Open Problems in the Universal Graph Theory*

Marko A. Rodriguez
Director of Engineering, DataStax Inc.
Project Management Committee, Apache TinkerPop
(Dated: May 25, 2017)

The universal graph is a theoretical construct capturing the idea that every aspect of reality can be modeled as a graph composed of vertices and edges and, as such, reality is a graph. From the physical world of atoms, people, and galaxies to the mental planes of thoughts, words, and knowledge, there exists a universal graph hosting all such structures. While this idea is enticing, there are still strides to be made in coming to terms with a reality that is not composed of atoms bound by spacetime, but instead, a graph composed of vertices united by edges. This letter presents three open problems in our understanding of the universal graph.

I. INTRODUCTION

The world of graphs is about representing everything we know as a collection of vertices and edges and is captured by the colloquial expression "everything is a graph." Graphs can be used to model the objective structures of our world – power grids, vascular systems, and the protein chains created by our DNA [10]. Graphs also enable us to model the more subjective and personal aspects of our lives – social ties [29], language structures [2], and knowledge bases [28]. From representing the social bonds that unite us to simulating the roads that enable us to keep those bonds strong, the world of graphs appear to touch every aspect of our reality. This reality is defined by the amalgamation of all graphs considered into a single, interconnected *universal graph* (colloquially known as The TinkerPop) [13]. This universal graph not only exists as vertices and edges encoded in silicon (in databases), but also in the atoms of our physical world (in nature). The universal graph is all relational quanta at every strata of existence. For the physicist, there are the lattices of spacetime [1], spin networks at Plank scale [21], and the cosmological maps of the galaxies and stars in our universe [6]. For the neurologist, atomic graphs are organized into neuronal vertices and dendritic edges yielding brain-like neural networks [22]. For the psychologist, the neurons and axons manifest language networks with word vertices projecting edges to the objects they denote in the atomic-spacetime graph [25]. References abound as the universal graph is all that surrounds regardless of the form it takes and no vertex nor edge has a special vantage as being more "real" than any other. Instead, the picture that emerges is one where (schema-motifed) subgraphs endogenous to the universal graph intricately link to one another, creating everything in front of and within us.

The universal graph's structure evolves when vertices and edges are added and removed. This evolution is guided by traversals executing in a parallel, distributed manner over the various regions of the graph like unleashed spasms of energy in a lightning storm. In the (sub)graph of atoms and spacetime, these traversals are the "laws of physics" (e.g. quantum and relativity) which expand spacetime, bind atoms into molecules, and gravitationally drive the galaxies into spiraling structures across the firmament. In the social (sub)graph of humanity, traversals identify similar behavioral and consumer patterns in order to unite person- and artifact-vertices via a type of "homophilic gravity" [9, 11]. Traversals are the processes that yield the universal flux and they exist at all scales and in all domains of the universal graph. However, this begs the question: what is the metaphysical status of the traversal with respects to the graph? Where are the traversals manifested? If the universal graph is truly universal, then the answer must be that traversals are topological features (subgraphs) of the universal graph as a true universal graph has nothing "outside" of its vertex and edge sets. This dove tailing of process and structure into one sparks the first of three open problems that must be addressed if there is to be any validity to the reality of the universal graph.

II. TIME: THE PROCESS OF CHANGE

A labeled graph is composed of vertices and edges and is defined as $G = (V, E, \lambda)$, where $E \subseteq V \times V$ and $\lambda : (V \cup E) \rightarrow \Sigma^*$ is a labeling function mapping every vertex and edge to a string.[1] For the sake of notational simplicity, an edge written as the triple (x, y, z) will denote the edge $(x, z) \in E$, where $\lambda((x, z)) = y$. In a

*Rodriguez, M.A.,"Open Problems in the Universal Graph Theory," GraphDay '17, 2(2), pages 1–5, San Francisco, CA, June 2017.

[1] The graph G is a directed, labeled, binary, multi-relational graph. This representation of a graph will be used throughout the remainder of this article. Without loss of generality, there exists a bijective, information preserving function that maps the more complex *property graph* (a vertex/edge-attributed extension of G) to G [4]. Furthermore, there exists a bijective function that takes G to an unlabeled form and a bijective function that takes the unlabeled form to an undirected form [16]. In a lossless manner, the universal graph can be represented simply as "dots and lines."

graph, vertices are typically understood to be the "data" (the matter) and edges, the "relationships" (the space). However, there is something missing from the above definition. There can be no structure without process. How was the graph created? How does it change? These questions are answered by processes known as traversals and their respective traversers. A traversal is composed of a series of functions/steps that map traversers in G from one location to another. In other words, traversals algorithmically direct a flow of traversers over the graph's topology, mutating the graph according to their inscribed program/rules. Viewed from this functional, executing frame of reference, every traversal is "outside" the graph. However, if the universal graph is all that exists, then the traversal must exist "inside" the graph. Graph-encoded traversals (i.e. procedural subgraphs) and their implications have been explored in Ripple [24], grammar-based random walkers [15], NenoFhat [17], and more generally in graph-based virtual machine computing [18].

