Appendix

Due to space limit, we present some minute details and relatively less important experimental results and analyses in this appendix.

Table 1: Data schema in the STATS dataset

Relation	Attribute		
users	**Id**	Integer	Primary Key
	Reputation	Integer	
	CreationDate	Timestamp	
	Views	Integer	
	UpVotes	Integer	
	DownVotes	Integer	
posts	Id	Integer	Primary Key
	PostTypeId	Integer	
	CreationDate	Timestamp	
	Score	Integer	
	ViewCount	Integer	
	OwnerUserId	Integer	
	AnswerCount	Integer	
	CommentCount	Integer	
	FavoriteCount	Integer	
	LastEditorUserId	Integer	
postHistory	Id	Integer	Primary Key
	PostHistoryTypeId	Integer	
	PostId	Integer	
	CreationDate	Timestamp	
	UserId	Integer	
comments	Id	Integer	Primary Key
	PostId	Integer	
	Score	Integer	
	CreationDate	Timestamp	
	UserId	Integer	
votes	Id	Integer	Primary Key
	PostId	Integer	
	VoteTypeId	Integer	
	CreationDate	Timestamp	
	UserId	Integer	
	BountyAmount	Integer	
badges	Id	Integer	Primary Key
	UserId	Integer	
	Date	Timestamp	
tags	Id	Integer	Primary Key
	Count	Integer	
	ExcerptPostId	Integer	

A FEATURE OF JOIN PREDICATES

We adapt an existing work [1] to map the join predicates $\boldsymbol{J}$ in the query clause $\mathtt{C}_i$ to the feature vector $\boldsymbol{\gamma}_i$. First, each attribute is uniquely encoded with a binary vector of dimension η. Then a join predicate P with format '$A_l = A_r$' is represented by a 2η-dim binary vector $\boldsymbol{\nu}_P$ with the first and second halves sub-vectors referring to attributes A_l and A_r, respectively. Suppose there are Δ possible join patterns, the join feature $\boldsymbol{\gamma}_i$ of a query clause $\mathtt{C}_i$ is a concatenation of Δ sub-vectors of dimension 2η, indicating which join patterns $\mathtt{C}_i$ captures their referred attributes. If $\boldsymbol{J}$ contains the kth join pattern P_k, the kth sub-vector of $\boldsymbol{\gamma}_i$ equals $\boldsymbol{\nu}_{P_k}$. Otherwise, this sub-vector is set to zero. Also, we treat the attributes in the left and right-hand sides of a join predicate as 'equivalent' and find all equivalence classes in $\boldsymbol{J}$. Afterward, we re-organize the join predicates based on the equivalence classes such that any two equivalent join predicate sets will be featurized to the same vector even if they contain different predicates.

B DETAILS OF THE STATS DATASET

The data schema and information regarding joins among relations in the STATS dataset are presented Table 1 and Fig. 1, respectively. Specifically, details about the relations discussed in Section 5.5 of the submitted paper are highlighted in bold.

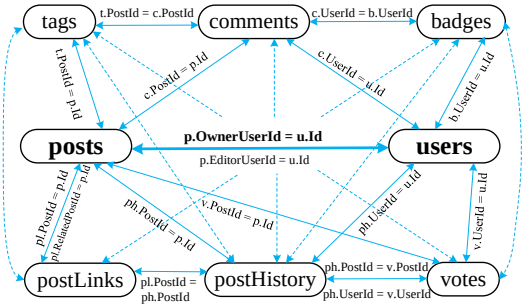

Fig. 1: Joins among relations in the STATS dataset

REFERENCES

[1] Pengfei Li, Wenqing Wei, Rong Zhu, Bolin Ding, Jingren Zhou, and Hua Lu. 2023. ALECE: An Attention-based Learned Cardinality Estimator for SPJ Queries on Dynamic Workloads. *Proc. VLDB Endow.* 17, 2 (2023), 197–210.