In [2]:

```python
user_mentions = spark.read.csv("s3://us-congress-tweets/user_mentions.csv", header=
True)
```

In [6]:

```
No module named 'networkx'
Traceback (most recent call last):
ModuleNotFoundError: No module named 'networkx'
```

In [7]:

```python
# user_mentions.printSchema()
user_mentions.show(20)
```

```
+------------------+------------------+-----+
|               src|               dst|count|
+------------------+------------------+-----+
|917194889275699201|        1249982359|    1|
|917570582555779072|         251918778|    1|
|         633674091|         432895323|   48|
|913222391416934402|         432895323|  261|
|         217574712|          47203904|    1|
|931278145679847424|         320757267|    1|
|897332217646522368|          92186819|    8|
|        2308265716|          18061669|    4|
|822826488500088832|        1249982359|  132|
|728645535566008320|958064770019741696|    1|
|827323185058033665|          14247236|   11|
|          36566383|826629809954553856|    2|
|913125313172918272|         432895323|   90|
|860643294367318016|         432895323|   52|
|866790059802079234|         432895323|  148|
|888055635828502530|          16056306|    2|
|        3141820397|         112006107|    1|
|836681162638532608|         262756641|    1|
|         157952243|         102391944|    1|
|         334871740|          16117316|   17|
+------------------+------------------+-----+
only showing top 20 rows
```

In [3]:

```
# seed_node = user_mentions.filter(user_mentions.src == 15764644)

user_mentions_sample = user_mentions.limit(1000)
```

In [70]:

```
user_mentions_sample.count()
```

1000

In [5]:

```
from graphframes import *
from graphframes.lib import Pregel
sc.setCheckpointDir("hdfs:///tmp/graphframes_checkpoint") # this is needed for any
 GraphFrames operation
```

In [28]:

```
seed_node.show()
```

```
+--------+------------------+-----+
|     src|               dst|count|
+--------+------------------+-----+
|15764644|         199325935|    5|
|15764644|        1092979962|    4|
|15764644|         381152398|    6|
|15764644|1079854463211397120|   2|
|15764644|        2916086925|    2|
|15764644|         294084341|    1|
|15764644| 716458790229581824|   1|
|15764644|        1051127714|    2|
|15764644| 855482971868069890|   3|
|15764644|        2966570782|    2|
|15764644|1076161611033968640|   2|
|15764644|         122453931|    1|
|15764644|          14857525|    1|
|15764644|         233783568|    3|
|15764644|1078355119920562176|   6|
|15764644| 827258161841135623|   1|
|15764644|        2244340904|    1|
|15764644|        1339931490|    1|
|15764644|          15751055|    1|
|15764644|         584912320|    2|
+--------+------------------+-----+
only showing top 20 rows
```

In [6]:

```
'''
edges = spark.createDataFrame([
    ['a', 'b'],
    ['b', 'd'],
    ['b', 'c'],
    ['c', 'b'],
    ['d', 'c'],
    ['d', 'a']
], ["src", "dst"]).cache()
edges.show()
'''
edges = user_mentions_sample.select("src","dst");
```

In [72]:

```
edges.show()
```

```
+------------------+------------------+
|               src|               dst|
+------------------+------------------+
|917194889275699201|        1249982359|
|917570582555779072|         251918778|
|         633674091|         432895323|
|913222391416934402|         432895323|
|         217574712|          47203904|
|931278145679847424|         320757267|
|897332217646522368|          92186819|
|        2308265716|          18061669|
|822826488500088832|        1249982359|
|728645535566008320|958064770019741696|
|827323185058033665|          14247236|
|          36566383|826629809954553856|
|913125313172918272|         432895323|
|860643294367318016|         432895323|
|866790059802079234|         432895323|
|888055635828502530|          16056306|
|        3141820397|         112006107|
|836681162638532608|         262756641|
|         157952243|         102391944|
|         334871740|          16117316|
+------------------+------------------+
only showing top 20 rows
```

In [7]:

```python
df_1_new = edges.select('src').distinct().alias('id')
df_2_new = edges.select('dst')

vertices = df_1_new.union(df_2_new)


vertices.show()
```

```
+-------------------+
|                src|
+-------------------+
|1092882255331160065|
|          313646333|
|  959535310328737792|
|           34633253|
|           20778273|
|1012493637840760832|
|          433908976|
|1042897077737611271|
|          275239133|
|           22001536|
|          861561900|
|1087405828410392579|
|           20164802|
|          546366633|
|  880629474286419968|
|          121847226|
|          172954545|
|  719742042616827904|
|          928111574|
|  890740621744390144|
+-------------------+
only showing top 20 rows
```

