ex54

August 18, 2022

```python
from pyspark import SparkContext, SparkConf
from pyspark.sql import SparkSession
from graphframes import GraphFrame

conf = SparkConf().setAppName("ex54")
sc = SparkContext(conf=conf)
ssql = SparkSession.builder.getOrCreate()
```

```python
[2]: edgesPath = "data/Ex54/data/edges.csv"
     vertexesPath = "data/Ex54/data/vertexes.csv"
     outputPath = "out54/"
```

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

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

```python
[5]: filteredEDF = eDF.filter("linktype='friend'")
```

```python
g = GraphFrame(vDF, filteredEDF)
```

```python
[10]: filteredG = g.find("(userX)-[]->(userY);!(userY)-[]->(userX)")
```

```python
[11]: filteredG.show(), filteredG.printSchema()
```

```
+---------------+---------------+
|          userX|          userY|
+---------------+---------------+
|{u4, David, 29}|{u1, Alice, 34}|
```

```
|{u1, Alice, 34}|  {u2, Bob, 36}|
+--------------+--------------+

root
 |-- userX: struct (nullable = false)
 |     |-- id: string (nullable = true)
 |     |-- name: string (nullable = true)
 |     |-- age: integer (nullable = true)
 |-- userY: struct (nullable = false)
 |     |-- id: string (nullable = true)
 |     |-- name: string (nullable = true)
 |     |-- age: integer (nullable = true)
```

[11]: (None, None)

[12]:
```python
formatteDFG = filteredG.selectExpr("userX.id as IDFriend", "userY.id as
 ↪IDNotFriend")
```

[13]:
```python
formatteDFG.write.csv(outputPath, header=True)
```