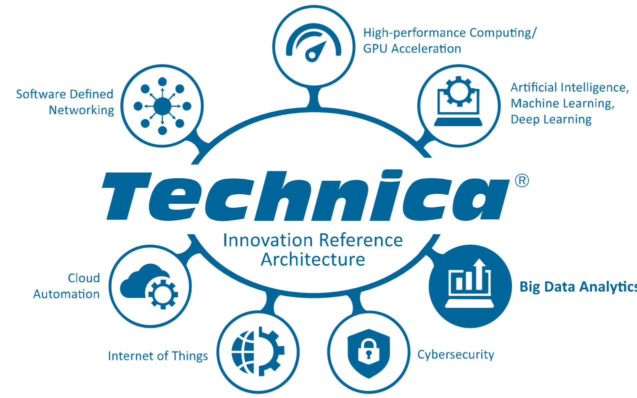
Technica’s Independent Research and Development (IR&D) team is focused on engineering innovative Big Data offerings that make use of Graphics Process Units (GPUs). The massive parallelism of GPUs allows significantly faster data processing and analysis without having to purchase expensive hardware and software licenses.

FUNL is Technica’s GPU accelerated graph analytics engine.

But… what is graph analytics? Why is it important? What are the best use cases for this approach to data analysis?

The Technica Innovation Platform White Paper Series presents advanced topics that will drive competitive advantage for next-generation IT over the next three-to-five years.




# WHAT IS GRAPH ANALYTICS? AND WHEN DO YOU USE IT?

## BACKGROUND

Since the early 1970s, the relational data modeling structure dominated the way in which data is modeled. Business empires like Oracle, Informix, and MySQL have been built on the relational model. Its pervasiveness causes us to think of data in terms of tables, primary keys, joins and how to write the best SQL queries. But, the relational model is not the only way in which to model data.

With the volume, variety, and velocity of data produced by Twitter, Facebook, Snapchat, and the like; the NoSQL movement gained steam. NoSQL, aided by technologies like JSON and Hadoop, simply captures data without the need to derive and enforce a data schema.

Modeling data with graphs is another, non-relational approach in which to derive meaning from data. Through various graph analytic algorithms, relationships latent to the naked eye may become readily identifiable. Whereas tables are the prism through which to view relationally modeled data, *relationships* are the keys to unlock the graph analytic kingdom.

## GRAPHS

Adam Kocoloski, CTO of IBM Cloud Data Services recently said at IBM InterConect 2016, “Sometimes people don’t know they’re dealing with a graph-shaped problem.”[1] Before we examine what constitutes a graph-shaped problem, a brief review of graphs is in order. Graphs are fundamental structures that represent complex structured information. The key elements of graphs are *vertices* (also

[1] http://www.ibmbigdatahub.com/blog/open-data-takeaways-ibm-interconnect-2016

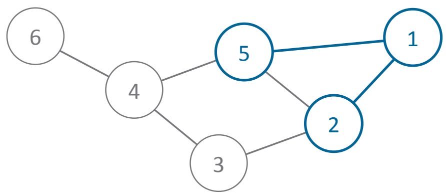



**Figure 1 – Undirected Graph**

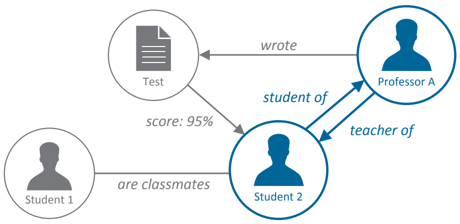



**Figure 2 – Directed Graph**

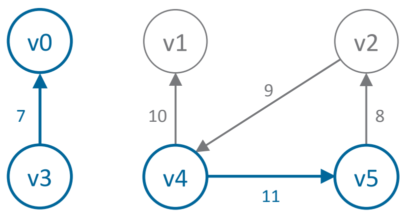



**Figure 3 – Directed, Weighted Graph**

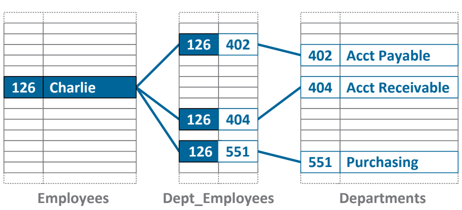



**Figure 4 – Relational Data Model**

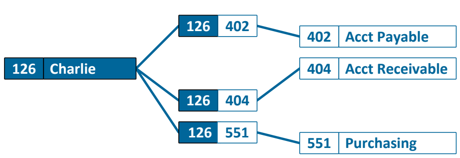



**Figure 5 – Graph Data Model**

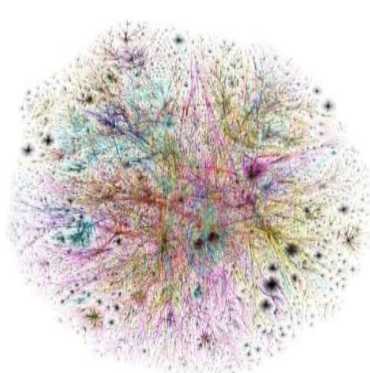


**Figure 6 – Facebook Friendships**

known as nodes) and relationships among vertices called *edges*. Edges in graphs are also called links.

