P³ — Partition, Pivot, and Prune:

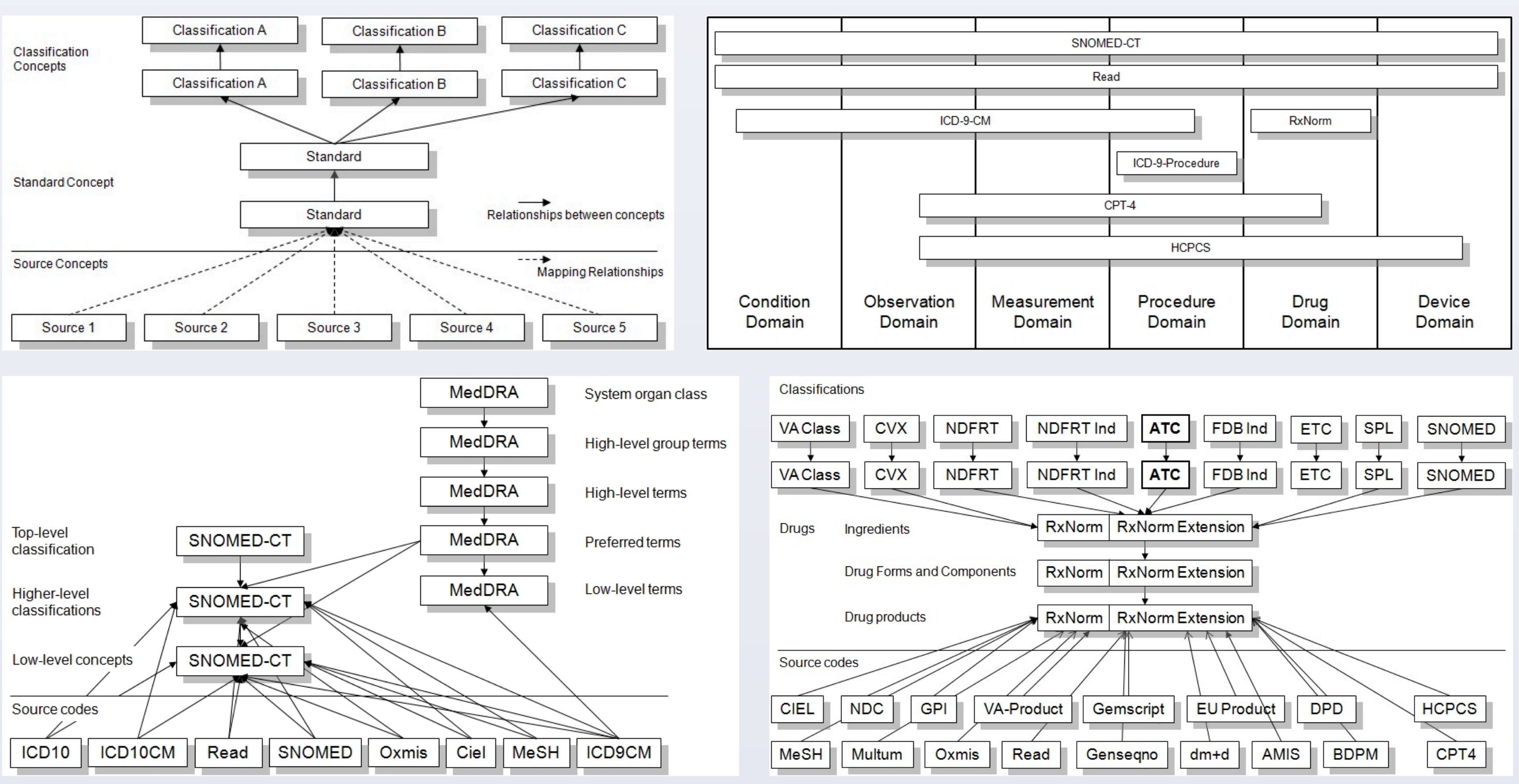
Aggregated Semantic Graphs for Analyzing Medical Terminologies

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

Integrated Vocabulary Collections for Coded Data

Comprehensive terminology collections, like the UMLS Metathesaurus or the OHDSI [2018] 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.

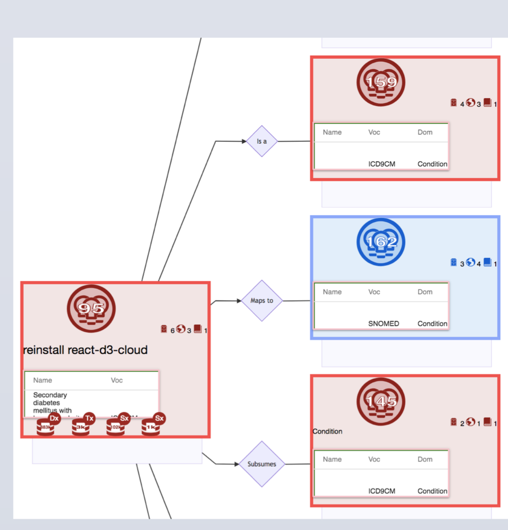


A default download of the OHDSI terminology system includes: 43 vocabularies, 40 domains, 4 million concepts, 5.5 million synonyms, 306 concept classes, 427 relationship types, 27.5 million concept relationships, 121 million concept ancestor relationships. The problem of designing useful, intuitive interfaces for navigating sprawling collections of complex, interconnected medical terminologies is hard.

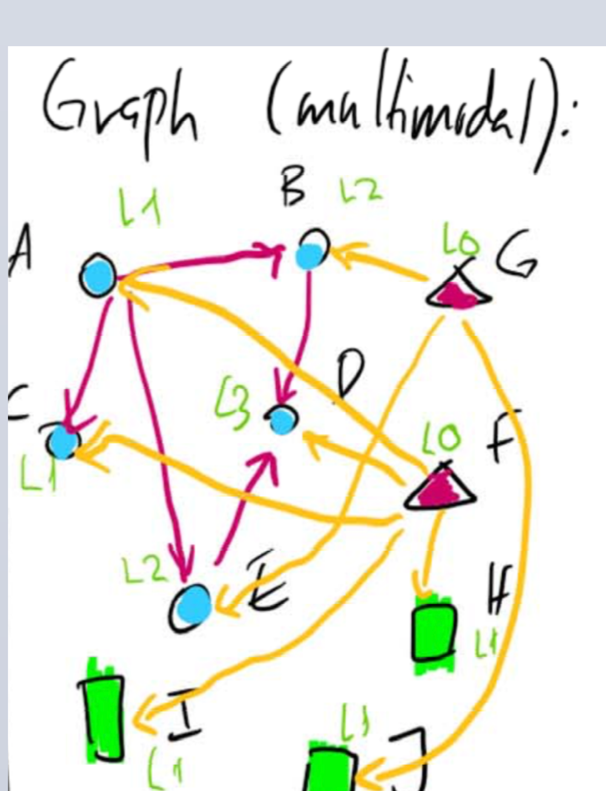
Graph Navigation and Semantic Networks

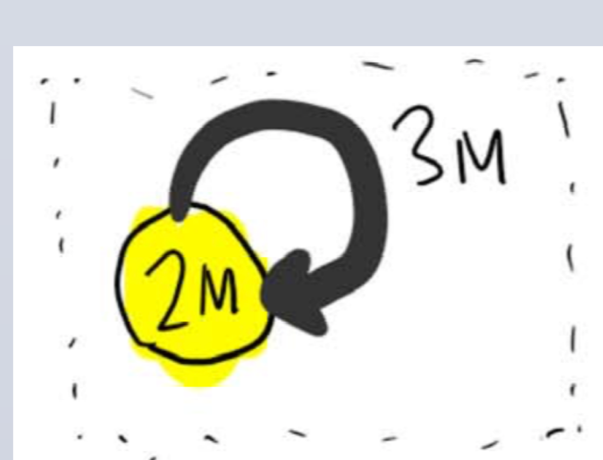
Generalized visualization approaches to graph or network navigation assume that the graph itself contains the information of interest. In the case of semantic networks like the OHDSI terminology system, the semantic graph holds complex metadata describing the coded clinical data actually being analyzed. The graph is present to link the clinical knowledge of analysts to the categories and terms embedded in the data by a multitude of people, organizations, and processes (standards bodies, clinicians, informaticians, software engineers, medical coders, billing procedures, etc.) throughout the production and capture of the data and the development of the infrastructures that supported it

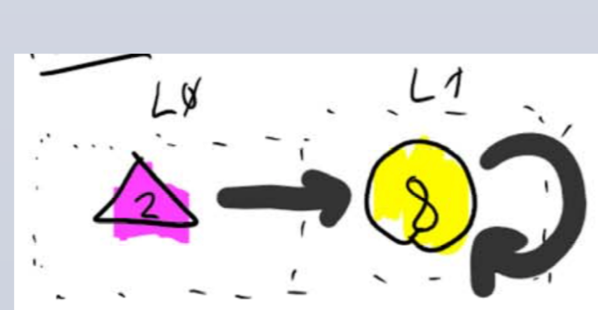
On one hand, this seems a sufficiently specific use case to justify custom design efforts. Our work began with this approach, as shown to the right. It became clear, however, that we would need new techniques for visual aggregation and navigation of graph data building on work from the visualization community, not specific to our use case. Our current designs use a standard node-link diagram representation, [vonLandesberger 2011] where nodes are rendered as labeled bubbles and links are rendered as arrows.

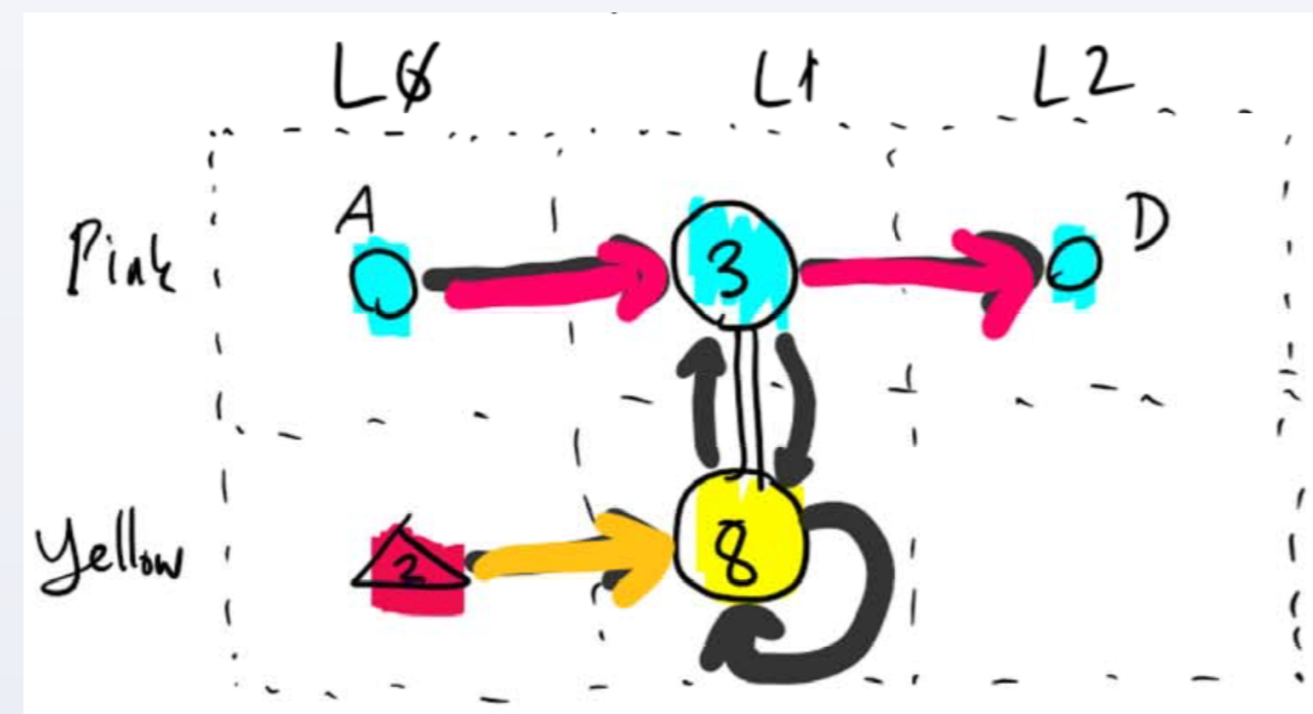


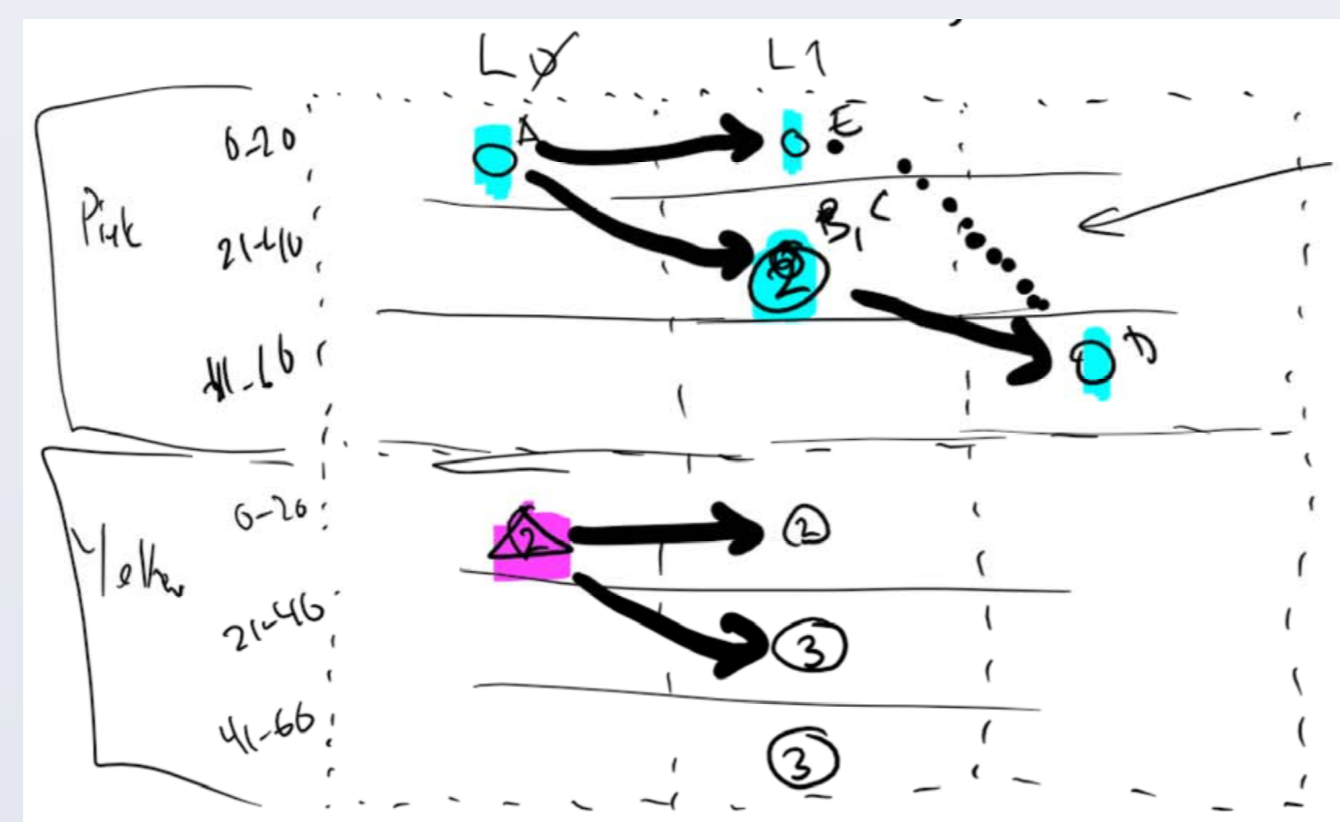
Named P³ for Partition, Pivot, and Prune, our method starts with a single aggregate *supernode* and superlink representing the entire graph. The P³ philosophy is that individual concepts and relations do not matter except in aggregate until the analyst drills down on specific local neighborhoods in the 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.

Our work generalizes the PivotGraph technique [Wattenberg 2006] 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 [von Landesberger 2011]; our work is mostly related to GraphDice, [Bezerianos 2010] which uses a similar attribute-based layout, as well as Matrix Cubes, [Bach 2014] which enables slicing graphs as stacked adjacency matrices.

References

1. Bach B, Pietriga E, Fekete J-D. Visualizing dynamic networks with matrix cubes. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems [Internet]. ACM; 2014 [cited 2018 Oct 24]. p. 877–86. Available from: https://dl.acm.org/citation.cfm?doid=2556288.2557010
2. Bezerianos A, Chevalier F, Dragicevic P, Elmqvist N, Fekete JD. GraphDice: A System for Exploring Multivariate Social Networks. Comput Graph Forum [Internet]. 2010 Aug 12;29(3):863–72. Available from: http://doi.wiley.com/10.1111/j.1467-8659.2009.01687.x
3. Munroe R. Standards [Internet]. xkcd. Jul 20, 2011 [cited 2017 Sep 22]. Available from: https://xkcd.com/927/, https://imgs.xkcd.com/comics/standards.png
4. OHDSI. Standard, Classification and Source Concepts [Internet]. Observational Health Data Sciences and Informatics. [cited 2018d Oct 29]. Available from: http://www.ohdsi.org/web/wiki/doku.php?id=documentation:vocabulary:standard_classification_and_source_concepts
5. von Landesberger T, Kuijper A, Schreck T, Kohlhammer J, van Wijk JJ, Fekete J-D, et al. Visual Analysis of Large Graphs: State-of-the-Art and Future Research Challenges. Comput Graph Forum [Internet]. 2011 Sep 21;30(6):1719–49. Available from: http://doi.wiley.com/10.1111/j.1467-8659.2011.01898.x
6. Wattenberg M. Visual exploration of multivariate graphs. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems [Internet]. ACM; 2006 [cited 2018 Oct 24]. p. 811–9. Available from: https://dl.acm.org/citation.cfm?doid=1124772.1124891

Contact

Sigfried Gold, sigfried@sigfried.org

UNIVERSITY OF MARYLAND COLLEGE OF INFORMATION STUDIES

HCIL

COMBINE