IC 1
IC 2
IC 3
IC 4
IC 5
IC 6
IC 7
IC 8
IC 9
IC 10
IC 11
IC 12
IC 13
IC 14

query	Interactive / complex / 1									
title	Trans	Transitive friends with certain name								
pattern				ate	- isLocate «opt work/ «opt study		locationCity: City name company: Company name university: University name	_isLocatedIn →	companyCountry: Country name universityCity: City name	
desc.	necte	Given a start Person, find Persons with a given first name (firstName) that the start Person is connected to (excluding start Person) by at most 3 steps via the knows relationships. Return Persons, including the distance (13), summaries of the Persons workplaces and places of study.								
	1	personId	ID							
params	2	firstName	String							
result	1 otherPerson.id 2 otherPerson.lastName 3 distanceFromPerson 4 otherPerson.birthday 5 otherPerson.creationDate 6 otherPerson.gender 7 otherPerson.browserUsed 8 otherPerson.locationIP 9 otherPerson.email 10 otherPerson.speaks 11 locationCity.name 12 universities 13 companies		String 32-bit Into Date DateTime String String String {Long Str {String}} String { <string} 32-bit="" into="" string="" {<string}="">} {<string>}</string></string}>	ing}	R R R R R R R R R A	{ <university.i companycountry<="" th="" universitycity="" {<company.name=""><th>y.name>} e, workAt.</th><th>yAt.classYear, workFrom,</th></university.i>	y.name>} e, workAt.	yAt.classYear, workFrom,		
sort	1 distanceFromPerson ↑ 2 otherPerson.lastName ↑ 3 otherPerson.id ↑		1							
limit	20									
CPs	2.1, 5.3, 8.2									
relevance	This query is a representative of a simple navigational query. It looks for paths of length 13 through the knows relation, starting from a given Person and ending at a Person with a given first name. It is interesting for several aspects. (1) It requires for a complex aggregation for returning the concatenation of universities, companies, languages and email information of the Person. (2) It tests the ability of the optimizer to move the evaluation of sub-queries functionally dependant on the Person, after the evaluation of the top-k. (3) Its performance is highly sensitive to properly estimating the cardinalities in each transitive path, and paying attention not to explore already visited Persons.									

IC 1	query	Interactive / complex / 2			
IC 2	title	Recent messages by your friends			
IC 3 IC 4 IC 5 IC 6	pattern	person: Person id = \$personId knows id firstName lastName lastName hasCreator creationDate < \$maxDate id content / imageFile creationDate			
IC 9	desc.	Given a start Person (person), find the most recent Messages from all of that Person's friends (friend nodes). Only consider Messages created before the given maxDate (excluding that day).			
IC 10 IC 11 IC 12	params	1 personId ID 2 maxDate Date			
IC 14	result	1 friend.id ID R 2 friend.firstName String R 3 friend.lastName String R 4 message.id ID R message.content or 5 message.imageFile (for Text R photos) 6 message.creationDate DateTime R			
	sort	1 message.creationDate ↓ 2 message.id ↑			
	limit	20			
	CPs	1.1, 2.2, 2.3, 3.2, 8.5			
	relevance	This is a navigational query looking for paths of length two, starting from a given Person, going to their friends and from them, moving to their published Posts and Comments. This query exercices both the optimizer and how data is stored. It tests the ability to create execution plans taking advantage of the orderings induced by some operators to avoid performing expensive sorts. This query requires selecting Posts and Comments based on their creation date, which might be correlated with their identifier and therefore, having intermediate results with interesting orders. Also, messages could be stored in an order correlated with their creation date to improve data access locality. Finally, as many of the attributes required in the projection are not needed for the execution of the query, it is expected that the query optimizer will move the projection to the end.			

