and Hodler (see 7).

[9] Needham and Hodler (see 7).

About the authors

Michael Eiden is a Cutter Expert and a member of the Arthur D. Little (ADL) AMP open consulting network. Dr. Eiden serves as Partner and Head of AI at ADL and is an expert in machine learning (ML) and artificial intelligence (AI) with more than 15 years' experience across different industrial sectors. He has designed, implemented, and productionized ML/AI solutions for applications in medical diagnostics, pharma, biodefense, and consumer electronics. Dr. Eiden brings along deep expertise in applying supervised, unsupervised, as well as reinforcement ML methodologies to a very diverse set of complex problem types. He has worked in various global technology hubs, such as Heidelberg (Germany), Cambridge (UK), and Silicon Valley (US), with clients ranging from small and medium-sized enterprises to globally active organizations. Dr. Eiden earned a doctorate in bioinformatics. He can be reached at experts@cutter.com.

Philippe Monnot is a Cutter Expert, a Data Scientist with ADL's Digital Problem Solving (DPS) practice, and a member of ADL's AMP open consulting network. He's passionate about solving complex challenges that impact people's livelihoods through the use of data, statistics, and ML. Mr. Monnot enjoys developing accessible solutions that customers will adopt through effective data storytelling and explainable AI. Before joining ADL, he worked in R&D, where he used ML to implement smart, scalable manufacturing processes to manufacture sustainable composite structures for the aerospace and oil and gas industries. He can be reached at experts@cutter.com.

Armand Rotaru is an AI/ML data scientist with ADL's Digital Problem Solving (DPS) practice and has been involved in a variety of projects that have a natural language processing (NLP) component, predominantly in the petrochemical, transportation, and biomedical sectors. He is also responsible for maintaining/expanding the NLP section of ADL's DPS Training Portal and mentoring junior team members. Mr. Rotaru has a master of science degree in AI from VU Amsterdam, a PhD in theoretical computer science from the School of Advanced Studies of the Romanian Academy, and a second PhD in experimental psychology from London's UCL. He can be reached at experts@cutter.com.

AMPLIFY

Anticipate, Innovate, Transform

Cutter Consortium, an Arthur D. Little community, is dedicated to helping organizations leverage emerging technologies and the latest business management thinking to achieve competitive advantage and mission success through our global research network. Cutter helps clients address the spectrum of challenges disruption brings, from implementing new business models to creating a culture of innovation, and helps organizations adopt cutting-edge leadership practices, respond to the social and commercial requirements for sustainability, and create the sought-after workplaces that a new order demands.

Since 1986, Cutter has pushed the thinking in the field it addresses by fostering debate and collaboration among its global community of thought leaders. Coupled with its famously objective "no ties to vendors" policy, Cutter's *Access to the Experts* approach delivers cutting-edge, objective information and innovative solutions to its community worldwide.

Amplify is published monthly by Cutter Consortium, an Arthur D. Little community, 37 Broadway Suite, Arlington, MA 02474-5552, USA

Founding Editor: Ed Yourdon
Publisher: Karen Fine Coburn
Group Publisher: Christine Generali
Production Manager: Linda Dias
Editors: Jennifer Flaxman, Tara K. Meads

CUTTER
AN ARTHUR D. LITTLE COMMUNITY

For more content, visit www.cutter.com