Note: these slides are from 1981!

Acyclic database schemes (of various degrees): a painless introduction

Ron Fagin
IBM Research
San Jose

First portion of talk based on research by:

Catriel Beeri, Hebrew University
Ron Fagin, IBM San Jose Research
Dave Maier, SUNY Stony Brook
Alberto Mendelzon, Toronto
Jeff Ullman, Stanford
Mihalis Yannakakis, Bell Labs

Database schema

A	B	C

B	C	D

A	E

A	B	C

B	C	D

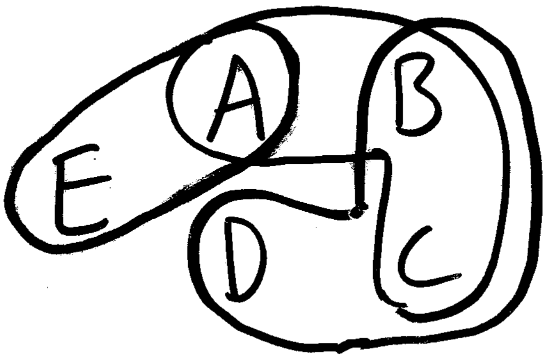
A	D

Database schema

A	B	C

B	C	D

A	E

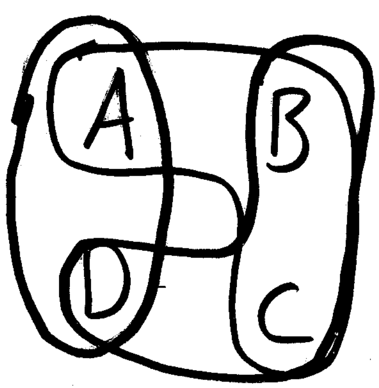


acyclic

A	B	C

B	C	D

A	D



cyclic

Many basic desirable properties of relational database schemes are equivalent to acyclicity.

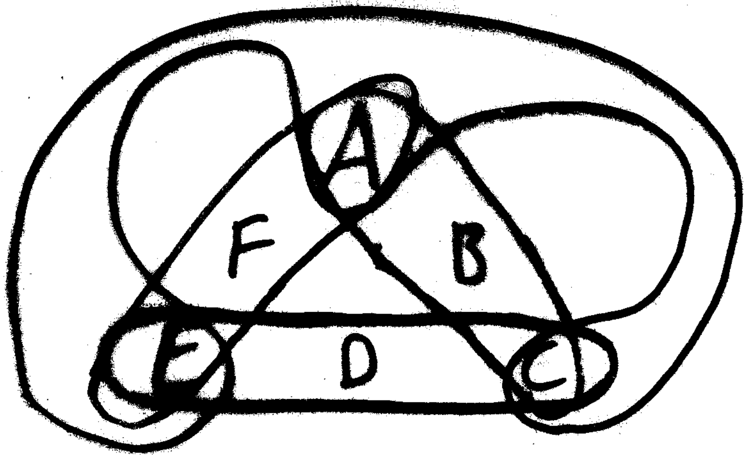
Acyclicity eliminates pathological cases.

Theory much more elegant in acyclic case.

More efficient algorithms available in acyclic case.

There is a simple algorithm for determining acyclicity.

Graham's algorithm

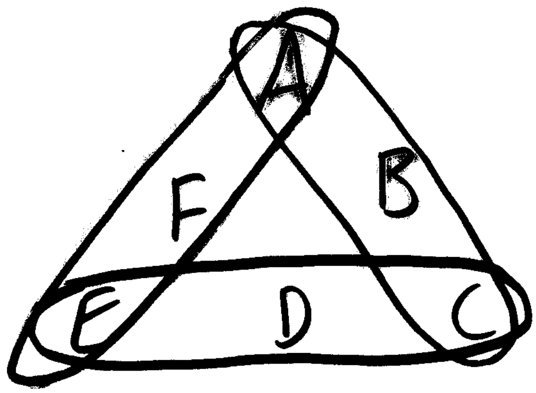


~~A B C~~

~~C D E~~

~~A E F~~

~~A~~ ~~C~~ ~~E~~



A ~~B~~ C

C ~~D~~ E

A E ~~F~~

<u>Theorem</u> Hypergraph acyclic iff Graham's algorithm succeeds.

Consistency

Question: Are the relations projections of a single relation?

Example 1:

A	B	C
0	0	1
1	0	1
2	3	4

B	C	D
0	1	1
3	4	5

A	D
0	1
1	1
2	5

Answer: Yes, projections of

A	B	C	D
0	0	1	1
1	0	1	1
2	3	4	5

Example 2:

A	B	C
0	0	0
1	1	1

B	C	D
0	0	0
1	1	1

A	D
0	1
1	0

Answer: No

Consistency

Question: Are the relations projections of a single relation?