In [79]:

```
vertices.show()
```

```
+------------------+
|                id|
+------------------+
|917194889275699201|
|917570582555779072|
|         633674091|
|913222391416934402|
|         217574712|
|931278145679847424|
|897332217646522368|
|        2308265716|
|822826488500088832|
|728645535566008320|
|827323185058033665|
|          36566383|
|913125313172918272|
|860643294367318016|
|866790059802079234|
|888055635828502530|
|        3141820397|
|836681162638532608|
|         157952243|
|         334871740|
+------------------+
only showing top 20 rows
```

In [13]:

```
vertices = vertices.selectExpr("src as id")
vertices.printSchema()
```

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

In [14]:

```
graph = GraphFrame(vertices, edges)
```

In [15]:

```
vertices = graph.outDegrees
vertices.show()
```

```
+-------------------+---------+
|                 id|outDegree|
+-------------------+---------+
| 986756647128150016|        1|
| 741932068347613185|        1|
|           43508994|        2|
|1065577571813867521|        1|
| 823163616966901760|        1|
|         2214354650|        1|
|1043932566053507072|        1|
|           18652786|        1|
| 823651149823688705|        1|
|          124822967|        1|
| 897596565589831682|        1|
|         2937150214|        1|
|1040418473849937920|        1|
|1067162687359119360|        1|
| 936842433186467840|        1|
|          142039570|        1|
|          453884603|        1|
| 996047496139984897|        1|
|          276443066|        1|
|          117457725|        1|
+-------------------+---------+
only showing top 20 rows
```

In [16]:

```
# Redefine Graph
graph = GraphFrame(vertices, edges)
```

In [8]:

In [53]:

```
+--------+---------+
|      id|outDegree|
+--------+---------+
|15764644|      342|
+--------+---------+
```

In [54]:

In [ ]:

```
import pyspark.sql.functions as F # needed for defining literals below and the sum
(...) function
ranks = graph.pregel  \
        .setMaxIter(1) \
        .withVertexColumn("rank",  F.lit(1.0),   Pregel.msg() * F.lit(0.85) + F.lit
(0.15)  ) \
        .sendMsgToDst(Pregel.src("rank")/Pregel.src("outDegree")) \
        .aggMsgs(F.sum(Pregel.msg()))
        .run()
ranks.show()
```

In [62]:

```
No module named 'pyrwr'
Traceback (most recent call last):
ModuleNotFoundError: No module named 'pyrwr'
```

