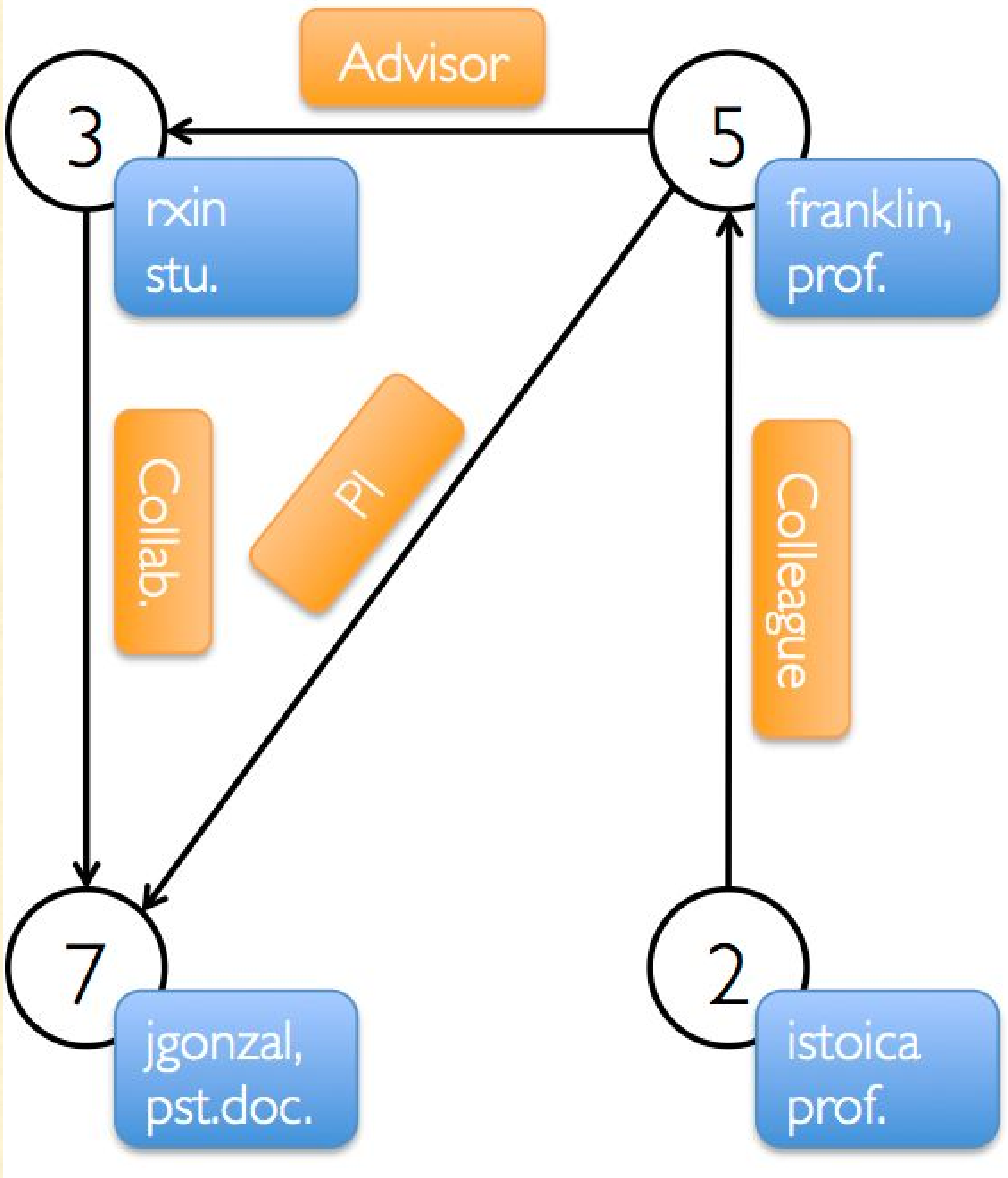
APACHE Spark™

GraphX

What is a graph

Property Graph



Vertex Table

Id	Property (V)
3	(rxin, student)
7	(jgonzal, postdoc)
5	(franklin, professor)
2	(istoica, professor)

