THE AGE OF THE GRAPH

Michael K. Bergman[1], Coralville, Iowa USA

August 12, 2012

AI3:::Adaptive Information blog

Virtually everywhere one looks we are in the midst of a transition for how we organize and manage information, indeed even relationships. Social networks and online communities are changing how we live and interact. NoSQL and graph databases — married to their near cousin Big Data — are changing how we organize and store information and data. Semantic technologies, backed by their ontologies and RDF data model, are showing the way for how we can connect and interoperate disparate information in ways only dreamed about a decade ago. And all of this, of course, is being built upon the infrastructure of the Internet and the Web, a global, distributed network of devices and information that is undoubtedly one of the most important technological developments in human history.

There is a shared structure across all of these developments — the graph. Graphs are proving to be the new universal paradigm for how we organize and manage information. Graphs have an inherently expandable nature, and one which can also capture any existing structure. So, as we see all of the networks, connections, relationships and links — both physical and informational — grow around us, it is useful to step back a bit and contemplate the universal graph structure at the core of these developments.

Understanding that we now live in the Age of the Graph means we can begin studying and using the concept of the graph itself to better analyze and manage our interconnected world. Whether we are trying to understand the physical networks of supply chains and infrastructure or the information relationships within ontologies or knowledge graphs, the various concepts underlying graphs and graph theory, themselves expressed through a rich vocabulary of terms, provide the keys for unlocking still further treasures hidden in the structure of graphs.
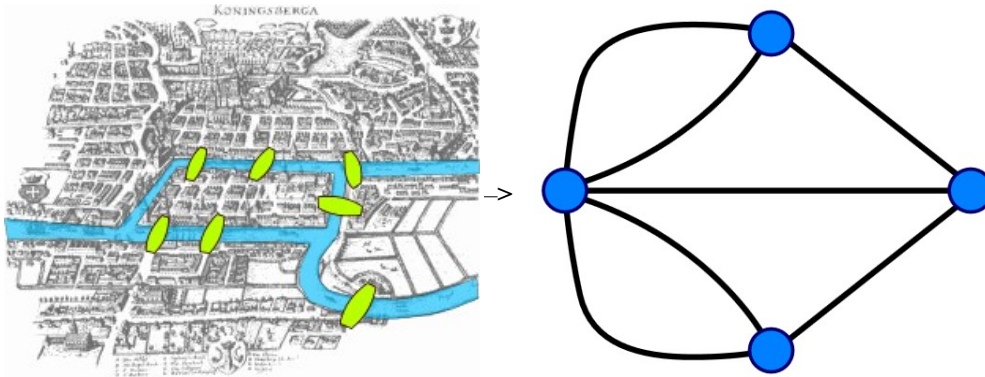
Graphs as a Concept

The use of "graph" as a mathematical concept is not much more than 100 years old. The beginning explication of the various classes of problems that can be addressed by graph theory probably is no older than 300 years. The use of graphs for expressing logic structures probably is not much older than 100 years, with the intellectual roots beginning with Charles Sanders Peirce [1]. Though likely trade routes and their affiliated roads and primitive transportation or nomadic infrastructures were perhaps the first expressions of physical networks, the emergence and prevalence of networks is a fairly recent phenomenon. The Internet and the Web are surely the catalyzing development that has brought graphs and networks to the forefront.

In mathematics, a graph is an abstract representation of a set of objects where pairs of the objects are connected. The objects are most often known as *nodes* or *vertices*; the connections between the objects are called *edges*. Typically, a graph is depicted in diagrammatic form as a set of dots or bubbles for the nodes, joined by lines or curves for the edges. If there is a logical relationship between connected nodes the edge is directed, and the graph is known as a directed graph. Various structures or topologies can be expressed through this conceptual graph framework. Graphs are one of the principle focuses of study in discrete mathematics [2]. The word "graph" was first used in the sense as a mathematical structure by J.J. Sylvester in 1878 [3].