IC 1
IC 2
IC 3
IC 4
IC 5
IC 6
IC 7
IC 8
IC 9
IC 10
IC 11
IC 12
IC 13
IC 14

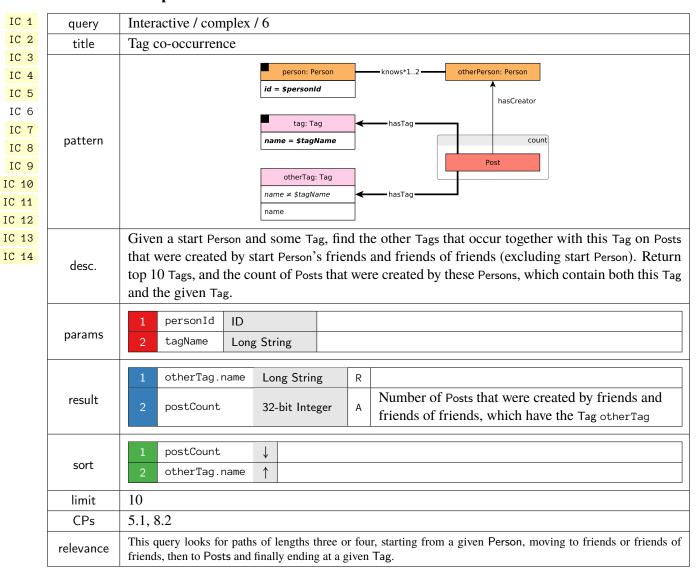
query	Interactive / complex / 3				
title	Friends and friends of friends that have been to given countries				
pattern	xCount = count xCountry isLocatedIn name = \$countryXName spartOf id = \$personId islocatedIn xCountry islocatedIn islocatedIn islocatedIn islocatedIn islocatedIn iname = \$countryY: Country islocatedIn iname = \$countryYName xneg islocatedIn iname = \$countryYName xneg islocatedIn iname = \$countryYName				
desc.	Given a start Person, find Persons that are their friends and friends of friends (excluding start Person) that have made Posts / Comments in both of the given Countries, CountryX and CountryY, within a given period. Only Persons that are foreign to Countries CountryX and CountryY are considered, that is Persons whose location is neither CountryX nor CountryY.				
params	1 personId ID 2 countryXName String 3 countryYName String 4 startDate Date Beginning of requested period 5 durationDays 32-bit Integer Duration of requested period, in days. The interval [startDate + durationDays) is closed-open				
result	1 otherPerson.id ID R 2 otherPerson.firstName String R 3 otherPerson.lastName String R 4 xCount 32-bit Integer A Number of Messages from Country Country X created by the Person within the given time 5 yCount 32-bit Integer A Number of Messages from Country Country Y created by the Person within the given time 6 count 32-bit Integer A count = xCount + yCount				
sort limit	1 xCount ↓ 2 otherPerson.id ↑				
CPs	2.1, 3.1, 5.1, 8.2, 8.5				
relevance	This query looks for paths of length two and three, starting from a Person, going to friends or friends of friends, and then moving to Messages. This query tests the ability of the query optimizer to select the most efficient join ordering, which will depend on the cardinalities of the intermediate results. Many friends of friends can be duplicate, then it is expected to eliminate duplicates and those people prior to access the Post and Comments, as well as eliminate those friends from CountryX and CountryY, as the size of the intermediate results can be severely affected. A possible structural optimization could be to materialize the number of Posts and Comments created by a Person, and progressively filter those people that could not even fall in the top 20 even having all their posts in the Countries CountryX and CountryY.				

IC 1	query	Interactive / complex / 4
IC 2	title	New topics
IC 3 IC 4 IC 5 IC 6 IC 7 IC 8 IC 9 IC 10	pattern	friend: Person wopt> hasCreator postCount = count Post \$startDate \le creationDate \le \$startDate \le \$star
IC 11 IC 12 IC 13	desc.	Given a start Person (personId), find Tags that are attached to Posts that were created by that Person's friends. Only include Tags that were attached to friends' Posts created within a given time interval, and that were never attached to friends' Posts created before this interval.
IC 14	params	1 personId ID 2 startDate Date 3 durationDays 32-bit Integer Duration of requested period, in days. The interval [startDate, startDate + durationDays) is closed-open
	result	1 tag.name Long String R 2 postCount 32-bit Integer A Number of Posts made within the given time interval that have tag
	sort	1 postCount ↓ 2 tag.name ↑
	limit	10
	CPs	2.3, 8.2, 8.5
	relevance	This query looks for paths of length two, starting from a given Person, moving to Posts and then to Tags. It tests the ability of the query optimizer to properly select the usage of hash joins or index based joins, depending on the cardinality of the intermediate results. These cardinalities are clearly affected by the input Person, the number of friends, the variety of Tags, the time interval and the number of Posts.

