

ex53

August 18, 2022

```
[ ]: from pyspark import SparkContext, SparkConf
     from pyspark.sql import SparkSession
     from graphframes import GraphFrame

     conf = SparkConf().setAppName("ex53")
     sc = SparkContext(conf = conf)
     ssql = SparkSession.builder.getOrCreate()
```

```
[19]: edgesPath = "data/Ex53/data/edges.csv"
      vertexesPath = "data/Ex53/data/vertexes.csv"
      outputPath = "out53/"
```

```
[4]: eDF = ssql.read.load(
      edgesPath,
      format="csv",
      header=True,
      inferSchema=True
    )

    vDF = ssql.read.load(
      vertexesPath,
      format="csv",
      header=True,
      inferSchema=True
    )
```

```
[5]: eDF.show(), eDF.printSchema()
     vDF.show(), vDF.printSchema()
```

```
+---+---+-----+
|src|dst|linktype|
+---+---+-----+
| u1| u2|  friend|
| u1| u4|  friend|
| u1| u5|  friend|
| u2| u1|  friend|
| u2| u3| follow|
| u3| u2| follow|
```

```

| u4| u1| friend|
| u4| u5| friend|
| u5| u1| friend|
| u5| u4| friend|
| u5| u6| follow|
| u6| u3| follow|
| u7| u6| follow|
+---+---+-----+

```

```

root
|-- src: string (nullable = true)
|-- dst: string (nullable = true)
|-- linktype: string (nullable = true)

```

```

+---+-----+---+
| id| name|age|
+---+-----+---+
| u1|Alice| 34|
| u2| Bob| 36|
| u3| John| 30|
| u4|David| 29|
| u5| Paul| 32|
| u6| Adel| 36|
| u7| Eddy| 60|
+---+-----+---+

```

```

root
|-- id: string (nullable = true)
|-- name: string (nullable = true)
|-- age: integer (nullable = true)

```

```
[5]: (None, None)
```

```
[6]: filteredEDF = eDF.filter("linktype='follow'")
```

```
[ ]: g = GraphFrame(vDF, filteredEDF)
```

```
[9]: nFollowerDF = g.inDegrees.withColumnRenamed("inDegree", "NFollowers")
nFollowerDF.printSchema(), nFollowerDF.show()
```

```

/home/webbelle/univenv/lib/python3.10/site-
packages/pyspark/sql/dataframe.py:127: UserWarning: DataFrame constructor is
internal. Do not directly use it.
  warnings.warn("DataFrame constructor is internal. Do not directly use it.")

```

```

root
|-- id: string (nullable = true)

```

```
|-- NFollowers: integer (nullable = false)
```

```
+---+-----+
| id|NFollowers|
+---+-----+
| u3|         2|
| u6|         2|
| u2|         1|
+---+-----+
```

```
[9]: (None, None)
```

```
[10]: maxFollowersDF = nFollowerDF\
      .agg({"NFollowers": "max"})\
      .withColumnRenamed("max(NFollowers)", "MaxNFollowers")
```

```
[15]: #se non selezionassi la colonna del DF, maxNFollowers conterrebbe SOLO un
      ↳ oggetto Row con quel valore
      maxNFollowers = maxFollowersDF.first().MaxNFollowers
      maxNFollowers
```

```
[15]: 2
```

```
[16]: nFollowerDFfinal = nFollowerDF.filter(nFollowerDF.NFollowers==maxNFollowers)
```

```
[17]: nFollowerDFfinal.show()
```

```
+---+-----+
| id|NFollowers|
+---+-----+
| u3|         2|
| u6|         2|
+---+-----+
```

```
[20]: nFollowerDFfinal.write.csv(outputPath, header=True)
```