The Gremlin graph traversal machine will be used to explain the "inside"/"outside"-dilemma more concretely [19]. The Gremlin machine can be defined as

$$(V, E) = G \longleftarrow_{\mu} \; t \in T \; _{\psi}\!\longrightarrow \Psi = (F, S)$$

The above definition has three primary components: the graph G (the data structure), the traversal Ψ (the program), and the traversers T (the central processing units). A traversal is defined as the graph $\Psi = (F, S)$, where F is a set of functions and $S \subseteq F \times F$ links these functions in an acyclic, directed manner. In applied graph computing, for execution efficiency, Ψ is not commingled with G, but instead exists as a program in its own address space (one not composed of vertices and edges, but instead as bytecode instructions in random access memory). However, given that Ψ is a graph in its abstract form, it is possible to encode Ψ as a collection of vertices and edges and thus, encode it as $\Psi \subset G$, where $F \subset V$ and $S \subset E$. At this point, from a computer engineering perspective, both the data memory and the program memory are co-located in G and what remains "outside" the graph are the traversers T (the processing units). Any traverser $t \in T$ has two primary projections: one into the graph denoted $\mu : T \rightarrow (V \cup E)$ (the traverser's location in the graph) and one into the traversal denoted $\psi : T \rightarrow F$ (the traverser's location in the traversal – i.e. its program counter). As such, the traverser lies at the intersection of the description of a process Ψ and its manifestation in the structure G. It is possible to represent a traverser as a vertex $t \in V$ and in doing so, the μ and ψ projections are encoded as edges in E, where (t, μ, v) states that traverser t is at vertex $v \in V$ and (t, ψ, f) states that traverser t is currently executing step $f \in S \subset V$. In summary, both the T and Ψ components of the Gremlin graph traversal machine can be encoded "inside" G – the universal graph.[2]

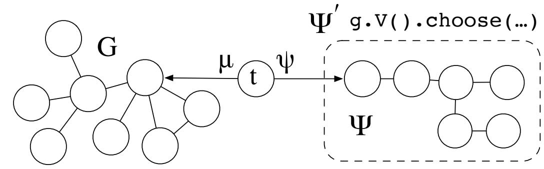

FIG. 1: Ψ encoded as a topological feature of the graph requires Ψ' "outside" of the graph to evaluate it.

In physics, there haven't been any popular suggestions as to where the "laws of physics" are stored – where are the algorithms of quantum and relativity theory [20]? The laws of physics, as presently understood, appear to exist "outside" the natural world impinging their rule in a transcendental manner – altering the movement of energy through spacetime accordingly. Any description of reality defining such a dichotomy entails a more general, encapsulating system to evolve it. In a computer science analog, a virtual machine requires the physical machine to execute it and the physical machine requires the physical world to execute it. What does the physical world require to execute it? In the universal graph story thus far, the "laws of the graph" are the traversals that traverse it. However, as previously demonstrated, these traversals (along with their respective traversers) can be represented "inside" the graph as vertices and edges, but in doing so, they become static structures no longer existing as executing processes, instead existing as representations of processes. For Ψ and T to resume their process-like nature, there must exist another traversal Ψ' and traverser set T' "outside" the graph that execute the graph-encoded Ψ and T (see Figure 1). To execute Ψ, Ψ' moves the graph-encoded traversers of Ψ through the graph by altering their μ-edges as directed by their ψ-edges. To the limit, in a recursive universal Turing machine manner [27], it is possible to then encode Ψ' in the graph, but then there must exist Ψ'' "outside" the graph executing Ψ' which in turn executes Ψ. The first open problem can be framed as such:

> How can the universal graph evolve without requiring processes to exist "outside" itself?

If the universal graph requires an influx of computing power from "outside" itself, then with each new traversal (Ψ) there must exist another traversal that follows to execute it (Ψ') so forth (Ψ'') and so on (Ψ''') *ad infinitum* (Ψ^{∞}). This would entail that the universal graph is growing as processes (traversals) continually bide for a structural (graph) representation in order to be within

[2] See "A Gremlin Implementation of the Gremlin Traversal Machine" at `https://www.datastax.com/dev/blog/a-gremlin-implementation-of-the-gremlin-traversal-machine`.

the vertex/edge address space of the universal graph [12]. However, while an ever expanding universal graph ensures, at the mathematical limit, that all aspects of reality are "inside" the graph, there is still a dichotomous-threshold where a-structural traversals rush in to fill the computational vacuum.[3]

III. THOUGHT: THE STRUCTURE OF EXPERIENCE

There is a flying toaster. This idea exists. Let us call M the materialist graph whose vertices are the quanta of spacetime and mass-energy. There are two types of edges in M, those that denote how spacetime is linked into a malleable lattice-like network and those that specify where each quanta of mass-energy is on that lattice [21]. The materialist philosophy of modern science places M as the foundational structure of reality. This philosophy would argue that the flying toaster (as qualia) is an emergent epiphenomenon of a neurotransmitter-facilitated propagation of energy through the inferior temporal gyrus, converging on the neurally-encoded concepts of "flying" and "toasters" [14]. The distributed, synchronous stimulation of these conceptual neuronal regions elicits the subjective experience of a flying toaster [23]. To date, the explanation ends there. The universal graph theory may take the explanation further. The theory agrees that the materialistic graph M does exist, but posits that M is a subgraph of G (i.e. $M \subset G$). Moreover, the qualia of a flying toaster must exist somewhere as it can be addressed (e.g. discussed and experienced). In which address space does it exist? According to the theory of the universal graph, the flying toaster experience is a vertex in $Q \subset G$ (a qualia graph). The flying toaster vertex in Q is adjacent to the synchronous neural vertices in M.[4] While the flying toaster is "outside" the graph M (as a semi-disjoint subgraph), the flying toaster is not "outside" the universal graph, but within the universal graph and able to be referenced and traversed as such – obviously given one's current "mind's eye"-view of a flying toaster. The second open problem can be framed as:

> How does the universal graph encode the qualia of thought as vertices and edges?

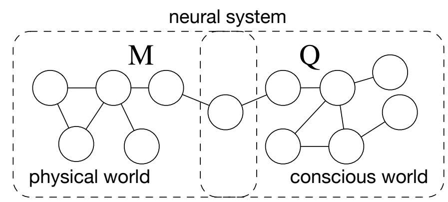

FIG. 2: The material graph M references the qualia graph Q via the brain.

Why can't the experience of a flying toaster be objectively seen in physical reality (M)? The short answer: physical reality is only one particular induced subgraph of the universal graph. While the human has an objective form in M, it also has a subjective form in Q and it is only through Q that M is ever known (see Figure 2). A "thought process" in M looks like a cascade of neural activity but to the human, a "thought process" in Q is much more rich – the qualia of consciousness. While there are many graph models for neural activity (e.g. neurons, axons, and action potentials), there are no known graph models of consciousness. Therefore, the open problem questions how experience is encoded – what are the quanta (vertices) of experience and how do they relate to each other (edges) within the universal graph?

IV. CREATIVITY: ALTERING THE STRUCTURE OF PROCESS

If the universal graph is all that there is, then our shared world's M-based "laws of physics"-traversal Ψ exists somewhere in the graph as vertices and edges. Such "laws of physics" must structurally exist because they ultimately have a reference to the atomic elements of spacetime (e.g. via a μ-edge from a graph-encoded traverser $t \in V$). Unable to manipulate the structural description of the "laws of physics," humans have instead manipulated M-encoded physical objects in order to have the "laws of physics"-traversal yield them utility (i.e. technology). However, given the "laws of physics"-traversal's structural form, it is theoretically possible (given the right technology) to alter its graph-encoding by adding and removing vertices and edges accordingly. This technology would be a mutating Ψ'-traversal tool with a $t' \in V$ μ-reference to the "laws of physics"-traversal subgraph, where $\Psi' \subset M$ (see Figure 3). Such process-aware feedback is known as *reflection* in computer science, whereby a program manipulates its executing self in order to, at runtime, alter its behavior [8]. In this

[3] This letter could have been entitled "Open Problems in the Universal Traversal Theory." It is possible to create a traversal (program) that does not operate on an existing graph, but instead generates the universal graph by adding vertices and edges to form space and its contents. However, such a "Big Bang"-traversal needs to be represented somewhere in order to exist and that requires some amount of structure (i.e. the traversal program – Ψ). Structure requires process to be a *dynamic form* and process requires structure to be a *static description*. This yin-yang is The TinkerPop.

[4] An argument to this statement is that the neural synchronous process is itself the flying toaster qualia. However, given that process has a structural encoding in the universal graph, then that neural synchronous process is a collection of vertices and edges and given the correlation between neural activity and conscious experience, those process vertices must be (implicitly) adjacent to the respective neuronal "grandmother cells" of "flying" and "toaster" [3].

way, it would be possible to rewrite the laws of physics and thus, alter the way in which the objective, manifest subgraph M evolves. The third and final open problem is therefore stated as such:

> If process is as malleable as structure, then how does one get reference to the "laws of physics" in order to alter the evolution of the physical world?

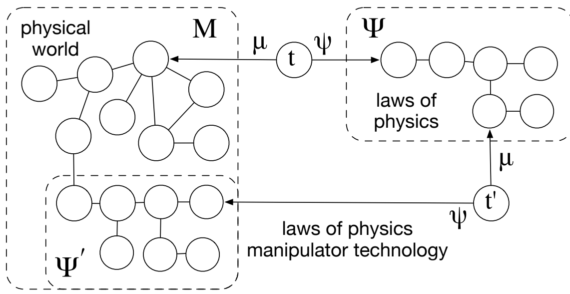

FIG. 3: The structural representation of the "laws of physics" being manipulated by a traversal technology in the physical world.

The notion that *everything* is simply composed of "dots and lines" leads to an age old question: why does the universal graph have this form and not another? Theoretical physics attempts to describe the flux of reality by means of equations. These equations, called theories, are based on closed world structures (e.g. numbers and matrices) and processes (e.g. algebraic manipulations). Mathematicians, on the other hand, are more interested not in the theories that govern reality, but in theories themselves and their respective entailments [7]. It is possible to define a set of axioms and then deduce theorems from them regardless if those axioms reflect aspects of the physical world. Many such explorations lead to logical inconsistencies and thus, are neither sound nor complete. However, some theories lead to harmonious, self-consistent worlds. The universal graph theory posits that such theoretical worlds exist. It is through the mathematician's Q-mind that a reference to these regions of the universal graph are bridged. A new theory/traversal/"laws of physics" could be created and if it were sound and complete, then it could be the means by which the evolution of M's Ψ is re-programmed – via the creativity of Q manifested in Ψ' [5].

A related problem is whether "emergent laws" that govern the evolution of large structures such as humans with social ties can be constructed from more "fundamental laws" that govern smaller structures such as particles propagating through spacetime. Thus, much like atoms compose to create people, once structurally-encoded, the "laws of physics"-traversals may more elegantly compose to create "social dynamics"-traversals (an all-encompassing elegance currently not witnessed across the seemingly disparate disciplines of science, art, and the humanities [26]). In fact, network science has identified similar network-growth patterns in numerous natural systems from the small- to the large-scale (e.g. preferential attachment) [10]. These fractal-like graph traversals may one day be realized as the general "scale-free laws of nature" executing at all (subgraph) scales and semantics within the universal graph. Such a generally applicable description of the evolution of the universe may lead to the discovery of their structural location in G and ultimate creative manipulation.

V. CONCLUSION

The universal graph is difficult to describe because it is everything that is describable. However, through the fields of graph theory and computing, a new world is emerging composed solely of dots and lines. Advances are being made by various engineering endeavors to provide graph technologies that enable the storage and processing of massive-scale graphs distributed across multi-machine compute clusters. Moreover, in application, graphs are being used to model numerous aspects of physics, chemistry, biology, electronics, computing, engineering, commerce, the social sciences and more. The more graphs pervade our descriptions and technologies of the world, the better fit we will be to expand our understanding of reality and in turn, alter it in ways consciousness has never dreamed.

Acknowledgments

This letter represents the author's current boundaries of thought after many years of exploring the wonderful world of graphs with his colleagues in academia (Vrije Universiteit Brussel and the Los Alamos National Laboratory), the open source engineering space (Apache TinkerPop and related software projects), and the commercial sector (DataStax and all the other businesses supporting this fascinating area of research and technology). More specifically, Stephen Mallette, Vadas Gintautas, Joshua Shinavier, Ulises Cerviño Beresi, and Clément Vidal provided valuable feedback on initial draft versions of this letter.

[1] A. Einstein. The foundation of the general theory of relativity. *Annalen der Physik*, 49(7):769–822, 1916.

[2] J. Fodor. *The Language of Thought.* Harvard University Press, 1975. ISBN 0-674-51030-5.

[3] C. G. Gross. Genealogy of the "grandmother cell". *The Neuroscientist*, 8(5):512–518, 2002. doi: 10.1177/107385802237175.

[4] O. Hartig. Reconciliation of RDF* and property graphs. Technical report, University of Waterloo, 2014. URL http://arxiv.org/abs/1409.3288.