The LDBC Social Network Benchmark – version 0.4.0-SNAPSHOT (2c2492c)

IC		1	
IC		2	
IC		3	
IC		4	
IC		5	
IC		6	
IC		7	
IC		8	
IC		9	
IC	1	0	
IC	1	1	
IC	1	2	
IC	1	3	
IC	1	4	

query	Interactive / complex / 5				
title	New groups				
pattern	person: Person id = \$personId hasMember				
desc.	Given a start Person, find the Forums which that Person's friends and friends of friends (excluding start Person) became Members of after a given date. For each Forum find the number of Posts that were created by any of these Persons. For each Forum and consider only those Persons which joined that particular Forum after the given date (minDate).				
params	1 personId ID 2 minDate Date				
result	1 forum.title Long String R 2 postCount 32-bit Integer A Number of Posts made in forum that were created by friends				
sort	1 postCount ↓ 2 forum.id ↑				
limit	20				
CPs	2.3, 3.3, 8.2, 8.5				
relevance	This query looks for paths of length two and three, starting from a given Person, moving to friends and friends of friends, and then getting the Forums they are members of. Besides testing the ability of the query optimizer to select the proper join operator, it rewards the usage of indexes, but their accesses will be presumably scattered due to the two/three-hop search space of the query, leading to unpredictable and scattered index accesses. Having efficient implementations of such indexes will be highly beneficial.				

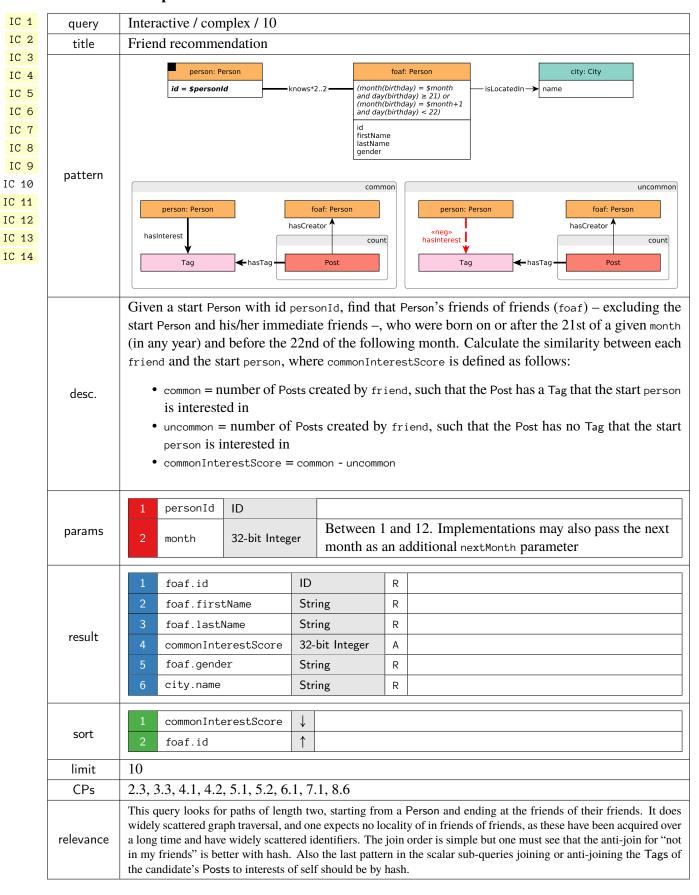


IC 1
IC 2
IC 3
IC 4
IC 5
IC 6
IC 7
IC 8
IC 9
IC 10
IC 11
IC 12
IC 13
IC 14

query	Interactive / complex / 7			
title	Recent likers			
pattern	person: Person id = \$personId hasCreator message: Message id content / imageFile friend: Person id firstName lastName likes creationDate			
desc.	Given a start Person, find the most recentlikes on any of start Person's Messages. Find Persons that liked (likes edge) any of start Person's Messages, the Messages they liked most recently, the creation date of that like, and the latency in minutes (minutesLatency) between creation of Messages and like. Additionally, for each Person found return a flag indicating (isNew) whether the liker is a friend of start Person. In case that a Person liked multiple Messages at the same time, return the Message with lowest identifier.			
params	1 personId ID			
result	1 friend.id ID R 2 friend.firstName String R 3 friend.lastName String R 4 likes.creationDate DateTime R 5 message.id ID R message.content or R R 6 message.imageFile (for photos) Text R 7 minutesLatency 32-bit Integer C Duration between creation of the Message and the creation of the like, in minutes 8 isNew Boolean C False if person and friend know each other, True otherwise			
sort	1 likes.creationDate ↓ 2 friend.id ↑			
limit	20			
CPs	2.2, 2.3, 3.3, 5.1, 8.1, 8.3			
relevance	This query looks for paths of length two, starting from a given Person, moving to its published messages and then to Persons who liked them. It tests several aspects related to join optimization, both at query optimization plan level and execution engine level. On the one hand, many of the columns needed for the projection are only needed in the last stages of the query, so the optimizer is expected to delay the projection until the end. This query implies accessing two-hop data, and as a consequence, index accesses are expected to be scattered. We expect to observe variate cardinalities, depending on the characteristics of the input parameter, so properly selecting the join operators will be crucial. This query has a lot of correlated sub-queries, so it is testing the ability to flatten the query execution plans.			