Example 1:

A	B	C
0	0	1
1	0	1
2	3	4

B	C	D
0	1	1
3	4	5

A	D
0	1
1	1
2	5

pairwise agreement

Necessary condition for consistency:
Relations agree pairwise

Determining consistency NP-complete
(Honeyman, Ladner, Yannakakis)

<u>Theorem</u> If schema acyclic, necessary + sufficient condition for consistency of instances is pairwise agreement.

<u>Theorem</u> If schema cyclic, there is always an example of relations passing the pairwise test but failing to be consistent.

Unrestricted case: determining consistency
NP-complete

Acyclic case: polynomial-time algorithm
for determining consistency

Semijoins

A	B	C	D	E	F	G
0	0	1	2	3	4	5
1	3	9	0	6	3	17
2	1	17	4	19	2	8

Relation r_1 at site 1

A	B	U	V	W	X	Y	Z
0	0	101	102	103	104	105	106
3	6	14	91	3	6	47	15

Relation r_2 at site 2

Desire join $r_1 \bowtie r_2$

Semijoins

A	B	C	D	E	F	G
0	0	1	2	3	4	5
1	3	9	0	6	3	17
2	1	17	4	19	2	8

Relation r_1 at site 1

A	B	U	V	W	X	Y	Z
0	0	101	102	103	104	105	106
3	6	14	91	3	6	47	15

Relation r_2 at site 2

Desire join $r_1 \bowtie r_2 = \{qrs : qr \in r_1 + qs \in r_2\}$

A	B	C	D	E	F	G	U	V	W	X	Y	Z
0	0	1	2	3	4	5	101	102	103	104	105	106

<u>Strategy 1</u> Ship r_2 to site 1, or r_1 to site 2, and take join

<u>Strategy 2</u> (semijoin strategy)

(a) Ship AB projection of r_2 to site 1
(b) Reduce r_1 (this forms semijoin $r_1 \ltimes r_2$)
(c) Ship reduced version of r_1 to site 2 and take join

Strategy 2 actually used in some distributed DB systems (SDD1, RAP).

Semijoin program

$$r_3 := r_3 \ltimes r_7$$
$$r_7 := r_7 \ltimes r_4$$
$$r_6 := r_6 \ltimes r_{17}$$
$$\vdots$$

<u>Theorem</u> If schema acyclic, there is a short semijoin program that fully reduces all of the relations

("Fully reduced" means that no tuple can be removed from any of the relations, without the join getting smaller)

Cyclic case

A	B	C
0	0	0
1	1	1

B	C	D
0	0	0
1	1	1

A	D
0	1
1	0

Join is empty, but no relation can be reduced by a semijoin

<u>Theorem</u> If schema cyclic, no semijoin program is guaranteed to fully reduce the relations.

Monotone joins

Example of nonmonotone join:

Say joining r_1, r_2, r_3, r_4.

$r_1 \bowtie r_2$: 1000 tuples

" $\bowtie r_3$: 1000 000 tuples

" $\bowtie r_4$: 10 tuples

<u>Monotone join</u>: a parenthesization, such as

$$(r_1 \bowtie r_4) \bowtie (r_2 \bowtie r_3),$$

such that each binary join that appears is over consistent relations. In above example:

r_1, r_4 consistent

r_2, r_3 consistent

$(r_1 \bowtie r_4), (r_2 \bowtie r_3)$ consistent

<u>Theorem</u> A database scheme is acyclic if and only if every pairwise consistent database over the scheme has a monotone join.

<u>Theorem</u> The following are equivalent

(1) $\{R_1, \ldots, R_n\}$ is an acyclic hypergraph

(2) Graham's algorithm succeeds with input $\{R_1, \ldots, R_n\}$

(3) The pairwise test decides consistency over $\{R_1, \ldots, R_n\}$ (i.e., every pairwise consistent database is globally consistent)

(4) $\{R_1, \ldots, R_n\}$ has a full semijoin reducer

(5) The join dependency $\bowtie\{R_1, \ldots, R_n\}$ is equivalent to a set of multivalued dependencies

(6) The reduction of the hypergraph $\{R_1, \ldots, R_n\}$ forms the maximal cliques of a chordal graph

(7) $\{R_1, \ldots, R_n\}$ has a space-efficient join strategy.

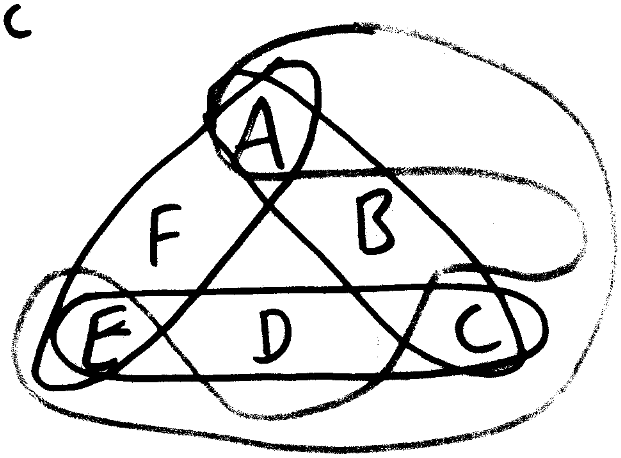
Other types of acyclicity

α-acyclic: what we've called "acyclic"

subhypergraph: a subset of the edges

subscheme: subset of the relation schemes

A subhypergraph of an α-acyclic hypergraph may be α-cyclic



<u>Defn.</u> A hypergraph (respectively, db scheme) is β-acyclic if every subhypergraph (respectively, db subscheme) is α-acyclic.