**Figure 1** portrays a basic graph. The numbered circular elements are the vertices. The lines denoting relationships among the vertices are the graph's edges. As a model for data, this basic graph conveys a large amount of information. For example, Vertex 1 has a relationship (because of the edges) with Vertex 5 and Vertex 2, but no relationship with Vertex 6.

Obviously, graphs can be more complicated than Figure 1. For example, the edges of Figure 1 are undirected. Directed edges would carry more information. **Figure 2** provides a look at a graph with directed edges.

Students 1 and 2, Professor A, and Test are the graph's vertices. The relationship between Student 2 and Professor A is bi-directional, conveying a two-way relationship. Additionally, some graphs employ weights to show the strength of the edge. **Figure 3** presents a graph of this nature.

In Figured 3, the directed edge, and therefore the relationship, between Vertex 4 and Vertex 5 is stronger (numeric value of 11) than the directed edge between Vertex 0 and Vertex 3 (numeric value of 7).

After examining the graphs in Figures 1-3, it becomes apparent that the relational model is much more hierarchical than the graph data model. While graph-like relationships can be conveyed in a relational database, a tortuous process of primary and secondary keys must be employed. Consider Figure 4.

**Figure 4** presents a relational data model in which Persons, i.e. employees, can be the member of more than one department in an organization. This many-to-many relationship must be portrayed with primary keys, foreign keys, and a Dept_Member table that serves as a bridge or join between the other tables. Now contrast Figure 4 with Figure 5. **Figure 5** portrays a graph data model of the same many-to-many relationship. Alice is much more intuitively associated with multiple departments in a graph data model.

The graphs in Figures 1-3 are ingestible by the human mind, but imagine a graph of two billion edges. The graph analytics employed by FUNL shine when the graphs are extremely large. **Figure 6** conveys the complexity of a large graph by portraying Facebook friendships.

Now we return to the original question: What are graph-shaped problems? Graph shaped problems are problems in which you are more concerned with relationships among entities, i.e. the vertices, than the vertices in isolation.

## GRAPH ANALYTIC TECHNIQUES

Given the structure of graphs, graph analytics typically employs one or more of the following methodologies to make sense out of the data.

- **Centrality analysis**: Identifies the most key or central vertices in a network. This is very useful for influence analysis.

- **Path analysis**: Identifies connections between a pair of vertices. This is helpful in understanding risks and exposure.
- **Community detection**: Identifies clusters or communities. This of great importance in understanding social and biological networks.
- **Sub-graph isomorphism**: Searches for patterns of relationships within portions of an overall graph. This is useful for validating hypotheses and searching for abnormal situations, such as hacker attack.

These techniques are applied to various use cases we will now examine in finer detail.

## GRAPH-SHAPED PROBLEMS

While relational analytics is good for examining one-to-one and one-to-many relationships, insights tend to be limited for many-to-many relationship use cases. Graph analytics shines when insight into many-to-many relationships is desired.

Graph analytics can be used to answer questions, like:

- Customers are associated with other customers in a community—who are the key influencers in this customer community?
- Terrorists support certain kinds of organizations and share about them in social media—what other organizations are gaining the most traction in similar terror networks?
- Patients have various diagnosis that relate to other patients with similar issues—how can we predict better outcomes if certain factors of care are modified?

For example, graph analytics could be utilized to not only ask questions about friends of a person (one-to-many) but friends of friends (many-to-many). Performing this friend of friend analysis would necessitate a very complex SQL query with numerous extended joins. Graph analytics can determine paths through the data, discovering complex relationships that are not obvious and are difficult to detect with relational analytics.

It is important to note that graph analytics does not replace relational analytics. Graphs are simply better for certain use cases. Generally, as exemplified by the questions above, graph analytics excels when questions are open-ended and non-obvious—unknown, unknowns.

