# LDBC SNB on Differential Dataflow

Frank McSherry
Materialize, Inc.
mcsherry@materialize.io

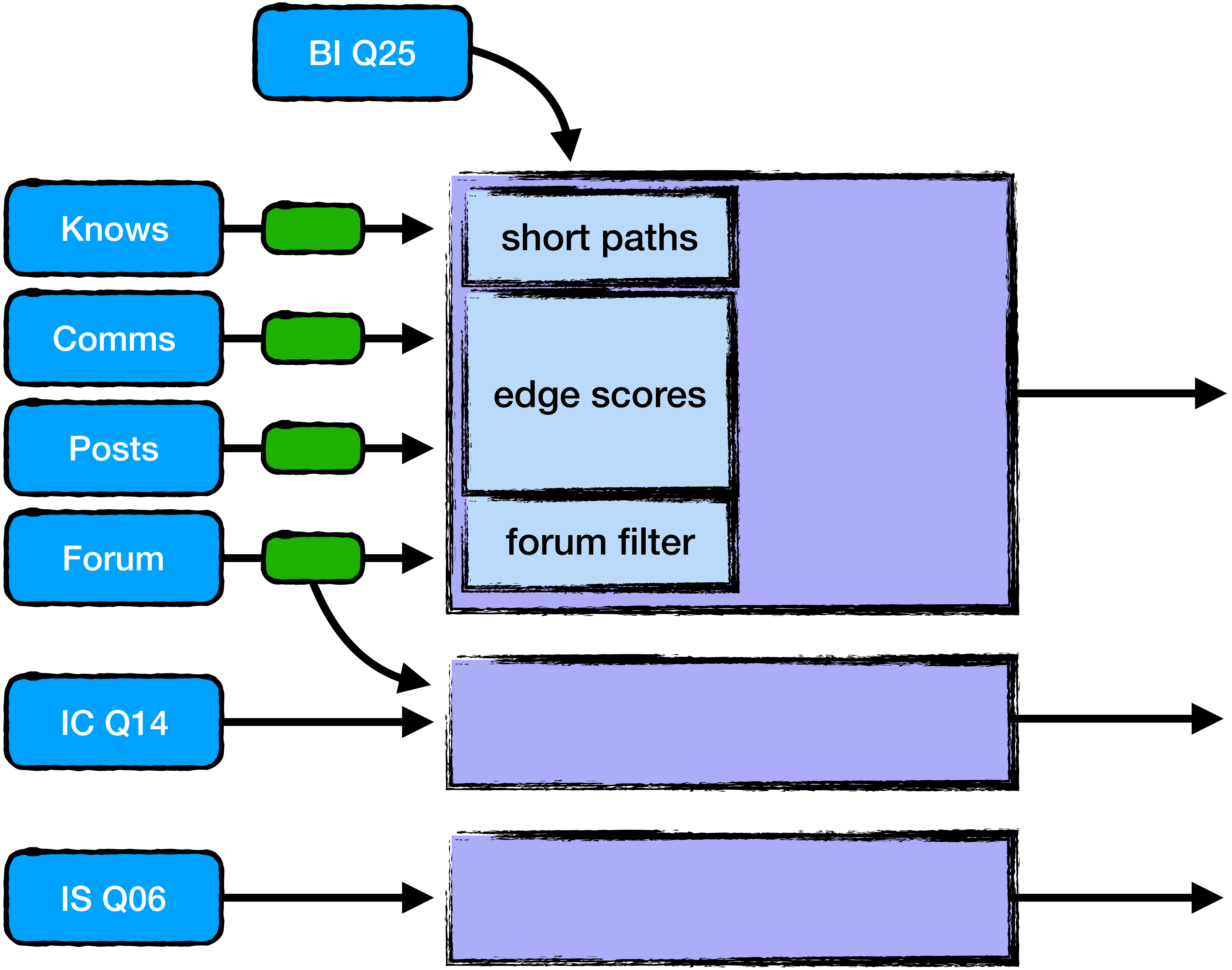
# Differential Dataflow

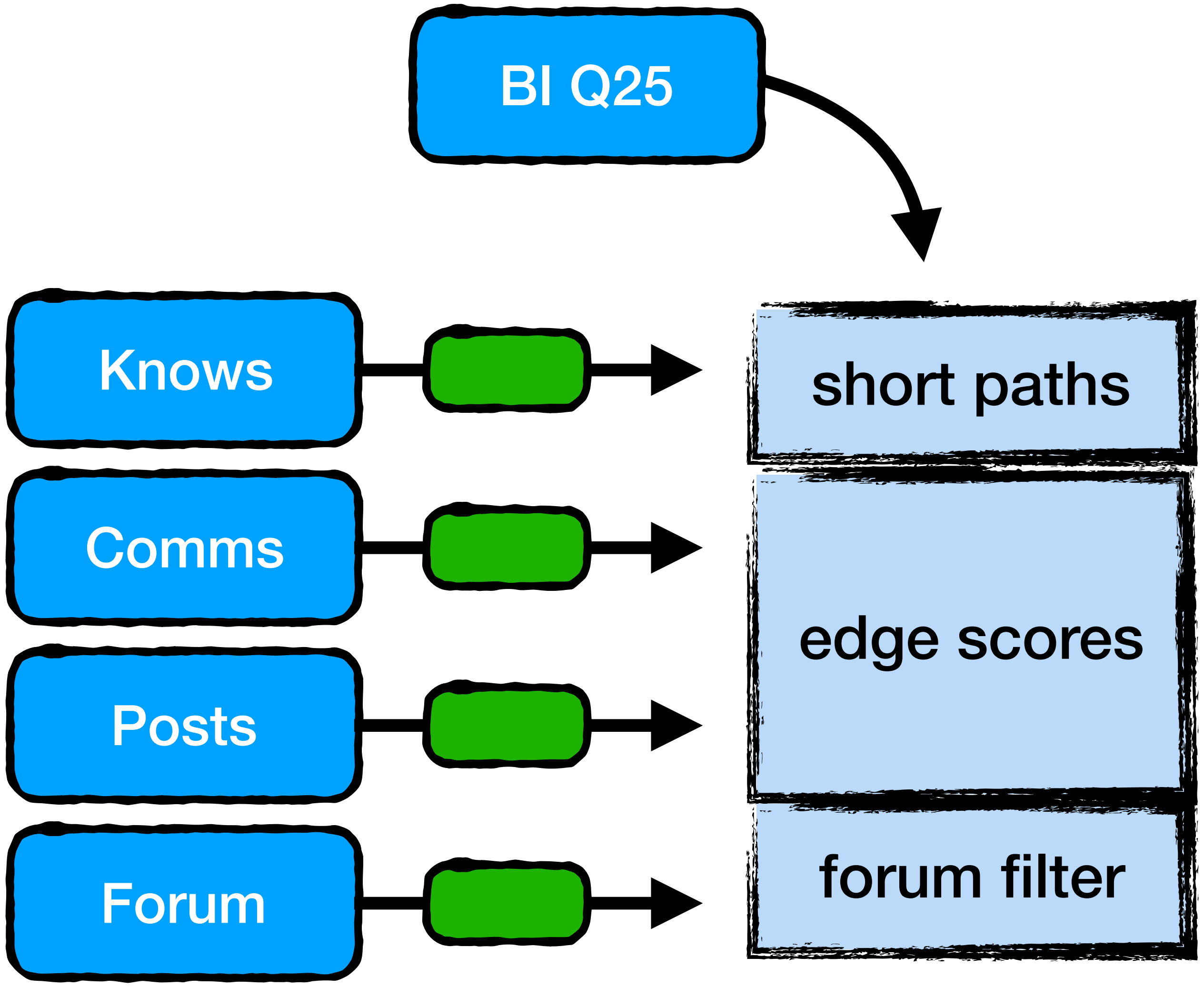
**A collection-oriented eDSL (in Rust) with:**

**Support for many operators: relational, map-reduce, (nested) mutual recursion.**

**Collections are described by streams of timestamped updates, and**

**Engine correctly updates all computations, in time proportional to the trace difference.**


BI Q25
Knows
Comms
Posts
Forum
IC Q14
IS Q06
short paths
edge scores
forum filter