In [91]:

```
results = graph.pageRank(resetProbability=0.15, maxIter=5, sourceId="313646333")
results.vertices.show()
```

```
An error occurred while calling o743.run.
: org.graphframes.NoSuchVertexException: GraphFrame algorithm given ver
tex ID which does not exist in Graph. Vertex ID 313646333 not contained
in GraphFrame(v:[id: string, outDegree: int], e:[src: string, dst: stri
ng])
        at org.graphframes.lib.GraphXConversions$.integralId(GraphXConv
ersions.scala:184)
        at org.graphframes.lib.PageRank$$anonfun$1.apply(PageRank.scal
a:129)
        at org.graphframes.lib.PageRank$$anonfun$1.apply(PageRank.scal
a:129)
        at scala.Option.map(Option.scala:146)
        at org.graphframes.lib.PageRank$.run(PageRank.scala:129)
        at org.graphframes.lib.PageRank.run(PageRank.scala:104)
        at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
        at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAcce
ssorImpl.java:62)
        at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMe
thodAccessorImpl.java:43)
        at java.lang.reflect.Method.invoke(Method.java:498)
        at py4j.reflection.MethodInvoker.invoke(MethodInvoker.java:244)
        at py4j.reflection.ReflectionEngine.invoke(ReflectionEngine.jav
a:357)
        at py4j.Gateway.invoke(Gateway.java:282)
        at py4j.commands.AbstractCommand.invokeMethod(AbstractCommand.j
ava:132)
        at py4j.commands.CallCommand.execute(CallCommand.java:79)
        at py4j.GatewayConnection.run(GatewayConnection.java:238)
        at java.lang.Thread.run(Thread.java:748)


Traceback (most recent call last):
  File "/mnt/tmp/spark-d4343136-a7d9-4dd4-b908-158defbfdf90/userFiles-c
cccc57c-8049-4c4e-a3e5-1a2c39288afb/graphframes_graphframes-0.7.0-spark
2.4-s_2.11.jar/graphframes/graphframe.py", line 371, in pageRank
    jgf = builder.run()
  File "/usr/lib/spark/python/lib/py4j-0.10.7-src.zip/py4j/java_gatewa
y.py", line 1257, in __call__
    answer, self.gateway_client, self.target_id, self.name)
  File "/usr/lib/spark/python/lib/pyspark.zip/pyspark/sql/utils.py", li
ne 63, in deco
    return f(*a, **kw)
  File "/usr/lib/spark/python/lib/py4j-0.10.7-src.zip/py4j/protocol.p
y", line 328, in get_return_value
    format(target_id, ".", name), value)
py4j.protocol.Py4JJavaError: An error occurred while calling o743.run.
: org.graphframes.NoSuchVertexException: GraphFrame algorithm given ver
tex ID which does not exist in Graph. Vertex ID 313646333 not contained
in GraphFrame(v:[id: string, outDegree: int], e:[src: string, dst: stri
ng])
        at org.graphframes.lib.GraphXConversions$.integralId(GraphXConv
ersions.scala:184)
        at org.graphframes.lib.PageRank$$anonfun$1.apply(PageRank.scal
a:129)
        at org.graphframes.lib.PageRank$$anonfun$1.apply(PageRank.scal
a:129)
        at scala.Option.map(Option.scala:146)
        at org.graphframes.lib.PageRank$.run(PageRank.scala:129)
```

```
        at org.graphframes.lib.PageRank.run(PageRank.scala:104)
        at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
        at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAcce
ssorImpl.java:62)
        at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMe
thodAccessorImpl.java:43)
        at java.lang.reflect.Method.invoke(Method.java:498)
        at py4j.reflection.MethodInvoker.invoke(MethodInvoker.java:244)
        at py4j.reflection.ReflectionEngine.invoke(ReflectionEngine.jav
a:357)
        at py4j.Gateway.invoke(Gateway.java:282)
        at py4j.commands.AbstractCommand.invokeMethod(AbstractCommand.j
ava:132)
        at py4j.commands.CallCommand.execute(CallCommand.java:79)
        at py4j.GatewayConnection.run(GatewayConnection.java:238)
        at java.lang.Thread.run(Thread.java:748)
```

In [67]:

No module named 'matplotlib'
Traceback (most recent call last):
ModuleNotFoundError: No module named 'matplotlib'

In [30]:

```
from graphframes.examples import Graphs

g = Graphs(sqlContext).friends()  # Get example graph
# g = graph

# Run PageRank personalized for vertex "a"
results3 = g.pageRank(resetProbability=0.15, maxIter=10)
```

In [ ]:

```
g
```

In [31]:

```
results3.edges.show()
```

```
+---+---+------------+------+
|src|dst|relationship|weight|
+---+---+------------+------+
|  a|  b|      friend|   1.0|
|  b|  c|      follow|   1.0|
|  e|  f|      follow|   0.5|
|  e|  d|      friend|   0.5|
|  c|  b|      follow|   1.0|
|  f|  c|      follow|   1.0|
|  d|  a|      friend|   1.0|
+---+---+------------+------+
```

In [32]:

```
results3.vertices.show()
```

```
+---+-------+---+------------------+
| id|   name|age|          pagerank|
+---+-------+---+------------------+
|  b|    Bob| 36|2.5724999999999993|
|  e| Esther| 32|              0.15|
|  a|  Alice| 34|         0.3316875|
|  f|  Fanny| 36|           0.21375|
|  d|  David| 29|           0.21375|
|  c|Charlie| 30|2.5183124999999995|
+---+-------+---+------------------+
```

In [16]:

```
user_mentions = spark.read.csv("s3://us-congress-tweets/user_mentions.csv", header=
True)
```

In [6]:

```
No module named 'networkx'
Traceback (most recent call last):
ModuleNotFoundError: No module named 'networkx'
```

In [7]:

```
# user_mentions.printSchema()
user_mentions.show(20)
```

```
+-----------------+-----------------+-----+
|              src|              dst|count|
+-----------------+-----------------+-----+
|917194889275699201|       1249982359|    1|
|917570582555779072|        251918778|    1|
|        633674091|        432895323|   48|
|913222391416934402|        432895323|  261|
|        217574712|         47203904|    1|
|931278145679847424|        320757267|    1|
|897332217646522368|         92186819|    8|
|        2308265716|         18061669|    4|
|822826488500088832|       1249982359|  132|
|728645535566008320|958064770019741696|    1|
|827323185058033665|         14247236|   11|
|         36566383|826629809954553856|    2|
|913125313172918272|        432895323|   90|
|860643294367318016|        432895323|   52|
|866790059802079234|        432895323|  148|
|888055635828502530|         16056306|    2|
|        3141820397|        112006107|    1|
|836681162638532608|        262756641|    1|
|        157952243|        102391944|    1|
|        334871740|         16117316|   17|
+-----------------+-----------------+-----+
only showing top 20 rows
```

In [69]:

```
# seed_node = user_mentions.filter(user_mentions.src == 15764644)

user_mentions_sample = user_mentions.limit(1000)
```

In [70]:

```
user_mentions_sample.count()
```

```
1000
```

In [1]:

```
from graphframes import *
from graphframes.lib import Pregel
sc.setCheckpointDir("hdfs:///tmp/graphframes_checkpoint") # this is needed for any
 GraphFrames operation
```

Starting Spark application

| ID | YARN Application ID | Kind | State | |
|---|---|---|---|---|
| 0 | application_1572570650052_0001 | pyspark | idle | Link (http://i 89.ec2.internal:20888/proxy/application_157257065( |

SparkSession available as 'spark'.

In [28]:

```
seed_node.show()
```

```
+--------+------------------+-----+
|     src|               dst|count|
+--------+------------------+-----+
|15764644|         199325935|    5|
|15764644|        1092979962|    4|
|15764644|         381152398|    6|
|15764644|1079854463211397120|    2|
|15764644|        2916086925|    2|
|15764644|         294084341|    1|
|15764644|  716458790229581824|    1|
|15764644|        1051127714|    2|
|15764644|  855482971868069890|    3|
|15764644|        2966570782|    2|
|15764644|1076161611033968640|    2|
|15764644|         122453931|    1|
|15764644|          14857525|    1|
|15764644|         233783568|    3|
|15764644|1078355119920562176|    6|
|15764644|  827258161841135623|    1|
|15764644|        2244340904|    1|
|15764644|        1339931490|    1|
|15764644|          15751055|    1|
|15764644|         584912320|    2|
+--------+------------------+-----+
only showing top 20 rows
```

In [71]:

```
'''
edges = spark.createDataFrame([
    ['a', 'b'],
    ['b', 'd'],
    ['b', 'c'],
    ['c', 'b'],
    ['d', 'c'],
    ['d', 'a']
], ["src", "dst"]).cache()
edges.show()
'''
edges = user_mentions_sample.select("src","dst");
```

In [72]:

```
edges.show()
```

```
+------------------+------------------+
|               src|               dst|
+------------------+------------------+
|917194889275699201|        1249982359|
|917570582555779072|         251918778|
|         633674091|         432895323|
|913222391416934402|         432895323|
|         217574712|          47203904|
|931278145679847424|         320757267|
|897332217646522368|          92186819|
|        2308265716|          18061669|
|822826488500088832|        1249982359|
|728645535566008320|958064770019741696|
|827323185058033665|          14247236|
|          36566383|826629809954553856|
|913125313172918272|         432895323|
|860643294367318016|         432895323|
|866790059802079234|         432895323|
|888055635828502530|          16056306|
|        3141820397|         112006107|
|836681162638532608|         262756641|
|         157952243|         102391944|
|         334871740|          16117316|
+------------------+------------------+
only showing top 20 rows
```

In [73]:

```python
df_1_new = edges.select('src').distinct().alias('id')
df_2_new = edges.select('dst')

vertices = df_1_new.union(df_2_new)

vertices.withColumnRenamed("src","DateOfBirth")
vertices.show()
```

```
+------------------+
|               src|
+------------------+
|1074837182253285378|
|         217517332|
|        3071880606|
| 746504825374810112|
|         440516000|
| 812408092205850624|
| 787375404839477252|
|        1969388486|
|        1677066894|
|        4663053613|
|1000786614417788932|
|          18544142|
| 962211528773484544|
|         268112428|
| 703059273022963713|
|         575649944|
| 787772818238865409|
|         140820196|
|          76964000|
|         217639183|
+------------------+
only showing top 20 rows
```