Modeling data through graphs is relevant for a number of use cases:

### IT Analytics

Graphs are used to represent networks of communication, data organization, computational devices, and the flow of computation. The link structure of a website can be represented by a directed graph, in which the vertices represent web pages and directed edges represent links from one page to another. Google built its search engine on the PageRank graph analytic algorithm in this manner.

### Influence Analysis

Influence analysis is a sweet spot for graph-shaped problems. The aforementioned PageRank algorithm is an example of influence analysis where more heavily trafficked web pages are ranked higher, i.e. more influential, all other things equal.

Everything is our digital world is connected in a network of relationships and influence among people, products, and technologies. Influence analysis can be used in any domain from analyzing traffic patterns, weather anomalies, which consumer is most influential, etc.

### Cybersecurity

Cisco collects data from all of its customer's firewalls, intrusion protections systems, and other security appliances. It allows them to better anticipate the vast world of evolving cyber threats including zero-day exploits, patterns of spear phishing, and other attack vectors.

Graph analytics helps make sense of this vast data. Cisco uses information it knows on 30M IP domains to learn about other unknown domains by finding hidden links. They analyze IP address, registrars, and domain names, looking for these connections.

### Social Network Analysis

Graph analytics is used extensively by Facebook and LinkedIn to suggest new connections. This is because graph analytics can detect communities that congregate around certain themes.

Graph theory is used in sociology as a way, for example, to measure actors' prestige or to explore rumor spreading, notably through the use of social network analysis software.

In the world of intelligence, numerous government agencies are interested in identifying threats through the detection of non-obvious patterns of relationships, and grouping communications buried in social media, email, texting and call detail records.

### Fraud Detection

Analyzing fraud patterns with graph analytics can help pinpoint and prevent adverse influence patterns. Usually, these sorts of influence patterns are anything but hierarchical, so they are well-suited to graphs.

### Graphing the IOT

The Internet of Things (IOT) will grow exponentially in the coming decade. Sensors tend to be deployed in nonhierarchical grids of great complexity. The collection of data from the endpoints can be analyzed to determine influence relationships that may be applicable for certain types of data exploration, like root analysis.

Connected cars and personal health trackers can be leveraged to determine shifting patterns of influence, for example, certain drivers are avoiding an intersection. Why?

### Semantic Search and Knowledge Discovery

Google employs a "Knowledge Graph" to enhance its search engine results with semantic-search capabilities. The vertices in the knowledge graph are concepts. The strength of the relationships among concepts help provide a weighting of related ideas in a non-hierarchical manner. For instance, corporate assets could be linked in similar knowledge graphs, allowing the company to offer access to information beyond keyword searches.

### Natural Language Understanding

Graph-theoretic methods have proven useful in linguistics, since natural language often lends itself well to discrete structure. Traditionally, syntax and compositional semantics follow tree-based structures, whose expressive power lies in the principle of compositionality, modeled in a hierarchical graph.

However, unstructured text can also be modeled in a non-hierarchical graph. Word associations from a text corpus are also suitable for graph analytics. For example, opinions can be mined from a corpus of product reviews. Graphs can also enable content recommendations based on keyword extraction.

### Life Sciences

In the medical field, organizations can use graph analytics to conduct research in healthcare fraud for bill-payers. Other potential graph analytics use cases include healthcare treatment efficacy and outcome analysis, analyzing drugs and side effects, and the analysis of proteins and gene pathways.

In the area of personalized healthcare, a startup called Lumiata intends to scale personalized medicine by leveraging machine learning and graphic analytics to empower nurses to carry more of the diagnostic chores, freeing doctors to focus on more urgent care needs.

### General Business Relationships

Graph Analytics can be used to address relationship-based problems in manufacturing, energy, gas exploration, travel, biology, conservation, computer chip design, chemistry, physics, higher education research, government, security, defense and many other fields. As with relational analytics, graph analytic use cases are inexhaustible.

## CONCLUSION

Graph analytics is a powerful paradigm for Big Data analysis that complements the older, more established relational analytics model powered by SQL. Graph analytics is used to answer questions that are many-to-many in nature. With the exponentially increasing nature of Big Data, graph analytics will only grow in importance.

Graph analytics offers IT the ability to analyze relationships or connections in Big Data. Since almost everything is connected in some way, shape, or form the only limit to graph analytics is our imagination.

Technica provides professional services, products, and innovative technology solutions to the Federal Government. We specialize in network operations and infrastructure; cyber defense and security; government application integration; systems engineering and training; and product research, deployment planning, and support.

Technica®
22970 Indian Creek Drive, Suite 500
Dulles, VA 20166
703.662.2000