Note: every subhypergraph of a β-acyclic hypergraph is β-acyclic.

Let $\mathcal{P}$ be any one of the equivalent desirable properties of α-acyclic db schemes

<u>Theorem</u> A database scheme is β-acyclic if and only if every subscheme enjoys property $\mathcal{P}$.

<u>Theorem</u> A hypergraph is β-cyclic if and only if there is a sequence

$$(S_1, x_1, S_2, x_2, \ldots, S_m, x_m, S_{m+1})$$

where

(i) $x_1, \ldots, x_m$ are distinct nodes;

(ii) $S_1, \ldots, S_m$ are distinct edges, and $S_{m+1} = S_1$;

(iii) $m \geq 3$, i.e., at least 3 edges are involved, and

(iv) x_i is in S_i and S_{i+1} and in no other S_j.

<u>Theorem</u> There is a polynomial-time algorithm for deciding β-acyclicity.

γ-acyclicity

$$\gamma\text{-acyclic} \underset{\not\Leftarrow}{\Rightarrow} \beta\text{-acyclic} \underset{\not\Leftarrow}{\Rightarrow} \alpha\text{-acyclic}$$

23A, RF4A

EMPWORK:

EMP	DEPT	SAL
Fagin	CS	$200K

DEPTINFO:

DEPT	CITY	MGR
CS	San Jose	Peled

EMPHOME:

EMP	STREET	CITY	CHILD
Fagin	162 Loma Alta	Los Gatos	Joshua

EMP-CITY relationships:

a. EMPHOME[EMP CITY]

b. (EMPWORK $\bowtie$ DEPTINFO)[EMP CITY]

EMPWORK:

EMP	DEPT	SAL
Fagin	CS	$200K

DEPTINFO:

DEPT	WORK CITY	MGR
CS	San Jose	Peled

EMPHOME:

EMP	STREET	HOME-CITY	CHILD
Fagin	162 Loma Alta	Los Gatos	Joshua

EMP-CITY relationships:

a. EMPHOME[EMP CITY]

b. (EMPWORK ⋈ DEPTINFO)[EMP CITY]

Theorem A db scheme is γ-acyclic iff for every consistent db over the scheme, and for every set X of attributes, there is a unique X-relationship.

Query: Find all employees who work in San Jose

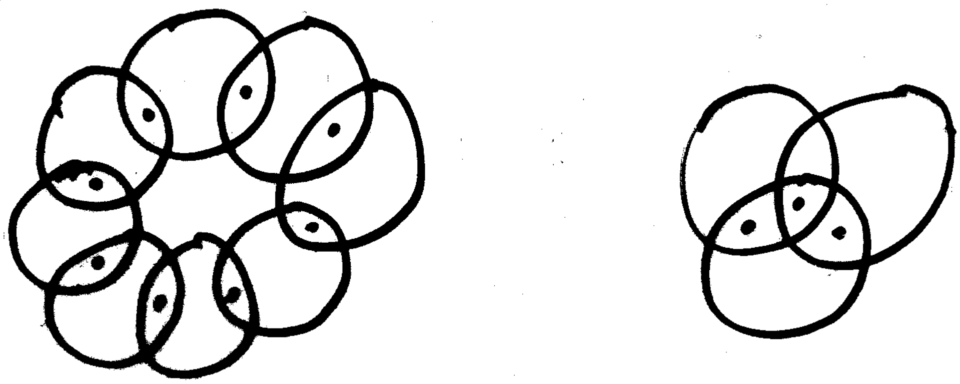
```
SELECT EMP
FROM EMPWORK, DEPTINFO
WHERE EMPWORK.DEPT = DEPTINFO.DEPT
AND DEPTINFO.WORKCITY = 'San Jose'
```

23B, RF4B

SELECT EMP WHERE WORKCITY = 'San Jose'

What is a γ-acyclic hypergraph?

Forbidden configurations:



- Algorithm for determining γ-acyclicity:

Do the following steps, repeatedly, in any order, until none can be applied:

1. Eliminate a node in exactly one edge
2. Eliminate an edge with exactly one node.
3. If 2 nodes in precisely the same edges, eliminate one.
4. Eliminate duplicate edges

A	B	C	D			
	B	C	D	E		
		C	D		F	
A						G

If everything eliminated, γ-acyclic; else, γ-cyclic

Recall that:

<u>Theorem</u> A database scheme is α-acyclic if and only if some join is monotone.

Recall that:

Theorem A database scheme is ~~α~~ γ-acyclic if and only if ~~some~~ every connected join is monotone.