IC 1	query	Interactive / complex / 8			
IC 2	title	Recent replies			
IC 3 IC 4 IC 5 IC 6 IC 7 IC 8 IC 9 IC 10	pattern	person: Person id = \$personId hasCreator Message commentAuthor: Person id firstName lastName hasCreator comment: Comment id content creationDate			
IC 11 IC 12 IC 13 IC 14	desc.	Given a start Person, find the most recent Comments that are replies to Messages of the start Person. Only consider direct (single-hop) replies, not the transitive (multi-hop) ones. Return the reply Comments, and the Person that created each reply Comment.			
	params	1 personId D			
	result	1 commentAuthor.id ID R 2 commentAuthor.firstName String R 3 commentAuthor.lastName String R 4 comment.creationDate DateTime R 5 comment.id ID R 6 comment.content Text R			
	sort	1 comment.creationDate ↓ 2 comment.id ↑			
	limit	20			
	CPs	2.4, 3.3, 5.3			
	relevance	This query looks for paths of length two, starting from a given Person, going through its created Messages and finishing at their replies. In this query there is temporal locality between the replies being accessed. Thus the top-k order by this can interact with the selection, i.e. do not consider older Posts than the 20th oldest seen so far.			

IC 1	query	Interactive / complex / 9			
IC 2	title	Recent messages by friends or friends of friends			
IC 3 IC 4 IC 5 IC 6 IC 7 IC 8 IC 9 IC 10	pattern	person: Person id = \$personId knows*12 knows*12 id firstName lastName hasCreator message: Message creationDate < \$maxDate id content / imageFile creationDate			
IC 12 IC 13 IC 14	desc.	Given a start Person, find the most recent Messages created by that Person's friends or friends of friends (excluding start Person). Only consider Messages created before the given maxDate (excluding that day).			
	params	1 personId ID 2 maxDate Date			
	result	1 otherPerson.id ID R 2 otherPerson.firstName String R 3 otherPerson.lastName String R 4 message.id ID R message.content or message.imageFile (for photos) 6 message.creationDate DateTime R			
	sort	1 message.creationDate ↓ 2 message.id ↑			
	limit	20			
	CPs	1.1, 1.2, 2.2, 2.3, 3.2, 3.3, 8.5			
	relevance	This query looks for paths of length two or three, starting from a given Person, moving to its friends and friends of friends, and ending at their created Messages. This is one of the most complex queries, as the list of choke points indicates. This query is expected to touch variable amounts of data with entities of different characteristics, and therefore, properly estimating cardinalities and selecting the proper operators will be crucial.			



IC 1	query	Interactive / complex / 11			
IC 2	title	Job referral			
IC 3 IC 4 IC 5 IC 6 IC 7 IC 8 IC 9 IC 10 IC 11 IC 12 IC 13	pattern	person: Person id = \$personId knows*12 workAt year(workFrom) company: Company name isLocatedIn country: Country name = \$name			
IC 14	desc.	Given a start Person, find that Person's friends and friends of friends (excluding start Person) who started working in some Company in a given Country, before a given date (year).			
	params	1 personId ID 2 countryName String 3 workFromYear 32-bit Integer			
	result	1 otherPerson.id ID R 2 otherPerson.firstName String R 3 otherPerson.lastName String R 4 company.name String R 5 workAt.workFrom 32-bit Integer R			
	sort	1 workAt.workFrom ↑ 2 otherPerson.id ↑ 3 company.name ↓			
	limit	10			
	CPs	1.3, 2.3, 2.4, 3.3, 4.2			
	relevance	This query looks for paths of length two or three, starting from a Person, moving to friends or friends of friends, and ending at a Company. In this query, there are selective joins and a top-k order by that can be exploited for optimizations.			

