



Habib University
shaping futures

CS343 Graph Data Science

Spring 2024

Aggregating Data

Muhammad Qasim Pasta

qasim.pasta@sse.habib.edu.pk

Count

- Returns the number of movies Tom Hanks acted in.
- Returns the number of movies released in 2000
- Returns the number of movies for each actor
- Output of the following?

```
MATCH (a:Person)-[:ACTED_IN]->(m:Movie)<-[:DIRECTED]-(d:Person)  
RETURN a.name, d.name,  
count(*) AS numMovies  
ORDER BY numMovies DESC
```

- Returns the number of actors who have acted in more than 5 movies.

WITH

- Returns the number of actors who have acted in more than 5 movies.

```
MATCH (a:Person)-[:ACTED_IN]->(m:Movie)
WITH a, count(m) AS movieCount
WHERE movieCount > 5
RETURN count(a) AS ActorsWithMoreThan5Movies
```

- Returns the average number of movies across all actors

```
MATCH (a:Person)-[:ACTED_IN]->(m:Movie)
WITH a, count(m) AS movieCount
RETURN avg(movieCount) AS AvgMoviesPerActor
```

- Average number of movies per year after 2010

Collect

- Returns the name of actors of all movies: movie name, actor names

```
MATCH (m:Movie)<-[:ACTED_IN]-(a:Person)  
RETURN m.title AS Movie, collect(a.name) AS Actors
```

- List of roles played by Tom Hanks
- Return the movies and their cast size

```
Match (m:Movie)-[:ACTED_IN]-(p:Person)  
RETURN m.title, size(collect(p.name))
```