Operations in databases (and elsewhere)

Def. The *projection* $P_{i_1 i_2 \dots i_m}$ maps $(a_1, a_2, ..., a_n)$ to $(a_{i_1}, a_{i_2}, ..., a_{i_m})$.
$P_{13}(1, 5, 25, 625) = (1, 25)$

Def.
R — relation of degree m
S — relation of degree n
join$_p(R, S)$, $p \leq \min(m, n)$
— relation of degree $m + n - p$

$$\{(a_1, a_2, ..., a_{m-p}, c_1, c_2, ..., c_p, b_1, b_2, ..., b_{n-p}) \mid (a_1, a_2, ..., a_{m-p}, c_1, c_2, ..., c_p) \in R \text{ and } (c_1, c_2, ..., c_p, b_1, b_2, ..., b_{n-p}) \in S\}.$$

Example:
$R \subseteq$ { (CPR #, Last Name, First, Date admitted, Diagnosis)}
$S \subseteq$ { (Date admitted, Diagnosis, Treatment)}

join$_2(R, S)$
adds treatment onto everyone in R.

If > 1 treatment for some date and diagnosis — each person admitted that day with that diagnosis will have one record inserted for every treatment.

Representing binary relations by matrices

R – a relation from $A = \{a_1, a_2, ..., a_m\}$ to $B = \{b_1, b_2, ..., b_n\}$.
Let $M_R = [m_{ij}]$ be an $m \times n$ matrix, where

$$M_{ij} = \begin{cases} 1 & \text{if } (a_i, b_j) \in R \\ 0 & \text{if } (a_i, b_j) \notin R \end{cases}$$

Example:
A={ Peter, Marianne, Niels} and
B={ vanilla, strawberry, chocolate, mint}.
R={ (Peter,vannilla), (Peter,chocolate), (Marianne,vanilla), (Marianne,strawberry), (Niels,vanilla), (Niels,chocolate), (Niels,mint)}

$$M_R = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \end{bmatrix}$$

If R is a relation on a set A, M_R is square.
R is reflexive iff $m_{ii} = 1 \quad \forall i$.
R is symmetric iff $M_R = M_R^t$
R is antisymmetric iff $m_{ij} = 1 \Rightarrow m_{ji} = 0$ for $i \neq j$.

Union: $R \cup S$
Def. $A = [a_{ij}]$, $B = [b_{ij}]$, $m \times n$, 0-1 matrices.
The *join* of A and B is $A \vee B = [a_{ij} \vee b_{ij}]$.
$M_{R \cup S} = M_R \vee M_S$.

Intersection: $R \cap S$
Def. $A = [a_{ij}]$, $B = [b_{ij}]$, $m \times n$, 0-1 matrices.
The *meet* of A and B is $A \wedge B = [a_{ij} \wedge b_{ij}]$.
$M_{R \cap S} = M_R \wedge M_S$.

Composite: $R \circ S$
Def. $A = [a_{ij}]$, $B = [b_{ij}]$, $m \times k$, 0-1 matrix.
$B = [b_{ij}]$, $k \times n$, 0-1 matrix.
The *Boolean product* of A and B is $A \odot B = [c_{ij} = (a_{i1} \wedge b_{1j}) \vee (a_{i2} \wedge b_{2j}) \vee ... \vee (a_{ik} \wedge b_{kj})]$.
$M_{R \circ S} = M_R \odot M_S$.

Powers of relations:

Def. A — $m \times m$, 0-1 matrix.

The r^{th} Boolean power of A is

$$\begin{aligned} A^{[0]} &= I_n \\ A^{[n+1]} &= A^{[n]} \odot A \end{aligned}$$

$M_{R^n} = M_R^{[n]}$.

Example:

$R = \{(1,1),(1,3),(1,5),(2,4),(3,4)\}$

1	0	1	0	1
0	0	0	1	0
0	0	0	1	0
0	0	0	0	0
0	0	0	0	0

Directed graphs of relations

Reflexive — loops at every vertex

Symmetric — all edges come in pairs one going in each direction

Antisymmetric — only 1 edge between any 2 vertices

Transitive — if $\exists$ a path from u to v, then $\exists$ an edge from u to v

Closure

P – property –transitivity

R – relation

S – *closure* of R w.r.t. P

The following must hold:

- $R \subseteq S$
- S has property P
- If $R \subseteq T$, T has P, then $S \subseteq T$

Example: Train network.