IC 1	query	Interactive / complex / 12		
IC 2	title	Expert search		
IC 3 IC 4 IC 5 IC 6 IC 7 IC 8 IC 9 IC 10 IC 11 IC 12 IC 13 IC 14	pattern	person: Person id = \$personId id = \$personId id firstName isSubclassOf *0 TagClass hasType collect(tag.name) tag: Tag name comment: Comment replyOf Post		
	desc.	Given a start Person, find the Comments that this Person's friends made in reply to Posts, considering only those Comments that are direct (single-hop) replies to Posts, not the transitive (multi-hop) ones. Only consider Posts with a Tag in a given TagClass or in a descendent of that TagClass. Count the number of these reply Comments, and collect the Tags that were attached to the Posts they replied to, but only collect Tags with the given TagClass or with a descendant of that TagClass. Return Persons with at least one reply, the reply count, and the collection of Tags.		
	params	1 personId ID 2 tagClassName Long String		
	result	1 friend.id ID R 2 friend.firstName String R 3 friend.lastName String R 4 tagNames {Long String} A 5 replyCount 32-bit Integer A		
	sort	1 replyCount ↓ 2 friend.id ↑		
	limit	20		
	CPs	3.3, 7.2, 7.3, 8.2		
	relevance	This query starts at a Person, moves to its friends, and the to their Comments and their root Posts. Then, it gets the Tag of each Post and checks whether it (directly or transitively) belongs to the specified TagClass. This can be thought of a bidirectional search between the Person and the TagClass. The difficulty of this query is determining the optimal direction of this traversal.		

IC 1	query	Interactive / complex / 13
IC 2	title	Single shortest path
IC 3		
IC 4	pattern	Person id = \$person2Id knows*0 id = \$person2Id
IC 5		a = \$personzia
IC 6		Given two Persons, find the shortest path between these two Persons in the subgraph induced by
IC 7		the knows relationships. Return the length of this path:
IC 8		
IC 9	desc.	• -1: no path found
IC 10		• 0: start person = end person
IC 11		• > 0: path found (start person ≠ end person)
IC 12		
IC 13		1 person1Id ID
IC 14	params	2 person2Id ID
	result	1 shortestPathLength 32-bit Integer C
	CPs	3.3, 7.2, 7.3, 7.5, 8.1, 8.6
	relevance	This query looks for a variable length path, starting at a given Person and finishing at an another given Person. Proper cardinality estimation and search space pruning, will be crucial. This query also allows for possible parallel implementations.

IC 1	query	Interactive / complex / 14
IC 2	title	Trusted connection paths
IC 3 IC 4 IC 5 IC 6 IC 7 IC 8 IC 9 IC 10 IC 11	pattern	Enumerate all unweighted shortest paths on knows edges from person1 to person2. person1: Person id = \$person2!d Case 1: Replies on Posts, weight += 1.0 × count(c) personA: Person hasCreator c: Comment replyOf → post: Post For each edge on the path, calculate a weight based on interactions between the pair of Persons of the edge, are calculated as a sum of cases #1 and #2 for the Persons (both ways), and the sum of these weights determine the total weight of each path. pl knows px knows py py pw knows pz Case 2: Replies on Comments, weight += 0.5 × count(c1) personA: Person hasCreator c1: Comment replyOf → c2: Comment
IC 13 IC 14	desc.	Given two Persons, find all (unweighted) shortest paths between these two Persons, in the subgraph induced by the knows relationship. Then, for each path calculate a weight. The nodes in the path are Persons, and the weight of a path is the sum of weights between every pair of consecutive Person nodes in the path. The weight for a pair of Persons is calculated based on their interactions: • Every direct reply (by one of the Persons) to a Post (by the other Person) contributes 1.0. • Every direct reply (by one of the Persons) to a Comment (by the other Person) contributes 0.5. Note that interactions are counted both ways (e.g. if Alice writes 2 Post replies and 1 Comment reply to Bob, while Bob writes 3 Post replies and 4 Comment replies to Alice, their interaction score is $2 \times 1.0 + 1 \times 0.5 + 3 \times 1.0 + 4 \times 0.5 = 7.5$). Return all the paths with shortest length, and their weights. Do not return any rows if there is no path between the two Persons.
	params	1 person1Id ID 2 person2Id ID
	result	personIdsInPath [ID] C identifiers representing an ordered sequence of the Persons in the path pathWeight 64-bit Float C
	sort	1 pathWeight ↓ The order of paths with the same weight is unspecified
	CPs	3.3, 5.3, 7.2, 7.3, 7.5, 7.7, 8.1, 8.2, 8.3, 8.6
	relevance	This query looks for a variable length path, starting at a given Person and finishing at an another given Person. This is a more complex query as it not only requires computing the path length, but returning it and computing a weight. To compute this weight one must look for smaller sub-queries with paths of length three, formed by the two Persons at each step, a Post and a Comment.