

ex52

August 18, 2022

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

     conf = SparkConf().setAppName("ex52")
     sc = SparkContext(conf = conf)
     ssql = SparkSession.builder.getOrCreate()
```

```
[2]: edgesPath = "data/Ex52/data/edges.csv"
     vertexesPath = "data/Ex52/data/vertexes.csv"
     outputPath = "out52/"
```

```
[3]: v = ssql.read.load(
      vertexesPath,
      format="csv",
      header=True,
      inferSchema=True
    )

     e = ssql.read.load(
      edgesPath,
      format="csv",
      header=True,
      inferSchema=True
    )
```

```
[4]: v.show(), v.printSchema()
     e.show(), e.printSchema()
```

```
+---+-----+---+
| 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)
```

```
+---+---+-----+
|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)
```

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

```
[5]: filteredEdges = e.filter("linktype='follow'")
```

```
[ ]: g = GraphFrame(v, filteredEdges)
```

```
[7]: finalDF = g.inDegrees.withColumnRenamed("inDegree", "NFollowers")
```

```
/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.")
```

```
[8]: finalDF.show(), finalDF.printSchema()
```

```
+---+-----+---+
```

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

```

root
|-- id: string (nullable = true)
|-- NFollowers: integer (nullable = false)

```

[8]: (None, None)

[9]: finalDF.write.csv(outputPath, header=True)