Cities = $\{A, B, C, D, E\}$

$\exists$ train from A to B and B to A

$\exists$ train from A to C and C to A

$\exists$ train from B to E and E to B

$\exists$ train from C to E and E to C

$\exists$ train from D to E and E to D

Thm. R a relation. G – digraph for R.
$\exists$ a path of length n from a to b in G
iff $(a,b) \in R^n$.

Pf. (by ind on n)

Basis step: $n = 1$.

$(a,b) \in R$ iff path of length 1. $\surd$

Inductive step: Assume true for some $n \geq 1$:
$\forall a,b \in V(G)$ $\exists$ a path of length n from a to b
iff $(a,b) \in R^n$.

$\exists$ a path of length $n+1$ from a to b

iff $\exists$ c s.t. $\exists$ a path of length n from a to c and an edge (c,b)

iff $\exists$ c s.t. $(a,c) \in R^n$ and $(c,b) \in R$ (IH)

iff $(a,b) \in R^{n+1}$. $\surd$

By induction, $\exists$ a path of length n from a to b in G iff $(a,b) \in R^n$. $\square$

Warshall's Algorithm

$W_0 = M_R$
$W_k = [w_{ij}^{(k)}]$ where

$$w_{ij}^{(k)} = \begin{cases} 1 & \text{if } \exists\ P = \{v_i, v_{s_1}, v_{s_2}, ..., v_{s_t}, v_j\} \\ & \text{s.t. } s_r \in \{1, ..., k\}\ \forall r \\ 0 & \text{otherwise} \end{cases}$$

procedure Warshall (M_R: $n \times n$, 0-1 matrix)
$W \leftarrow M_R$
for $k \leftarrow 1$ **to** n
  for $i \leftarrow 1$ **to** n
    for $j \leftarrow 1$ **to** n
      $w_{ij} \leftarrow w_{ij} \vee (w_{ik} \wedge w_{kj})$
return $(W = M_{R^*})$

$2n^3$ bit operations.

Equivalence relation —
reflexive, symmetric, transitive

Example: $R = \{(a,b) \mid a \equiv b \pmod m\}$

Equivalence class of x —
all elements related to x
$[a]_R = \{s \mid (a,s) \in R\}$
$b \in [a]_R$ is a *representative* of the class.

An equivalence relation *partitions* a set.

Thm. R — equivalence relation on A
The following are equivalent:

1. aRb
2. $[a] = [b]$
3. $[a] \cap [b] \neq \emptyset$

Pf. $(1 \Rightarrow 2)$ Suppose $x \in [a]$.
Then, $(a,x) \in R$.
R reflexive and $(a,b) \in R \Rightarrow (b,a) \in R$.
R transitive, so $(b,a) \in R$, $(a,x) \in R$
$\Rightarrow (b,x) \in R$. Thus, $x \in [b]$, so $[a] \subseteq [b]$.
Similarly $[b] \subseteq [a]$, so $[a] = [b]$.

$(2 \Rightarrow 3)$ $a \in [a]$.
$[a] = [b] \Rightarrow a \in [b] \Rightarrow a \in [a] \cap [b]$.
Thus, $[a] \cap [b] \neq \emptyset$.

$(3 \Rightarrow 1)$ Suppose $x \in [a] \cap [b]$.
$(a,x) \in R \ \wedge\ (b,x) \in R$.
R symmetric, so $(x,b) \in R$.
R transitive, so $(a,b) \in R$. $\square$

Def. R a relation on S.
R is a *partial order* if it is
reflexive, antisymmetric, transitive.
(S,R) — *partially ordered set, poset.*

Example: Containment of line segments on the real line.

a, b are *comparable* in (S,R) if $(a,b) \in R$ or $(b,a) \in R$. Otherwise they are *incomparable*.