Edge Table

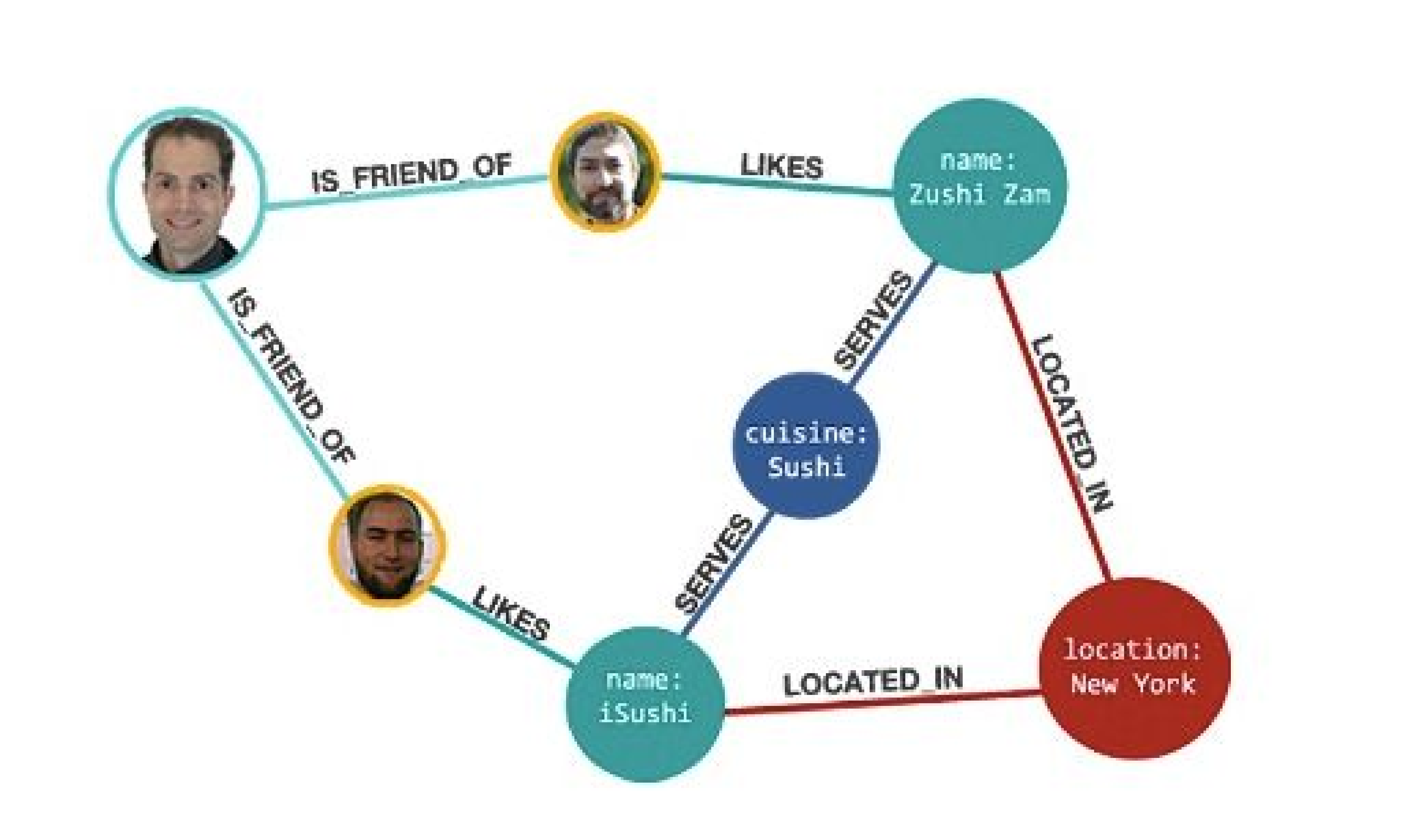
SrcId	DstId	Property (E)
3	7	Collaborator
5	3	Advisor
2	5	Colleague
5	7	PI