In [79]:

```
vertices.show()
```

```
+------------------+
|                id|
+------------------+
|917194889275699201|
|917570582555779072|
|         633674091|
|913222391416934402|
|         217574712|
|931278145679847424|
|897332217646522368|
|        2308265716|
|822826488500088832|
|728645535566008320|
|827323185058033665|
|          36566383|
|913125313172918272|
|860643294367318016|
|866790059802079234|
|888055635828502530|
|        3141820397|
|836681162638532608|
|         157952243|
|         334871740|
+------------------+
only showing top 20 rows
```

In [82]:

```
# vertices = vertices.selectExpr("src as id")
vertices.printSchema()
```

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

In [85]:

```
graph = GraphFrame(vertices, edges)
```

In [86]:

```
vertices = graph.outDegrees
vertices.show()
```

```
+-------------------+---------+
|                 id|outDegree|
+-------------------+---------+
|1092882255331160065|        1|
|          313646333|        1|
|  959535310328737792|        1|
|           34633253|        1|
|           20778273|        1|
|1012493637840760832|        1|
|          433908976|        1|
|1042897077737611271|        1|
|          275239133|        1|
|           22001536|        1|
|          861561900|        1|
|1087405828410392579|        1|
|           20164802|        1|
|          546366633|        1|
|  880629474286419968|        1|
|          121847226|        1|
|          172954545|        1|
|  719742042616827904|        1|
|          928111574|        2|
|  890740621744390144|        1|
+-------------------+---------+
only showing top 20 rows
```

In [88]:

```
graph = GraphFrame(vertices, edges)
```

In [8]:

In [53]:

```
+--------+---------+
|      id|outDegree|
+--------+---------+
|15764644|      342|
+--------+---------+
```

In [54]:

In [ ]:

```python
import pyspark.sql.functions as F # needed for defining literals below and the sum
(...) function
ranks = graph.pregel  \
        .setMaxIter(1) \
        .withVertexColumn("rank",  F.lit(1.0),   Pregel.msg() * F.lit(0.85) + F.lit
(0.15)  ) \
        .sendMsgToDst(Pregel.src("rank")/Pregel.src("outDegree")) \
        .aggMsgs(F.sum(Pregel.msg()))
        .run()
ranks.show()
```

In [62]:

```
No module named 'pyrwr'
Traceback (most recent call last):
ModuleNotFoundError: No module named 'pyrwr'
```

In [91]:

```
results = graph.pageRank(resetProbability=0.15, maxIter=5, sourceId="313646333")
results.vertices.show()
```

```
An error occurred while calling o743.run.
: org.graphframes.NoSuchVertexException: GraphFrame algorithm given ver
tex ID which does not exist in Graph. Vertex ID 313646333 not contained
in GraphFrame(v:[id: string, outDegree: int], e:[src: string, dst: stri
ng])
        at org.graphframes.lib.GraphXConversions$.integralId(GraphXConv
ersions.scala:184)
        at org.graphframes.lib.PageRank$$anonfun$1.apply(PageRank.scal
a:129)
        at org.graphframes.lib.PageRank$$anonfun$1.apply(PageRank.scal
a:129)
        at scala.Option.map(Option.scala:146)
        at org.graphframes.lib.PageRank$.run(PageRank.scala:129)
        at org.graphframes.lib.PageRank.run(PageRank.scala:104)
        at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
        at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAcce
ssorImpl.java:62)
        at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMe
thodAccessorImpl.java:43)
        at java.lang.reflect.Method.invoke(Method.java:498)
        at py4j.reflection.MethodInvoker.invoke(MethodInvoker.java:244)
        at py4j.reflection.ReflectionEngine.invoke(ReflectionEngine.jav
a:357)
        at py4j.Gateway.invoke(Gateway.java:282)
        at py4j.commands.AbstractCommand.invokeMethod(AbstractCommand.j
ava:132)
        at py4j.commands.CallCommand.execute(CallCommand.java:79)
        at py4j.GatewayConnection.run(GatewayConnection.java:238)
        at java.lang.Thread.run(Thread.java:748)


Traceback (most recent call last):
  File "/mnt/tmp/spark-d4343136-a7d9-4dd4-b908-158defbfdf90/userFiles-c
cccc57c-8049-4c4e-a3e5-1a2c39288afb/graphframes_graphframes-0.7.0-spark
2.4-s_2.11.jar/graphframes/graphframe.py", line 371, in pageRank
    jgf = builder.run()
  File "/usr/lib/spark/python/lib/py4j-0.10.7-src.zip/py4j/java_gatewa
y.py", line 1257, in __call__
    answer, self.gateway_client, self.target_id, self.name)
  File "/usr/lib/spark/python/lib/pyspark.zip/pyspark/sql/utils.py", li
ne 63, in deco
    return f(*a, **kw)
  File "/usr/lib/spark/python/lib/py4j-0.10.7-src.zip/py4j/protocol.p
y", line 328, in get_return_value
    format(target_id, ".", name), value)
py4j.protocol.Py4JJavaError: An error occurred while calling o743.run.
: org.graphframes.NoSuchVertexException: GraphFrame algorithm given ver
tex ID which does not exist in Graph. Vertex ID 313646333 not contained
in GraphFrame(v:[id: string, outDegree: int], e:[src: string, dst: stri
ng])
        at org.graphframes.lib.GraphXConversions$.integralId(GraphXConv
ersions.scala:184)
        at org.graphframes.lib.PageRank$$anonfun$1.apply(PageRank.scal
a:129)
        at org.graphframes.lib.PageRank$$anonfun$1.apply(PageRank.scal
a:129)
        at scala.Option.map(Option.scala:146)
        at org.graphframes.lib.PageRank$.run(PageRank.scala:129)
```

```
        at org.graphframes.lib.PageRank.run(PageRank.scala:104)
        at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
        at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAcce
ssorImpl.java:62)
        at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMe
thodAccessorImpl.java:43)
        at java.lang.reflect.Method.invoke(Method.java:498)
        at py4j.reflection.MethodInvoker.invoke(MethodInvoker.java:244)
        at py4j.reflection.ReflectionEngine.invoke(ReflectionEngine.jav
a:357)
        at py4j.Gateway.invoke(Gateway.java:282)
        at py4j.commands.AbstractCommand.invokeMethod(AbstractCommand.j
ava:132)
        at py4j.commands.CallCommand.execute(CallCommand.java:79)
        at py4j.GatewayConnection.run(GatewayConnection.java:238)
        at java.lang.Thread.run(Thread.java:748)
```

In [67]:

```
No module named 'matplotlib'
Traceback (most recent call last):
ModuleNotFoundError: No module named 'matplotlib'
```

In [1]:

```
from graphframes.examples import Graphs

# g = Graphs(sqlContext).friends()  # Get example graph
g = graph

# Run PageRank personalized for vertex "a"
results3 = g.pageRank(resetProbability=0.15, maxIter=2)
```

Starting Spark application

| ID | YARN Application ID | Kind | State | |
|----|---------------------|------|-------|---|
| 1 | application_1572620638272_0002 | pyspark | idle | Link (http://136.ec2.internal:20888/proxy/application_15726206? |

SparkSession available as 'spark'.

```
name 'graph' is not defined
Traceback (most recent call last):
NameError: name 'graph' is not defined
```

In [111]:

```
results3.vertices.show()
```

```
+---+---------+--------+
| id|outDegree|pagerank|
+---+---------+--------+
+---+---------+--------+
```

In [113]:

```
from graphframes.examples import Graphs
g = Graphs(sqlContext).friends()  # Get example graph

graph.vertices.show()

results3 = edge.pageRank(resetProbability=0.15, maxIter=2, sourceId="a")
```

```
'DataFrame' object has no attribute 'createDataFrame'
Traceback (most recent call last):
  File "/mnt/tmp/spark-d4343136-a7d9-4dd4-b908-158defbfdf90/userFiles-c
cccc57c-8049-4c4e-a3e5-1a2c39288afb/graphframes_graphframes-0.7.0-spark
2.4-s_2.11.jar/graphframes/examples/graphs.py", line 41, in friends
    v = sqlContext.createDataFrame([
  File "/usr/lib/spark/python/lib/pyspark.zip/pyspark/sql/dataframe.p
y", line 1301, in __getattr__
    "'%s' object has no attribute '%s'" % (self.__class__.__name__, nam
e))
AttributeError: 'DataFrame' object has no attribute 'createDataFrame'
```