Def. If every pair of elements in S is comparable, S is a *totally ordered set* (*linearly ordered set*, *chain*), and R is a *total order* (*linear order*).

$\mathbb{Z}$ and $\mathbb{R}$ are totally ordered under $\leq$.

Def. A totally ordered set is *well-ordered* if every nonempty set has a least element.

$\mathbb{Z}$ is well ordered.

Hasse diagrams —

- start with the digraph for the partial order
- remove loops
- remove edges implied by transitivity
- remove arrows (they all go upwards)

An element $a \in (S, R)$ is *maximal* if $(a, b) \in R \;\Rightarrow\; a = b$. It is *minimal* if $(b, a) \in R \;\Rightarrow\; a = b$.

An element $a \in (S, R)$ is the *greatest element* if $\forall\ b \in S$, $(b, a) \in R$. It is the *least element* if $\forall\ b \in S$, $(a, b) \in R$.

An element $a \in S$ is an *upper bound* of $A \subseteq S$ if $\forall u \in A$, $(u, a) \in R$. The *least upper bound* is less than or equal to any other upper bound of A.

An element $a \in S$ is an *lower bound* of $A \subseteq S$ if $\forall u \in A$, $(a, u) \in R$. The *greatest lower bound* is greater than or equal to any other lower bound of A.

Lattice — a poset in which every pair of elements has both a least upper bound and a greatest lower bound.

Example: $(P(S), \subseteq)$, where S is a set.
The least upper bound of A and B is $A \cup B$.
The greatest lower bound of A and B is $A \cap B$.

Huffman Encoding

Text files in ASCII waste space.
ASCII uses 8 bits for every character.
Redundancy in all natural languages.

Use short bit strings for the most frequent n-tuples and longer strings for others.
One can represent English text using $\approx$ 1.5 bits/character on average.

Huffman codes are prefix codes.
Also called *prefix-free codes*.
No code representing a character is a prefix of another code.

Example: "0110" and "011010" could not both be codes in a prefix-free encoding;
the first is a prefix of the second.

Prefix-free encoding makes decoding easy.

Example: Suppose $f(a) = 01$, $f(b) = 0010$, $f(c) = 0011$, $f(d) = 10$, $f(e) = 11$.
010011111000110111 decodes as
01 0011 11 10 0011 01 11: *acedcae*.
Need just one pass through the string.

To do Huffman encoding, build a binary tree.
k possible characters.
Probability of character c_i is p_i.
$0 < p_i \leq 1\ \forall\ i$, and $\sum_{i=1}^{k} p_i \;=\; 1$.
One leaf in the binary tree for each character.
Each node in the tree contains a key,
a probability.

procedure Huffman($p_1, p_2, ..., p_k$)

Create 1 tree for each character.
Put the probabilities for the characters in roots.

Sort the roots according to their probabilities.

while $\exists > 1$ tree **do**
  Remove the 2 trees whose roots have the smallest probabilities, p_i and p_j.
  Make these two trees the left and right subtrees of a new root.
  Give the new root probability $p_i + p_j$.
  Insert the new tree, maintaining sorted order for roots.
end do

return pointer to root of remaining tree

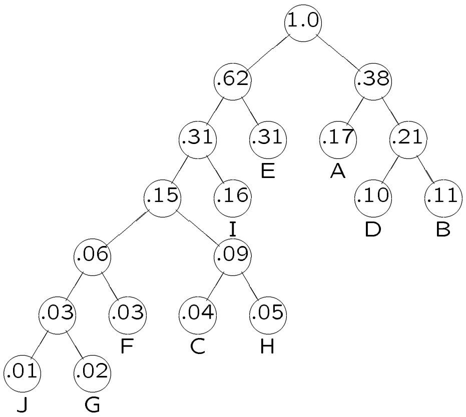
Each pass through the loop decreases the number of existing trees by one.
So there are $k-1$ passes through the loop.

Get short paths from the root to the leaves with frequently occurring characters.

The path from the root to a leaf will define the encoding.

Edge going to a left child — 0
Edge going to a right child — 1

A	B	C	D	E	F	G	H	I	J
.17	.11	.04	.10	.31	.03	.02	.05	.16	.01



The encoding of the different characters is as follows:

A	B	C	D	E
10	111	00010	110	01

F	G	H	I	J
00001	000001	00011	001	000000

Codes are paths in the tree, so no code is a prefix of another.

Algebra

The definition of a group

Def. A *group* is a set G closed under a binary operation $\odot$ such that

- *Associative Law:*
  $\forall x, y, z \in G,\ x \odot (y \odot z) = (x \odot y) \odot z.$
- *Identity:* $\exists e \in G$, the *identity*:
  $\forall x \in G,\ e \odot x = x \odot e = x.$
- *Inverse:*
  $\forall x \in G,\ \exists y \in G$ s.t. $x \odot y = y \odot x = e.$
  $y = x^{-1}$ is the *inverse* of x.

Examples: The integers $\mathbb{Z}$, the reals $\mathbb{R}$, and the rationals $\mathbb{Q}$ are groups under addition.

Example: $\mathbb{R} - \{0\}$ under multiplication.

Def. For finite groups, the number of elements in a group G, written $|G|$, is the *order* of the group.

Example: $\mathbb{Z}_n$, the integers modulo n, under addition. The order of $\mathbb{Z}_n$ is n.

Example: $\mathbb{Z}_n^*$, the positive integers less than n which are relatively prime to n, under multiplication. The order of $\mathbb{Z}_n^*$ is $\phi(n)$, where ϕ is the Euler ϕ-function.

The above examples are all *abelian groups* — the operation is commutative: $x \odot y = y \odot x$ for all x, y in the group.
Not all groups are abelian.

Example: S_n, the symmetric group on n letters, is the set of permutations of the set $\{1, 2, ..., n\}$.
S_n is not an abelian group.

Any permutation can be written as a product of cycles.

A *transposition* is a cycle of length 2, (ij). A permutation is *even* iff it can be expressed as a product of an even number of transpositions.

The *symmetric group* of a set X is Sym(X).

Subgroups

Def. Let G be a group, and $H \subseteq G$.
H is a *subgroup* of G if H itself is a group w.r.t. the operation in G. The *order* of a subgroup is its cardinality.

Suppose G is a group and $H \subseteq G$. Then H is a subgroup of G iff the following hold:

- $\forall x, y \in H$, $x \odot y \in H$.
  (H is closed under the group operation.)
- The identity is in H.
- $\forall x \in H$, $x^{-1} \in H$.

Example: Any group is a subgroup of itself.

Example: If e is the identity in G, $\{e\}$ is a subgroup of G.

Example: The even integers are a subgroup of the integers under addition.

Example: $\left\{ \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 2 & 3 & 4 & 5 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 1 & 4 & 3 & 5 \end{pmatrix} \right\}$ is a subgroup of S_5.

Example: A_n, the set of all even permutations on n letters, is a subgroup of S_n.
It is called the *alternating group* on n letters.

Example: The set $\{0, 5, 10\}$ is a subgroup of $\mathbb{Z}_{15}$ under addition.

Example: The set $\{1, 2, 4\}$ is a subgroup of $\mathbb{Z}_7^*$.

Thm. Suppose S is a nonempty collection of subgroups of a group G. Let $H' = \bigcap_{H \in S} H$. Then H' is a subgroup of G.

Generators

Def. Let H be a subset of a group G, and let S be the collection of all subgroups of G which contain H. Then, $<H> = \bigcap_{G' \in S} G'$ is the *subgroup generated by* H.

$H \subseteq G$.
$<H> = \{h_1 \odot h_2 \odot ... \odot h_n \mid h_i \text{ or } h_i^{-1} \in H \quad \forall i\}$.

Def. A group or subgroup is said to be *cyclic* if it is generated by a single element.
Such an element is a *generator* or *primitive element*.

Def. The *order* of an element of a group G is the order of the subgroup that element generates.

Thm. Suppose G is a group with identity e, and $g \in G$ has finite order m. Then m is the least positive integer such that $g^m = e$.