Examples of graph computations

- Finding common friends
- Finding the page rank
- And Many more…

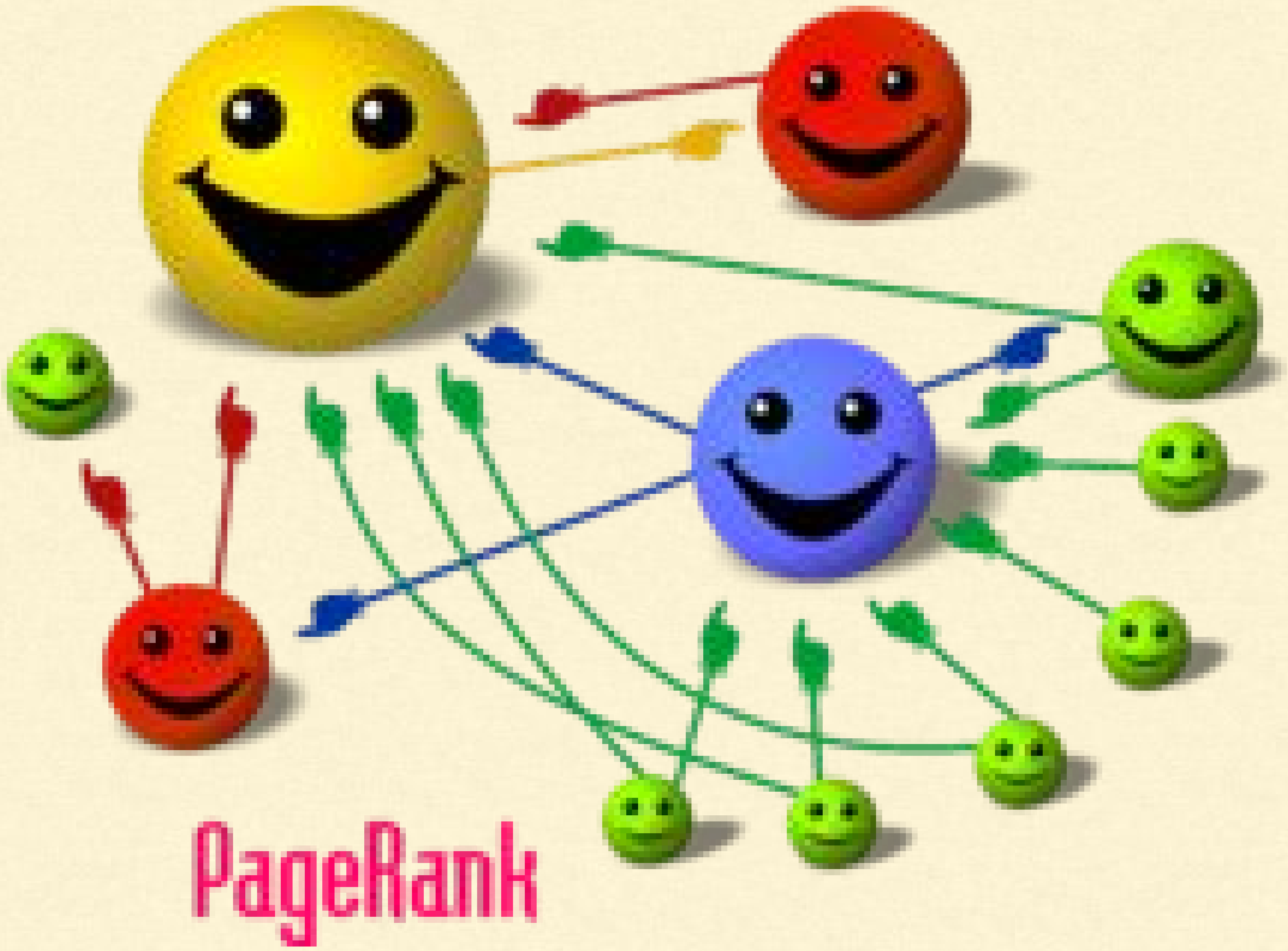
Examples of graph computations

Finding common friends



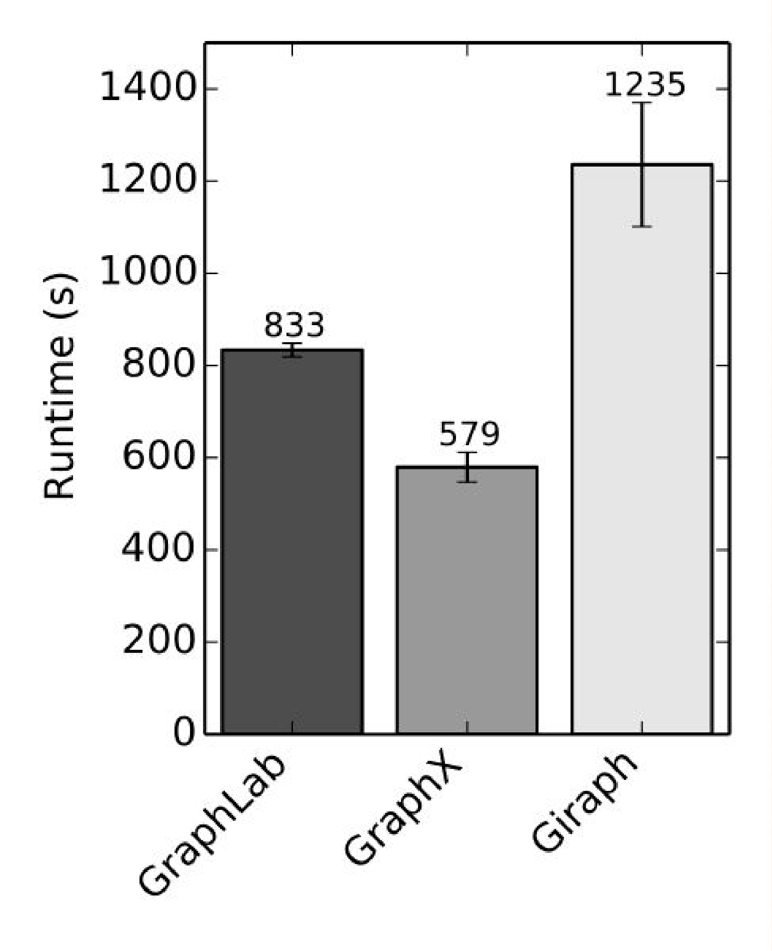
Examples of graph computations

Finding Page Rank



GraphX

- Unifies Graph Computation
  - ETL
  - Exploratory analysis
  - Iterative
- View the same Data as Graph and Collections
- Transform and join graphs with RDDs efficiently
- Extends the Spark RDD by introducing a new Graph abstraction



GraphX

Has library of algorithms

- **PageRank**
  - If important pages link you, you are more important
- **Connected Components**
  - Clusters amongst your facebook friends
- **Triangle Counting**
  - Triangles passing through each vertex => measure of clustering.
- **Label propagation**
- **SVD++**
- **Strongly connected components**

