

## 1 Databases (tgg22)

Suppose that we have a relational database with the following tables.

| Table                    | Primary Key    |
|--------------------------|----------------|
| Movies(mid, title, year) | mid            |
| People(pid, name)        | pid            |
| Genres(gid, genre)       | gid            |
| ActsIn(pid, mid)         | pid, mid       |
| HasRole(pid, mid, role)  | pid, mid, role |
| HasGenre(gid, mid)       | gid, mid       |

In tables `ActsIn` and `HasRole`, `pid` is a foreign key into `People` and `mid` is a foreign key into `Movies`. In table `HasGenre`, `mid` is a foreign key into `Movies` and `gid` is a foreign key into `Genres`.

Note that this database is similar to, but not the same as, the examples used in lectures and the database used for practicals.

- (a) For the table `ActsIn`, carefully explain what is meant by saying that `pid` is a foreign key into `People`. [2 marks]
- (b) Discuss potential problems this database might suffer due to data redundancy. [2 marks]
- (c) Write an SQL query that produces triples of the form `genre1, genre2, total` that count the number of movies associated with a pair of distinct genres. Each pair of genres should only appear once in the result. That is, if the triple `genre1, genre2, total` appears in the result, then the triple `genre2, genre1, total` should not. [5 marks]
- (d) Suppose that `kid` is the `pid` associated with Kevin Bacon. Write SQL that returns every `pid` for actors with a Bacon number of 2. This SQL should not include views. [5 marks]
- (e) Simplify the SQL of Part (d) using views. [6 marks]