Solution 3

Solved by Aaron Colin Foote, acf502, 4258770

(i) Find full information about the applicants who applied for a position offered by a given employer.
(ii) Find full information about the applicants who posses a give skill.
(iii) Find full information about the skills possessed by a given applicant.
(iv) Find full information about the positions applied by a given applicant.
(v) Find full information about employers who advertise more than a given number positions.

(1) SELECT statements implementing the queries (i), (ii), (iii), (iv) and (v):

i.
```
SELECT DISTINCT A.anumber, fname, lname, dob, A.city, A.state, A.phone,
    A.fax, A.email
FROM APPLICANT A
JOIN APPLIES
        ON A.anumber = APPLIES.anumber
JOIN POSITION
        ON APPLIES.pnumber = POSITION.pnumber
JOIN EMPLOYER
        ON POSITION.ename = EMPLOYER.ename
WHERE EMPLOYER.ename = 'University of New South Wales';
```

ii.
```
SELECT DISTINCT A.anumber, fname, lname, dob, city, state, phone, fax,
    email
FROM APPLICANT A
JOIN SPOSSESSED
        ON A.anumber = SPOSSESSED.anumber
JOIN SKILL
        ON SPOSSESSED.sname= SKILL.sname
WHERE SKILL.sname = 'driving';
```

iii.
```
SELECT *
FROM SKILL
JOIN SPOSSESSED
        ON SKILL.sname = SPOSSESSED.sname
JOIN APPLICANT
        ON SPOSSESSED.anumber = APPLICANT.anumber
WHERE APPLICANT.anumber = 000001;
```

iv.
```
SELECT *
FROM POSITION
JOIN APPLIES
        ON POSITION.pnumber = APPLIES.pnumber
JOIN APPLICANT
        ON APPLIES.anumber = APPLICANT.anumber
WHERE APPLICANT.anumber = 000001;
```

v.
```
SELECT E.ename, city, state, phone, fax, email, web
FROM EMPLOYER E
JOIN POSITION
        ON E.ename = POSITION.ename
GROUP BY E.ename, city, state, phone, fax, email, web
HAVING COUNT(*) > 1;
```

(2) Computations of costs and benefits that lead to construction of clustering graph:

SKILL	50 data blocks
SREQUIRED	200 data blocks
APPLICANT	500data blocks
EMPLOYER	300 data blocks
EMPLBY	5000 data blocks
POSITION	500 data blocks
SPOSSESSED	300 data blocks
SNEEDED	600 data blocks
APPLIES	1000 data blocks

Assume that join of:

JOIN A: APPLICANT and APPLIES needs 3 * (500 + 1000) = 4500 read block operations
JOIN B: APPLIES and POSITION needs 3 * (1000 + 500) = 4500 read block operations
JOIN C: POSITION and EMPLOYER needs 3 * (500 + 300) = 2400 read block operations
JOIN D: APPLICANT and SPOSSESSED needs 3 * (500 + 300) = 2400 read block operations
JOIN E: SPOSSESSED and SKILL needs 3 * (300 + 50) = 1050 read block operations

Then, the benefits from clustering of:

JOIN A 5 times per day: 5 * 4500 – 5 * (500 + 1000) = 15000 reads per day
JOIN A 10 times per day: 10 * 4500 – 10 * (500 + 1000) = 30000 reads per day

JOIN B 5 times per day: 5 * 4500 – 5 * (1000 + 500) = 15000 reads per day
JOIN B 10 times per day: 10 * 4500 – 10 * (1000 + 500) = 30000 reads per day

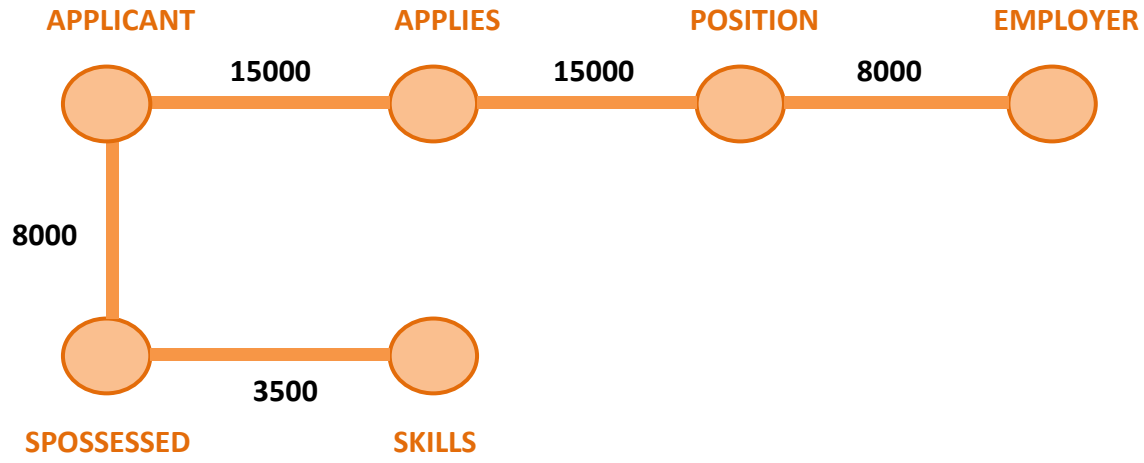
JOIN C 5 times per day: 5 * 2400 – 5 * (500 + 300) = 8000 reads per day
JOIN C 10 times per day: 10 * 2400 – 10 * (500 + 300) = 16000 reads per day

JOIN D 5 times per day: 5 * 2400 – 5 * (500 + 300) = 8000 reads per day
JOIN D 10 times per day: 10 * 2400 – 10 * (500 + 300) = 16000 reads per day

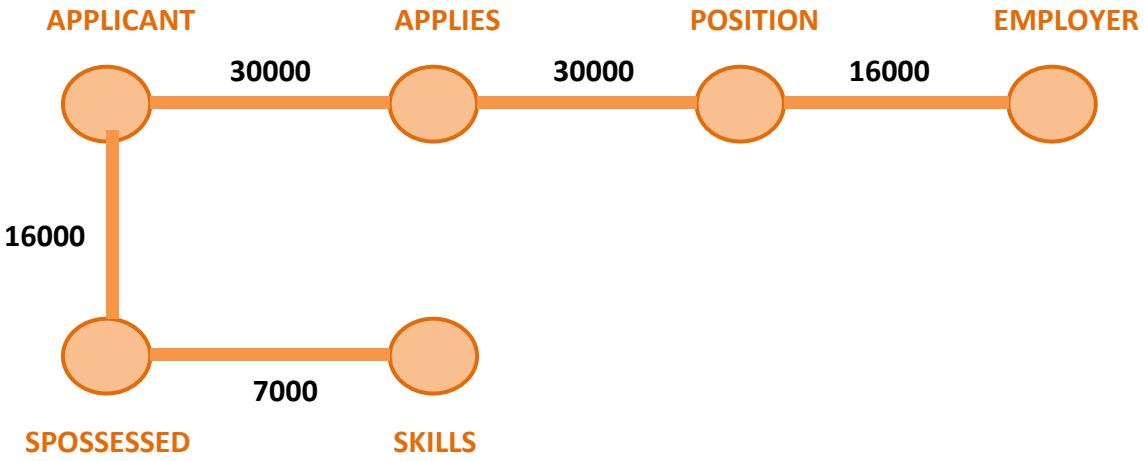
JOIN E 5 times per day: 5 * 1050 – 5 * (300 + 50) = 3500 reads per day
JOIN E 10 times per day: 10 * 1050 – 10 * (300 + 50) = 7000 reads per day

(3) A drawing of a clustering graph:

For queries i and ii (5 times per day):



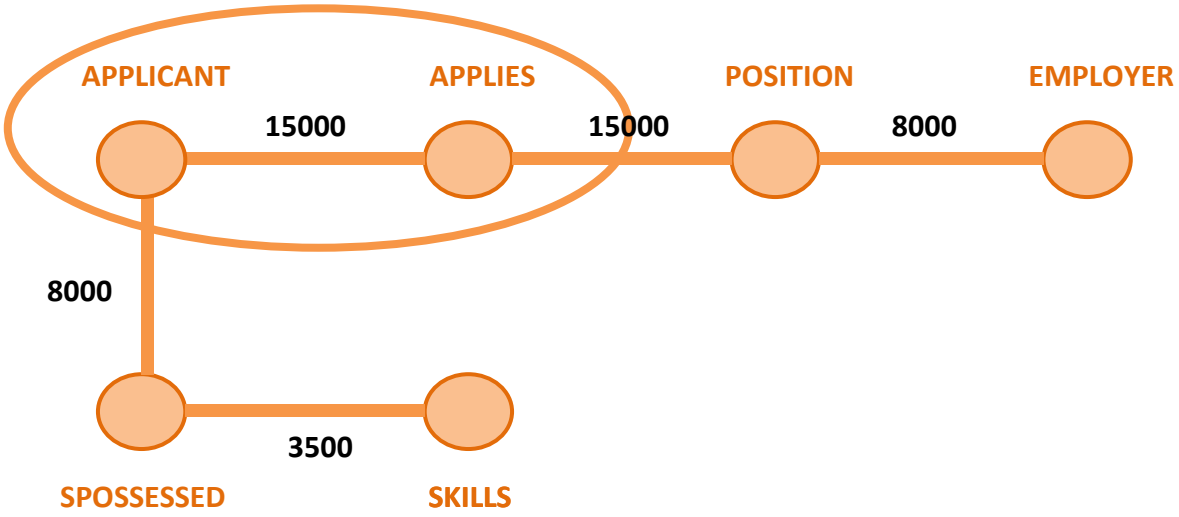
For queries iii, iv and v (10 times per day):



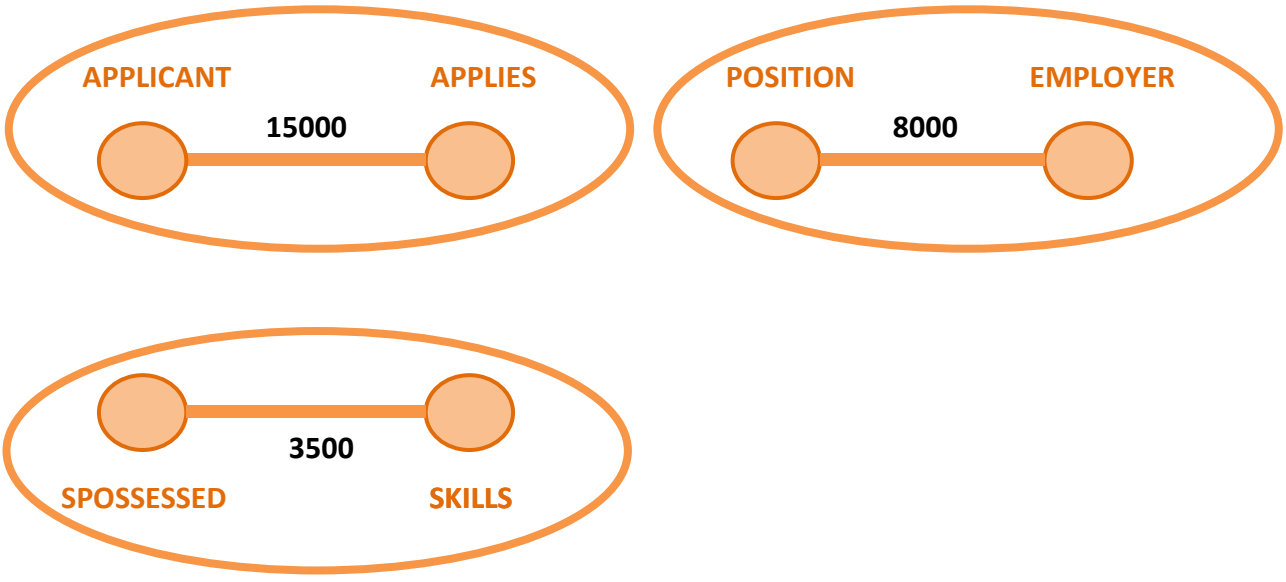
(4) Suboptimal clustering that improved performance of the queries (i), (ii), (iii), (iv) and (v):

Note that I am following algorithm written on lecture notes:

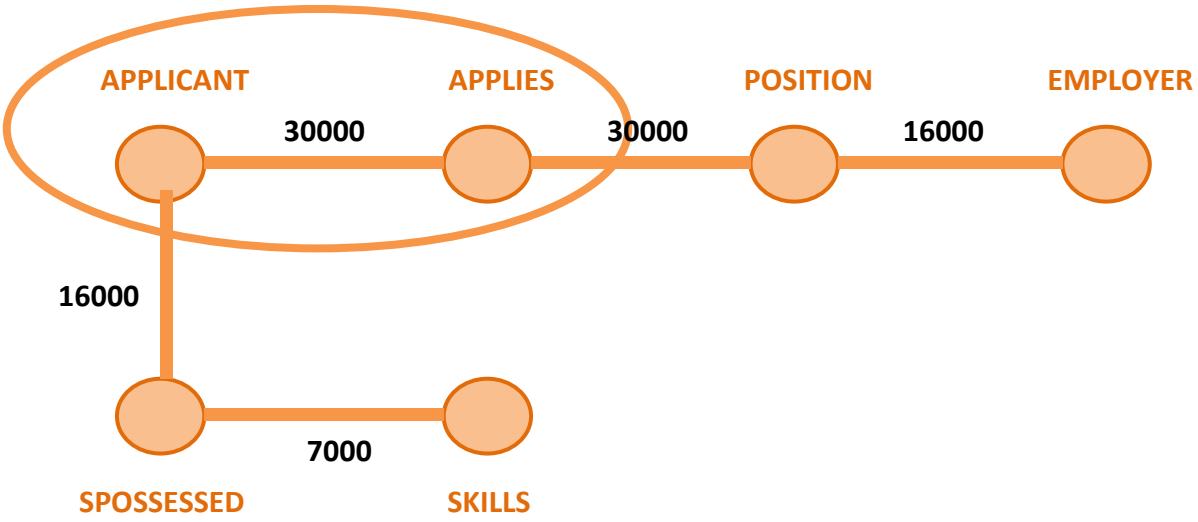
For queries i and ii (5 times per day):



Remove edges that are intersected by circle:



For queries iii, iv and v (10 times per day):



Remove edges that are intersected by circle: