LHS

Stanford SNAP Dataset

Synthetic

LR LC RC

LD

RD

LNNZ

RNNZ

SPMM Scenario Generation

NNZ < INT_MAX

Yes →

LR × RC < INT_MAX

Yes

Mem usage < Executor_Mem

← Yes

LC × RC < INT_MAX

Yes ↓

Filtering

LR	LC	RC	LD	RD	LNNZ	RNNZ

C

D-optimal
Federov ($Max|D^T D|$)

LR	LC	RC	LD	RD	LNNZ	RNNZ

D

Optimal Scenario Selection

SPMM(D)

Add Feature Combination

LR, LC, RC, LD, RD, LNNZ, RNNZ, LR*LC, LC*RC, LR*RC
LR*LC*RC, LD*RD, LR*RC*LD*RD, LR*LC*RC*LD*RD
LNNZ*RNNZ, Spark SM*DM, Breeze SM*SM

Training Dataset Generation