

Hypergraphs as a Conceptual Model for Mediating Translations between Data Models

Aruna Bansal (2018ANZ8379)

Abstract of the thesis

Hypergraphs are rich candidate structures that provide a natural mathematical way to represent the richness of diversified data and complex relationships in hierarchical, relational, navigational, and semi-structured settings, e.g., bibliographic paper submissions, mHealth, and social media applications. The representation and modeling of information in terms of entities and relationships are challenging in hypergraphs. Further, representing and formulating queries on such hypergraphs, and obtaining results for those queries are other challenges. Some hypergraph-oriented representations and data models are mentioned in the literature, but they do not pay enough attention to entities and relationships. Additionally, the existing approaches work with low-level models and do not unify the essence of other data models.

In this thesis, a *Hypergraph Data Model* (or *HgDM*) proposes a notion of *schematically well-formed typed nested* hypergraphs that integrates the benefits of various data models for representing data, metadata, and complex relationships. The HgDM is formally presented as a high-level hypergraph data model that can be directly represented in object-oriented frameworks. The queries are considered diagrams of ideas, and objects in a database are matched to the diagrams.

Hypergraphs have a trivial embedding from other data models, such as hyperedges can directly represent relations of a relational model, edges of graphs and tree structures, and objects of object-oriented databases. However, generalizing hypergraphs in other data models and vice versa in a lossless way is another challenge for which the thesis proposes a solution by presenting translation patterns. The thesis proposes *Hypergraph Mediator* (or *HgMed*) for hypergraph-based mediation to other data models. The HgMed presents a notion of simulation relations between schematic diagrams of different data models modulo path morphisms. On the basis of this, the correctness of representations and translations between schema and schematic queries can be determined. In order to preserve structural properties that may be otherwise lost in a translation, we use annotations that help compensate for the lossiness of translations.

Furthermore, hypergraph-oriented applications and databases may operate in distributed environments with a requirement for availability while coping with high

network latencies. Replication is the commonly used approach to achieving a high degree of availability, facilitating local query processing. However, replication requires expensive (often infeasible) concurrency control to ensure consistency. Moreover, administrative and security policies may prohibit certain parts of the database from being fully replicated at certain sites. The thesis proposes a solution to fully or partially replicate hypergraphs across multiple replicas as a data type in conformance with certain distribution policies while ensuring data availability, network latency, and a weak notion of consistency. In this direction, the thesis presents *Hypergraphs as Conflict-free Replicated Data Type* (or *HgCRDT*), and *Hypergraphs as Conflict-free Partially Replicated Data Type* (or *HgCPRDT*) that ensure strong eventual consistency.

Apart from these works, the thesis proposes and presents *Hypergraph-oriented GraphQL* (or *HGQL*) that semantically and syntactically integrates the richness of hypergraphs in GraphQL (a query language used for APIs). Traditional GraphQL supports graphs and hierarchical structures. The thesis enhances the functionality of GraphQL by embracing the richness of hypergraphs so that GraphQL can be employed with hypergraph-oriented databases.

Additionally, the thesis presents details of a prototype implementation system (named *Hypergraph Queriable* system or *HQ*) that demonstrates the applicability of the above approaches.

In brief, the thesis aims at presenting *hypergraphs* as a unifying conceptual data model within which different data models (i.e., hierarchical, graph, and relational data models) can be expressed, keeping most of their strengths while avoiding most of their shortcomings. The thesis also aims at leveraging this conceptual hypergraph data model to generalize other data models by mediating to and from hypergraphs in a lossless way. The thesis represents hypergraphs as a data type to update information among multiple replicas of a distributed domain. Therefore, the thesis proposes and presents HgDM for hypergraph-based representation and modeling; HgMed for hypergraph-based mediation to other data models; HgCRDT and HgCPRDT for updating hypergraphical information in distributed settings; HGQL for integrating schematic hypergraphs into GraphQL; and HQ, a prototype system that implements our proposed approaches.