[5] W. Kelvin, J. Joule, and V. Regnault. *On the Dynamical Theory of Heat: With Numerical Results Deduced from Mr. Joule's Equivalent of a Thermal Unit and M. Regnault's Observations on Steam.* Society, 1851.

[6] D. Krioukov, M. Kitsak, R. Sinkovits, D. Rideout, D. Meyer, and M. Boguñá. Network cosmology. *Nature Scientific Reports*, 2(793), 2012. URL https://www.nature.com/articles/srep00793.

[7] S. Mac Lane. *Categories for the Working Mathematician.* Springer, 1998.

[8] P. Maes. *Computational Reflection.* PhD thesis, Vrije Universiteit Brussel, Brussels, Belgium, January 1987.

[9] M. McPherson, L. Smith-Lovin, and J. Cook. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27:415–444, 2001.

[10] M. Newman, A.-L. Barabasi, and D. J. Watts. *The Structure and Dynamics of Networks.* Princeton University Press, May 2006.

[11] M. E. J. Newman. Assortative mixing in networks. *Physical Review Letters*, 89(20):208701, 2002. URL http://arxiv.org/abs/cond-mat/0205405/.

[12] P. J. E. Peebles and B. Ratra. The cosmological constant and dark energy. *Review of Modern Physics*, 75:559–606, April 2003. doi: 10.1103/RevModPhys.75.559. URL https://arxiv.org/abs/astro-ph/0207347.

[13] R. Rado. Universal graphs and universal functions. *Acta Arithmetica*, 9:331–350, 1964.

[14] M. Riesenhuber and T. Poggio. Hierarchical models of object recognition in cortex. *Nature Neuroscience*, 2:1019–1025, 1999.

[15] M. A. Rodriguez. Grammar-based random walkers in semantic networks. *Knowledge-Based Systems*, 21(7):727–739, 2008. doi: 10.1016/j.knosys.2008.03.030. URL http://arxiv.org/abs/0803.4355.

[16] M. A. Rodriguez. Mapping semantic networks to undirected networks. *International Journal of Applied Mathematics and Computer Science*, 5(1):39–42, February 2008. URL http://arxiv.org/abs/0804.0277.

[17] M. A. Rodriguez. *Emergent Web Intelligence: Advanced Semantic Technologies*, chapter General-Purpose Computing on a Semantic Network Substrate, pages 57–104. Advanced Information and Knowledge Processing. Springer-Verlag, June 2010. ISBN 78-1-84996-076-2. URL http://arxiv.org/abs/0704.3395.

[18] M. A. Rodriguez. The RDF virtual machine. *Knowledge-Based Systems*, 24(6):890–903, August 2011. URL http://arxiv.org/abs/0802.3492.

[19] M. A. Rodriguez. The Gremlin graph traversal machine and language. In *Proceedings of the 15th Symposium on Database Programming Languages*, DBPL 2015, pages 1–10. ACM, October 2015. ISBN 978-1-4503-3902-5. doi: 10.1145/2815072.2815073. URL http://arxiv.org/abs/1508.03843.

[20] M. A. Rodriguez and J. H. Watkins. Quantum walks with Gremlin. In L. Bender, editor, *Proceedings of the GraphDay Conference*, volume 1, pages 1–16, Austin, Texas, January 2016. URL https://arxiv.org/abs/1511.06278.

[21] C. Rovelli and L. Smolin. Spin networks and quantum gravity. *Physical Review D*, 52:5743–5759, 1995. doi: 10.1103/PhysRevD.52.5743. URL https://arxiv.org/abs/gr-qc/9505006.

[22] D. E. Rumelhart and J. L. McClelland. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition.* MIT Press, July 1993. ISBN 0262181231.

[23] M. N. Shadlen and J. A. Movshon. Synchrony unbound: A critical evaluation of the temporal binding hypothesis. *Neuron*, pages 67–77, 1999.

[24] J. Shinavier. Functional programs as Linked Data. In *3rd Workshop on Scripting for the Semantic Web*, Innsbruck, Austria, 2007.

[25] J. F. Sowa. *Principles of Semantic Networks: Explorations in the Representation of Knowledge.* Morgan Kaufmann, San Mateo, CA, 1991.

[26] R. P. Taylor. Order in Pollock's Chaos. *Scientific American*, December 2002.

[27] A. M. Turing. On computable numbers, with an application to the entscheidungsproblem. *Proceedings of the London Mathematical Society*, 42(2):230–265, 1937.

[28] P. Wang. *Rigid Flexibility: The Logic Of Intelligence.* Springer, 2006.

[29] S. Wasserman and K. Faust. *Social Network Analysis: Methods and Applications.* Cambridge University Press, Cambridge, UK, 1994.