P^3 — Partition, Pivot, and Prune: Aggregated Semantic Graphs for Analyzing Medical Terminologies

Sigfried Gold, MFA, MA[1], Harry Hochheiser, PhD[2],
Richard D. Boyce, PhD[2], Niklas Elmqvist, PhD[1]
[1]University of Maryland, College Park, MD, USA; [2]University of Pittsburgh, PA, USA

Motivation.   A *controlled medical terminology* is a systematically organized and machine-readable collection of medical concepts, terms, and codes used to record clinical information, such as observations, diagnoses, treatments, pharmaceuticals, and tests. These terminologies—ICD-10-CM, LOINC, RxNorm, SNOMED-CT—are central to electronic health records (EHRs) for computer-supported recording of clinical patient data, and have become a cornerstone of health system interoperability. As a result, medical terminologies are not only employed during the practice of care by clinicians, but they are also used by clinical researchers to define *in silica* virtual cohorts from large EHR databases for the purpose of conducting retrospective studies. However, difficulties in forming queries across multiple, complex medical terminologies present substantial barriers to achieving complete and accurate cohort definitions.

Comprehensive terminology systems, like the UMLS Metathesaurus or the OHDSI vocabulary system, bring together multiple, overlapping, sometimes redundant vocabularies, along with crosswalks between them, into collections for general analytic use. Users of a comprehensive terminology system must navigate a forbiddingly large and complex space of terms and relationships to form virtual cohorts. Existing approaches to this problem either force the user to rely on their own knowledge or on published cohort definitions, or to manually traverse the hierarchy using rudimentary search and filter operations based on textual and tabular interfaces. Although visual approaches such as node-link diagrams have been tried for semantic navigation, they have generally been abandoned due to visual clutter. Our goal is to provide clinical researchers with better support for navigating the entire space of terms and relationships.

Contribution.   We propose a dynamic graph aggregation interface method for navigating large-scale semantic graphs representing collections of linked medical terminologies. We use a standard node-link diagram representation,[1] where nodes are rendered as labeled bubbles and links are rendered as arrows. Named P^3 for Partition, Pivot, and Prune, our method starts with a single aggregate *supernode* and superlink representing the entire graph. The P^3 philosophy is that individual concepts and relations do not matter except in aggregate until the analyst drills down on specific local neighborhoods in the semantic graph. A sequence of *partition* operations will split the supernode into constituent parts based on intrinsic or computed attributes. The resulting supernodes can then be arranged into a table using a *pivot* operation where supernodes in the same row and column share the same attribute values, and superlinks connect supernodes in adjacent table cells. Finally, the *prune* operation provides a filtering capability to hide or delete data not relevant to the task. For example, using P^3 with a patient EHR database, a SNOMED-CT graph could be partitioned based on the "causes" relationship for bone fractures, pruned to only include the causes, and partitioned again based on patient age. Pivoting the result in a table using rows for fracture causes and columns for age group will give a visual breakdown of how geriatric patients often break their hips when merely falling in their homes or everyday lives due to osteoporosis, whereas younger people generally only break their legs in more serious accidents.

Related Work.   Our work generalizes the PivotGraph technique[2] for aggregated multivariate graphs by enabling partitioning and pivoting on any number of attributes (rather than just two) using a hierarchical table. Additional visualization techniques for multivariate graphs exist[1]; our work is mostly related to GraphDice,[3] which uses a similar attribute-based layout, as well as Matrix Cubes,[4] which enables slicing graphs as stacked adjacency matrices.

References

[1] von Landesberger T, Kuijper A, Schreck T, Kohlhammer J, Van Wijk JJ, Fekete JD, et al. Visual Analysis of Large Graphs: State-of-the-Art and Future Research Challenges. Computer Graphics Forum. 2011;30(6):1719–1749.

[2] Wattenberg M. Visual exploration of multivariate graphs. In: Proceedings of the ACM Conference on Human Factors in Interactive Computing Systems. ACM; 2006. p. 811–819.

[3] Bezerianos A, Chevalier F, Dragicevic P, Elmqvist N, Fekete JD. GraphDice: A System for Exploring Multivariate Social Networks. Computer Graphics Forum. 2010;29(3):863–872.

[4] Bach B, Pietriga E, Fekete JD. Visualizing Dense Dynamic Networks with Matrix Cubes. In: Proceedings of the ACM Conference on Human Factors in Interactive Computing Systems. ACM; 2014. p. 877–886.