---

[1]Email: mike@mkbergman.com

As representative of various data models, particularly in our company's own interests in the Resource Description Framework (RDF) model, the nodes can represent "nouns" or subjects or objects (depending on the direction of the links) or attributes. The edges or connections represent "verbs" or relationships, properties or predicates. Thus, the simple "triple" of the basic statement in RDF (consisting of *subject – predicate – object*) is one of the constituent barbells that make up what becomes the eventual graph structure.

The manipulation and analysis of graph structures comes under the rubric of graph theory. The first recognized paper in that field is the *Seven Bridges of Königsberg*, written by Leonhard Euler in 1736. The objective of the paper was to find a walking path through the city that would cross each bridge once and only once. Euler proved that the problem has no solution:

Euler's approach represented the path problem as a graph, by treating the land masses as nodes and the bridges as edges. Euler's proof postulated that if every bridge has been traversed exactly once, it follows that, for each land mass (except for the ones chosen for the start and finish), the number of bridges touching that land mass must be even (the number of connections to a node we now call "degree"). Since that is not true for this instance, there is no solution. Other researchers, including Leibniz, Cauchy and L'Huillier applied this approach to similar problems, leading to the origin of the field of topology.

Later, Cayley broadened the approach to study tree structures, which have many implications in theoretical chemistry. By the 20th century, the fusion of ideas coming from mathematics with those coming from chemistry formed the origin of much of the standard terminology of graph theory.

The Theory of Graphs

Graph theory forms the core of network science, the applied study of graph structures and networks. Besides graph theory, the field draws on methods including statistical mechanics from physics, data mining and information visualization from computer science, inferential modeling from statistics, and social structure from sociology. Classical problems embraced by this realm include the four color problem of maps, the traveling salesman problem, and the six degrees of Kevin Bacon.

Graph theory and network science are the suitable disciplines for a variety of information structures and many additional classes of problems. This table lists many of these applicable areas, most with links to still further information from Wikipedia:

Graph Structures	Graph Problems
Data structures • linked data structures	**Enumeration**

Graph Structures	**Graph Problems**

- arrays
- B-trees
- tables

Tree structures
- grammatical structures
- binary trees
- evolutionary trees
- work breakdown structures
- nested sets
- decision trees
- taxonomies
- family trees
- tries
- tree automata
- abstract syntax trees
- phylogenetic trees

List structures
- incidence lists
- adjacency lists
- linked lists
- queues

Matrix structures
- incidence matrices
- adjacency matrices
- Laplacian matrices
- distance matrices
- lattices

Path structures
- paths
- simple paths
- loops
- Eulerian paths
- Hamiltonian paths

Networks
- engineered networks
- information networks
- biological networks
- cognitive and semantic networks
- social networks
- Bayesian networks
- complex networks
- scale-free networks
- small-world networks

- graphical enumeration

Subgraphs, induced subgraphs, and minors
- subgraph isomorphism problem
- clique problem
- finding induced subgraphs
- the independent set problem
- minor containment problem
- the reconstruction conjecture

Search and navigation
- tree traversal
- breadth-first search
- depth-first search

Graph coloring
- four-color theorem
- strong perfect graph theorem
- Erdős–Faber–Lovász conjecture (unsolved)
- total coloring conjecture (unsolved)
- list coloring conjecture (unsolved)
- Hadwiger conjecture (unsolved)

Subsumption and unification
- operations between graphs
- unification of graphs
- automatic theorem proving
- modeling linguistic structure

Route (path) problems
- Hamiltonian path and cycle problems
- minimum spanning tree
- route inspection problem (also called the "Chinese Postman Problem")
- Seven Bridges of Königsberg
- shortest path problem
- Steiner tree
- three-cottage problem
- traveling salesman problem
- critical path analysis
- percolation

Matrix manipulations (many)

Network flow
- max flow min cut theorem

Visibility graph problems

Graph Structures	Graph Problems

- museum guard problem

Covering problems
- set cover problem
- vertex cover problem

Logic structures
- Logic graphs
- Ontologies
- Conceptual graphs
- Mind maps

Random graphs

Weighted graphs

Sparse/dense graphs

Graph structure
- degree distribution
- power laws
- clustering coefficients
- assortativity or disassortativity
- community structure
- hierarchical structure
- clustering
- reciprocity
- graph partitioning

Graph classes
- enumerating the members of a class
- characterizing a class in terms of forbidden substructures
- ascertaining relationships among classes
- deciding membership in a class
- finding representations for members of a class

Graphs are among the most ubiquitous models of both natural and human-made structures. They can be used to model many types of relations and process dynamics in physical, biological and social systems. Many problems of practical interest can be represented by graphs. This breadth of applicability makes network science and graph theory two of the most critical analytical areas for study and breakthroughs for the foreseeable future. I touch on this more in the concluding section.

Graphs as Physical Networks

Surely the first examples of graph structures were early trade and nomadic routes. Here, for example, are the trade routes of the Radhanites dating from about 870 AD [4]:

It is not surprising that routes such as these, or other physical networks as exemplified by the bridges of Königsberg, were the stimulus for early mathematics and analysis related to efficient use of networks. Minimizing the time to complete a trade circuit or visiting multiple markets efficiently has clear benefits. These economic rationales apply to a wide variety of modern, physical networks, including:

- Telecommunications networks
- Global shipping networks
- Transport networks
- Electrical networks and grids and,
- Computer networks.

Of course, included in the latter category is the Internet itself. It is the largest graph in existence, with an estimated 2.2 billion users and their devices all connected in one way or another in all parts of the globe [5].

Graphs as Natural Systems

Graphs and graph theory also have broad applicability to natural systems. For example, graph theory is used extensively to study molecular structures in chemistry and physics. A graph makes a natural model for a molecule, where vertices represent atoms and edges bonds. Similarly, in biology or ecology, graphs can readily express such systems as species networks, ecological relationships, migration paths, or the spread of diseases. Graphs are also proper structures for modeling biological and chemical pathways.

Some of the exemplar natural systems that lend themselves to graph structures include:

- Chemical reaction networks
- Gene regulatory networks
- Spin networks
- Neural networks
- Ecological networks, and
- Petri nets (chemistry).

As with physical networks, a graph representation for natural systems provides real benefits in computer processing and analysis. Once expressed as a graph, all graph algorithms and perspectives from graph theory and network science can be brought to bear. Statistical methods are particularly applicable to representing connections between interacting parts of a system, as well to representing the physical dynamics of natural systems.

Graphs as Social Networks

Parallel with the growth of the Internet and Web has been the growth of social networks. Social network analysis (SNA) has arguably been the single most important driver for advances in graph theory and analysis algorithms in recent years. New and interesting problems and challenges — from influence to communities to conflicts — are now being elucidated through techniques pioneered for SNA.

Second only in size to the Internet has been the graph of interactions arising from Facebook. Facebook had about 900 million users as of May 2012, half of which accessed the service via mobile devices [6]. Facebook famously embraced the graph with its own Open Graph protocol, which makes it easy for users to access and tie into Facebook's social network. A representation of the Facebook social graph as of December 2010 is shown in this well-known figure:

The suitability of the graph structure to capture relationships has been a real boon to better understanding of social and community dynamics. Many new concepts have been introduced as the result of SNA, including such things as influence, diversity, centrality, cliques and so forth. (The opening diagram to this article, for example, models *centrality*, with blue the maximum and red the minimum.)

Particular areas of social interaction that lend themselves to SNA include:

- Social networks
- Military conflicts and terrorism
- Value networks
- Project networks
- Workflows, and
- Business ecosystems.

Entirely new insights have arisen from SNA including finding terrorist leaders, analyzing prestige, or identifying keystone vendors or suppliers in business ecosystems.

Graphs as Information Representations

Given the ubiquity of graphs as representations of real systems and networks, it is certainly not surprising to see their use in computer science as as means for information representation. We already saw in the table above the many data structures that can be represented as graphs, but the paradigm has even broader applicability.

The critical breakthroughs have come through using the graph as a basis for data models and logic models. These, in turn, provide the basis for crafting entire graph-based vocabularies and languages. Once such structures are embraced, it is a natural extension to also extend the mindset to graph databases as well.

Some of the notable information representations that have a graph as their basis include:

- Graph-structured data
- Graph databases
- RDF data model
- OWL ontology language
- Common logic
- Conceptual graphs, and
- Peirce's symbolic or relational logic, as well as his use of triadic relations [1].

Graphs as Knowledge Representations

A key point of graphs noted earlier was their inherent extensibility. Once graphs are understood as a great basis for representing both logic and data structures, it is a logical next step to see their applicability extend to knowledge representations and knowledge bases as well.

Graph-theoretic methods have proven particularly useful in linguistics, since natural language often lends itself well to discrete structure. So, not only can graphs represent syntactic and compositional structure, but they can also capture the interrelationships of terms and concepts within those languages. The usefulness of graph theory to linguistics is shown by the various knowledge bases such as WordNet (in various languages) and VerbNet.

Domain ontologies are similar structures, capturing the relationships amongst concepts within a given knowledge domain. These are also known as knowledge graphs, and Google has famously just released its graph of entities to the world [7]. Semantic networks and neural networks are similar knowledge representations.

What all of these examples show is the nearly universal applicability of graphs, from the abstract to the physical, from the small to the large, and every gradation between. We also see how basic graph structures and concepts can be built upon with more structure. This breadth points to the many synergies and innovations that may be transferred from diverse fields to advance the usefulness of graph theories.

Graphs as a Guiding Paradigm

Despite the many advances that have occurred in graph theory and the increased attention from social network analysis, many, many graph problems remain some of the hardest in computation. Optimizations, partitioning, mapping, inferencing, traversing and graph structure comparisons remain challenging. And, some of these challenges are only growing due to the growth in the size of networks and graphs.

Applying the lessons of the Internet in such areas as non-relational databases, distributed processing, and big data and map reduce-oriented approaches will help some in this regard. We're learning how to divide and conquer big problems, and we are discovering data and processing architectures more amenable to graph-based problems.

The fact we have now entered the Age of the Graph also bodes that further scrutiny and attention will lead to more analytic breakthroughs and innovation. We may be in an era of Big Data, but the structure underlying all of that is the graph. And that reality, I predict, will result in accelerated advances in graph theory.

Acknowledgements

This article was originally posted on the *AI3::Adaptive Information* Web site at http://www.mkbergman.com/1020/the-age-of-the-graph/. This version has been edited and reformatted slightly for PDF distribution. We thank Cognonto Corporation for making this content freely available.

---

[1] For a fairly broad discussion of Peirce in relation to these topics, see M.K. Bergman, 2012. "Give Me a Sign: What Do Things Mean on the Semantic Web?," in *AI3:::Adaptive Innovation* blog, January 24, 2012. See http://www.mkbergman.com/994/give-me-a-sign-what-do-things-mean-on-the-semantic-web/.

[2] Topics in discrete mathematics, which are all applicable to graphing techniques and theory, include theoretical computer science, information theory, logic, set theory, combinatorics, probability, number theory, algebra, geometry, topology, discrete calculus or discrete analysis, operations research, game theory, decision theory, utility theory, social choice theory, and all discrete analogues of continuous mathematics.

[3] See reference 1 in the Wikipedia entry on graph theory.

[4] According to Wikipedia, the Radhanites were medieval Jewish merchants involved in trade between the Christian and Islamic worlds during the early Middle Ages (approx. 500–1000 AD). Many trade routes previously established under the Roman Empire continued to function during that period largely through their efforts. Their trade network covered much of Europe, North Africa, the Middle East, Central Asia and parts of India and China.

[5] See the article on the Internet in Wikipedia for various size estimates.

[6] See the article on the Facebook in Wikipedia for various size estimates.

[7] For my discussion of the Google Knowledge Graph, see M.K. Bergman, 2012. "Deconstructing the Google Knowledge Graph," in *AI3:::Adaptive Innovation* blog, May 18, 2012. See http://www.mkbergman.com/1009/deconstructing-the-google-knowledge-graph/.

[8] UMBEL (the *Upper Mapping and Binding Exchange Layer*) is designed to help content interoperate on the Web. It provides two functions: a) it is a broad, general reference structure of 25,000 concepts, which provides a scaffolding to link and interoperate other datasets and domain vocabularies, and b) it is a base vocabulary for the construction of other concept-based domain ontologies, also designed for interoperation.