BI Q25

Knows

Comms

Posts

Forum

short paths

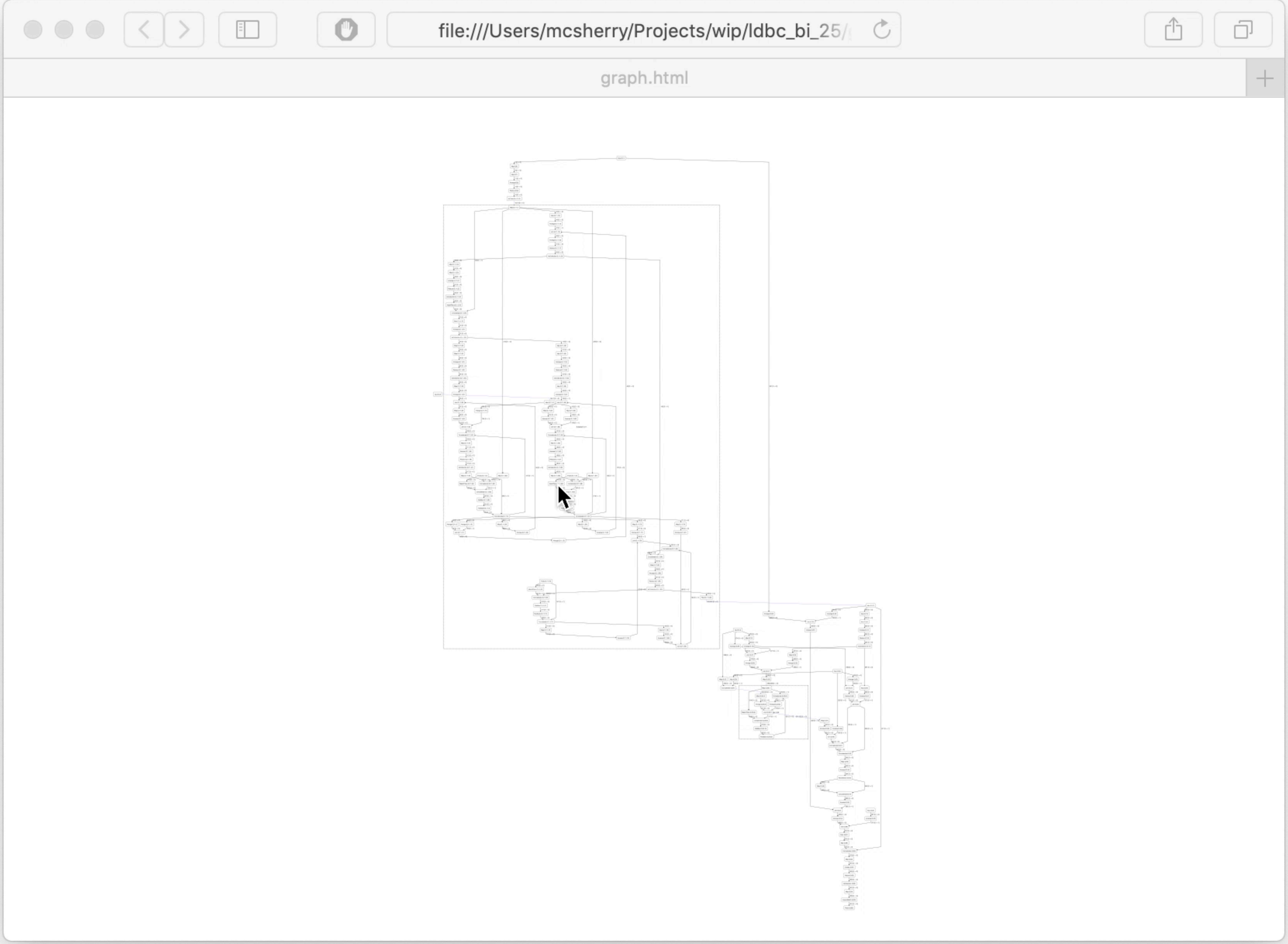
edge scores

forum filter

**125 lines of code (17 ;s)**
**(bidirectional Dijkstra)**

**50 lines of code (5 ;s)**
**(iterative roll-up)**

**13 lines of code (1 ;)**
**(joins, count, probe)**


file:///Users/mcsherry/Projects/wip/ldbc_bi_25/
graph.html

# Performance

**Experiment: Maintain 100 queries with updates (SF1, loaded by CreationDate).**

| Batch size | Elapsed | Refresh | Throughput |
|---|---|---|---|
| **10^3** | 229.490s | 14.37Hz | 14,380 elt/s |
| **10^4** | 109.740s | 3.01Hz | 30,071 elt/s |
| **10^5** | 55.024s | 0.60Hz | 59,973 elt/s |
| **10^6** | 42.780s | 0.07Hz | 77,138 elt/s |
| **10^7** | 38.375s | 0.03Hz | 85,993 elt/s |

# Thoughts

**LDBC has a variety of different queries, which is great (for me) but perhaps intimidating.**

**Should there be reference stats for individual queries, so that folks can attack easier ones?**

**Keep the queries hard, reach for stretch goals.**
**(e.g., millisecond microbatches)**
**(e.g., query maintenance, sharing)**