In [107]:

```
results3.edege.show()
```

```
'GraphFrame' object has no attribute 'edge'
Traceback (most recent call last):
AttributeError: 'GraphFrame' object has no attribute 'edge'
```

In [106]:

```
graph.vertices.show()
```

```
+-------------------+---------+
|                 id|outDegree|
+-------------------+---------+
|          202602593|        1|
|1101888810395275264|        1|
|         3985859614|        1|
|1094586924621283329|        1|
| 911775858045599744|        1|
| 800848709818806273|        1|
|          544411686|        1|
| 773540449965510656|        1|
|           32267724|        1|
|          279358548|        1|
| 825488267340632065|        1|
|         4816438815|        1|
|          801055502|        1|
|           41713818|        1|
|1025772687203999744|        1|
|          326957691|        1|
|          523509739|        1|
|          172829760|        1|
|          843606578|        1|
|1042286660203896832|        1|
+-------------------+---------+
only showing top 20 rows
```

In [109]:

```
results3 = edges.pageRank(resetProbability=0.15, maxIter=2, sourceId="15764644")
```

```
'DataFrame' object has no attribute 'pageRank'
Traceback (most recent call last):
  File "/usr/lib/spark/python/lib/pyspark.zip/pyspark/sql/dataframe.p
y", line 1301, in __getattr__
    "'%s' object has no attribute '%s'" % (self.__class__.__name__, nam
e))
AttributeError: 'DataFrame' object has no attribute 'pageRank'
```

In [114]:

```
user_mentions_sample.printSchema()
```

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

In [ ]:

```
print(1)
```

In [ ]:

In [113]:

```
from graphframes.examples import Graphs
g = Graphs(sqlContext).friends()  # Get example graph

graph.vertices.show()

results3 = edge.pageRank(resetProbability=0.15, maxIter=5, sourceId="a")
```

```
'DataFrame' object has no attribute 'createDataFrame'
Traceback (most recent call last):
  File "/mnt/tmp/spark-d4343136-a7d9-4dd4-b908-158defbfdf90/userFiles-c
cccc57c-8049-4c4e-a3e5-1a2c39288afb/graphframes_graphframes-0.7.0-spark
2.4-s_2.11.jar/graphframes/examples/graphs.py", line 41, in friends
    v = sqlContext.createDataFrame([
  File "/usr/lib/spark/python/lib/pyspark.zip/pyspark/sql/dataframe.p
y", line 1301, in __getattr__
    "'%s' object has no attribute '%s'" % (self.__class__.__name__, nam
e))
AttributeError: 'DataFrame' object has no attribute 'createDataFrame'
```

In [107]:

```
results3.edege.show()
```

```
'GraphFrame' object has no attribute 'edge'
Traceback (most recent call last):
AttributeError: 'GraphFrame' object has no attribute 'edge'
```

In [106]:

```
graph.vertices.show()
```

```
+------------------+---------+
|                id|outDegree|
+------------------+---------+
|         202602593|        1|
|1101888810395275264|       1|
|        3985859614|        1|
|1094586924621283329|       1|
| 911775858045599744|       1|
| 800848709818806273|       1|
|         544411686|        1|
| 773540449965510656|       1|
|          32267724|        1|
|         279358548|        1|
| 825488267340632065|       1|
|        4816438815|        1|
|         801055502|        1|
|          41713818|        1|
|1025772687203999744|       1|
|         326957691|        1|
|         523509739|        1|
|         172829760|        1|
|         843606578|        1|
|1042286660203896832|       1|
+------------------+---------+
only showing top 20 rows
```

In [109]:

```
results3 = edges.pageRank(resetProbability=0.15, maxIter=2, sourceId="a")
```

```
'DataFrame' object has no attribute 'pageRank'
Traceback (most recent call last):
  File "/usr/lib/spark/python/lib/pyspark.zip/pyspark/sql/dataframe.p
y", line 1301, in __getattr__
    "'%s' object has no attribute '%s'" % (self.__class__.__name__, nam
e))
AttributeError: 'DataFrame' object has no attribute 'pageRank'
```

In [114]:

```
user_mentions_sample.printSchema()
```

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

In [ ]:

```
print(1)
```

In [ ]: