GQS: Graph query system for patterns under homomorphism and subgraph isomorphism

Thomas Fannes, Ashraf Kibriya, Jan Ramon

Contact: thomas.fannes@cs.kuleuven.be

Motivation

Homomorphism and subgraph isomorphism are NP-complete

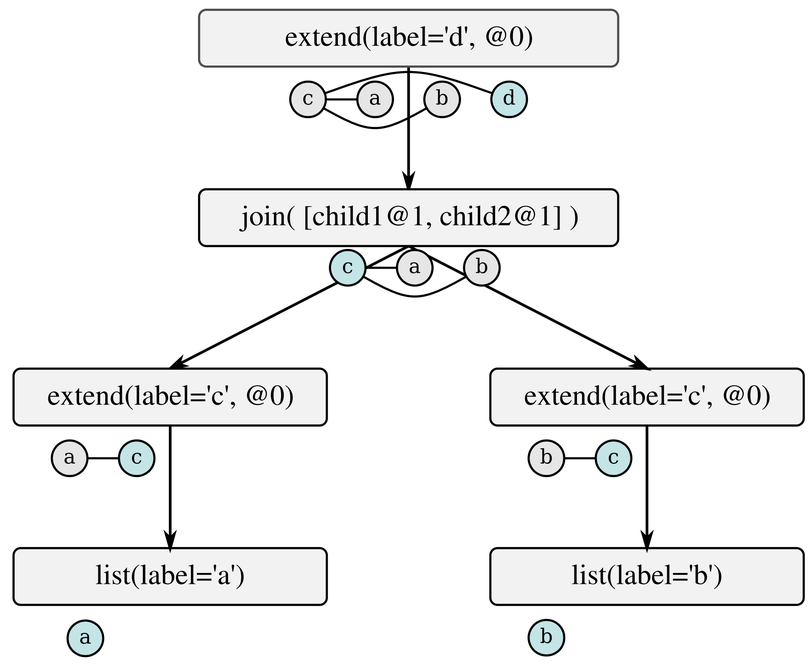
Basic complexity: For a network G and a pattern P of bounded treewidth with treewidth(P) ≤ fixed parameter w:
- homomorphism: polynomial in V(G) and V(P), exponential in w
- subgraph isomorphism: polynomial in V(G), exponential in w, mildly exponential in |V(P)| ~ $O(2^{|V(P)|})$

➡ Efficient query system for homomorphism and subgraph isomorphism

GQS building blocks

Query operator:
- represents a pattern P
- with a target value function T
- works on a network G
- outputs a stream (e, T(e)):
  e is an embedding of P in G ∧ T(e) ≠ 0
- can be based on child query operators

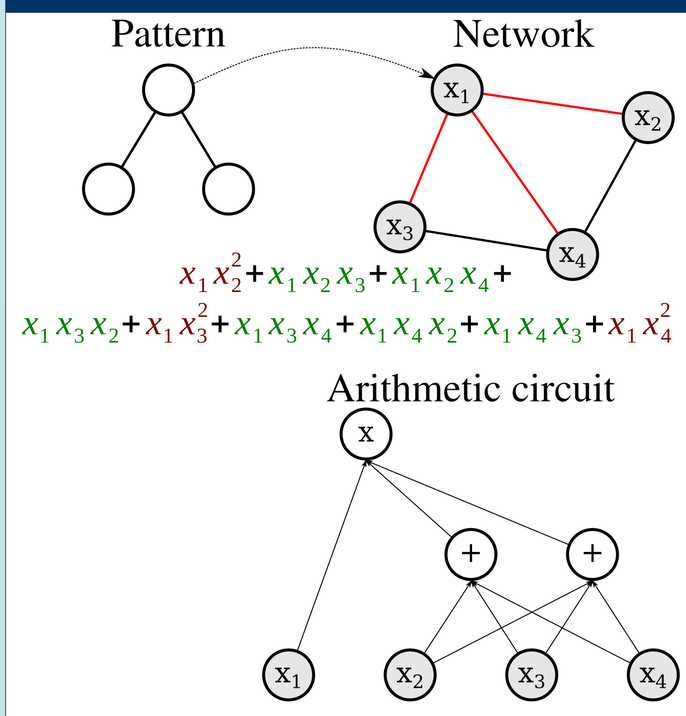
Query Plan: Rooted **directed acyclic graph** of query operators



Query operator categories:
- List operator: list embedding of **vertices** or **edges**
- Extend operator: extends a child operator by **adding a new vertex**
- Select operator: **Filters** the stream of embedding based on a criterium
- Join operator: **Joins** two child operators on a **subset** of their pattern vertices
- Project operator: Removes pattern vertices from output stream → **stream cardinality reduction**

A query plan outputs a stream of embeddings of pattern P under homomorphism

GF2 algebra for subgraph isomorphism



homomorphic embeddings of pattern P are represented with polynomial:
- each embedding a term
- each network vertex v ∈ V(G) → x_v
- pattern can be represented by circuit of size O(|V(G)| |V(P)|)
- non-multilineair term are non-isomorphic

Use $GF(2^l)\mathbb{Z}_2^k$ algebra as target value:
- one embedding ↔ one term
- squares (or higher) are evaluated to zero
- randomized approach:
  T(e) ≠ 0 → isomorphic embedding of P in G
  T(e) = 0 → Pr [no isomorphic embedding of P in G] < δ

with k = |V(P)|, l = 3 + $\log_2 k$

[Koutis I, Williams R (2009) Limits and applications of group algebras for parameterized problems]

GQS properties

- C++ templated system → compile your query
- preliminary tests on runtime efficiency:
  Automatic compiled query ≈ manually written query

In pipeline:
- query plan optimizations
- search query plan space for most efficient query plan

GQS properties

	Tree	**Graph**
homomorphism	default	tree decomposition
Subgraph isomorphism	GF2Algebra	tree decomposition + GF2Algebra

DTAI Declaratieve Talen en Artificiële Intelligentie

erc Starting Grant 240186 MiGraNT

K.U.Leuven
Department of Computer Science