Example: The set $\{2\}$ generates the subgroup $\{1, 2, 4\}$ of $\mathbb{Z}_7^*$.
Thus, it is a cyclic subgroup.
The order of the element 2 is 3.
The set $\{3\}$ generates all of $\mathbb{Z}_7^*$,
so it is a cyclic group.

Fact. $\mathbb{Z}_p^*$ is cyclic whenever p is prime.

Lagrange's Theorem

Def. Let $H < G$, $x \in G$. $Hx = \{h \odot x \mid h \in H\}$ is a *right coset* of H in G.

Lemma Let $H < G$. All right cosets of H contain $|H|$ elements.

The relation $R =$
$\{(a, b) \mid a \text{ and } b \text{ are in the same right coset of } H\}$
is an equivalence relation.

Lemma Let H be a subgroup of G, $x, y \in G$. Then either $Hx = Hy$ or $Hx \cap Hy = \emptyset$.

These two lemmas tell us that the group G must be a disjoint union of right cosets of any subgroup H, all of which have the same size.

Thm. [Lagrange] If G is a finite group and $H < G$, then the order of H divides the order of G.

Corollary Let G be a finite group and $g \in G$. The order of the element g divides the order of the group G.

Corollary Suppose $|G| = n$ and $g \in G$. Then $g^n = e$, where e is the identity in G.
Pf. Let s be the order of the subgroup generated by g. By a previous theorem, $g^s = e$, where e is the identity. By Lagrange's Theorem, s divides n, so there is an integer c such that $n = sc$. Note that $g^n = g^{sc} = (g^s)^c = e^c = e$, so the corollary follows. □

Thm. [Fermat's Little Theorem] If p is prime and a is an integer not divisible by p, then $a^{p-1} \equiv 1 \pmod{p}$.

Using this same group theory, we can show that encryption with RSA, followed by decryption with RSA yields the original message, if the message is less than n and relatively prime to n.

$n = pq$ where p and q are large primes.
$|\mathbb{Z}_n^*| = \phi(n) = (p-1)(q-1)$.
If the message M is less than n and relatively prime to n, $M \in \mathbb{Z}_n^*$.
$ed \equiv 1 \pmod{(p-1)(q-1)}$,
so $\exists k \in \mathbb{Z}$ s.t. $ed = 1 + k(p-1)(q-1)$.

If $C \equiv M^e \pmod{n}$, then $C^d \equiv (M^e)^d \equiv M^{ed} \equiv M^{1+k(p-1)(q-1)} \equiv M \cdot M^{k(p-1)(q-1)} \equiv M \cdot (M^{\phi(n)})^k \equiv M \cdot 1^k \equiv M \pmod{n}$.

Rings and fields

Def. A *ring* is a set R closed under two binary operations $+$ and $\bullet$ s.t.

- R1. R is an abelian group with respect to the operation $+$.
- R2. The operation $\bullet$ is associative.
- R3. [Distributive Laws] $\forall x, y, z \in R$, the following hold:

$$\begin{aligned} x \bullet (y+z) &= x \bullet y + x \bullet z \\ (y+z) \bullet x &= y \bullet x + z \bullet x \end{aligned}$$

The first operation $+$ is called addition and the second operation $\bullet$ is called multiplication.

Example: $\{0\}$ is the *trivial ring*.
$0 + 0 = 0$ and $0 \bullet 0 = 0$.
A *nontrivial ring* is a ring with more than one element.

The identity element with respect to addition is called *zero*, and all other elements are called *nonzero elements*.

If the ring R has an identity element i with respect to multiplication, then for all $x \in G$, $i \bullet x = x \bullet i = x$, and R is said to be a *ring with identity*. This identity is denoted by 1.

The ring R is *commutative* if for all $x, y \in R$, $x \bullet y = y \bullet x$.

Examples: $\mathbb{Z}$ and $\mathbb{R}$ are both commutative rings with identity.

Def. A *field* is a nontrivial commutative ring with identity in which every nonzero element has a multiplicative inverse.

Examples: $\mathbb{R}$ is a field and $\mathbb{Q}$ is a field, but $\mathbb{Z}$ is not.

Example: $\mathbb{Z}_n$ is a field when n is prime. It is a ring when n is composite, but not a field.