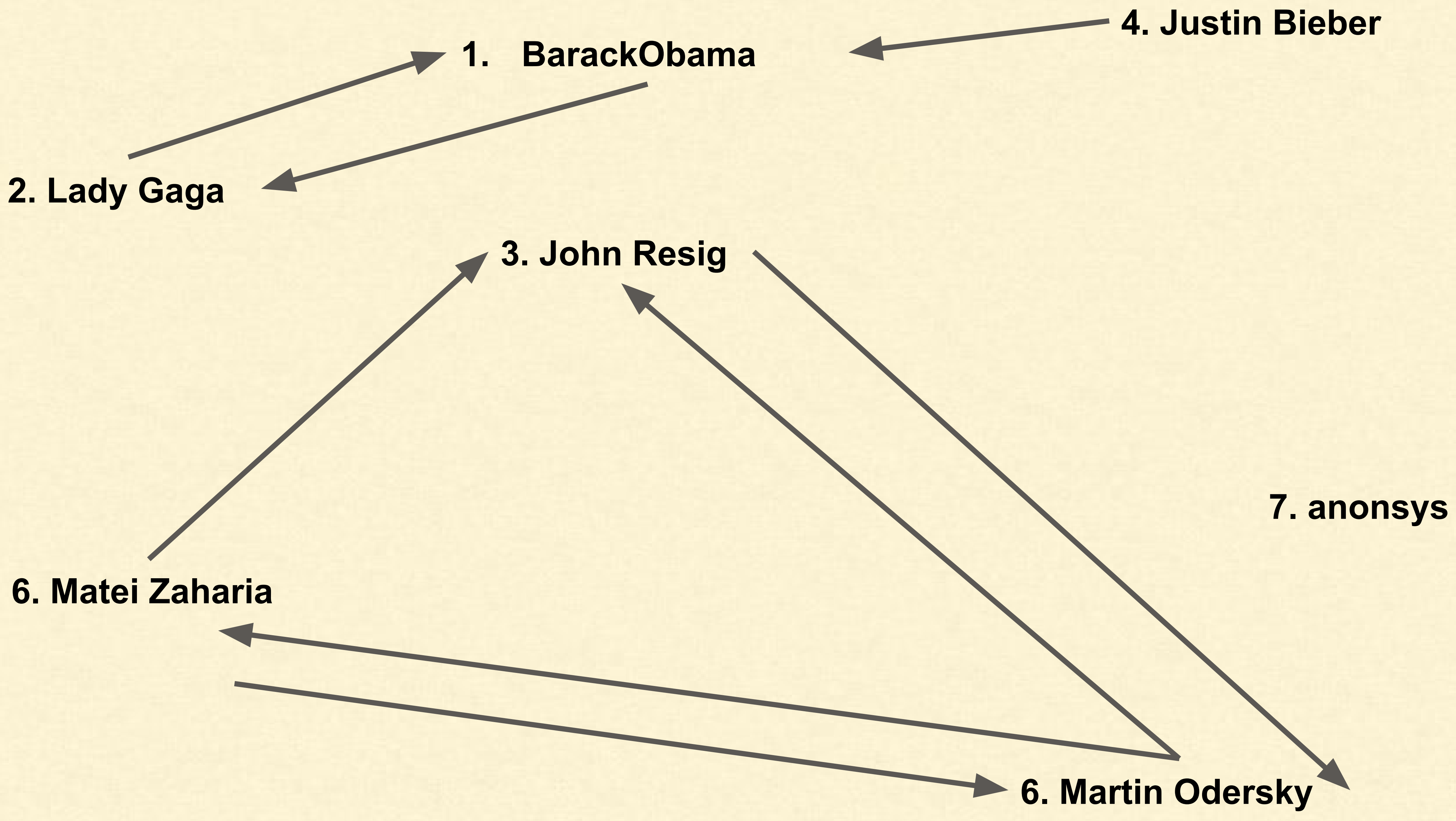
GraphX

Provides set of fundamental operations

- subgraph
- joinVertices
- aggregateMessages
- And more….

https://spark.apache.org/docs/latest/graphx-programming-guide.html

GraphX - Pagerank



*PR(A) = 0.15 + 0.85 * (rank of node / outgoing)*

GraphX - Pagerank

```
$ hadoop fs -cat /data/spark/graphx/followers.txt
2 1
4 1
1 2
6 3
7 3
7 6
6 7
3 7
```

https://github.com/cloudxlab/bigdata/blob/master/spark/examples/graphx/pagerank.scala

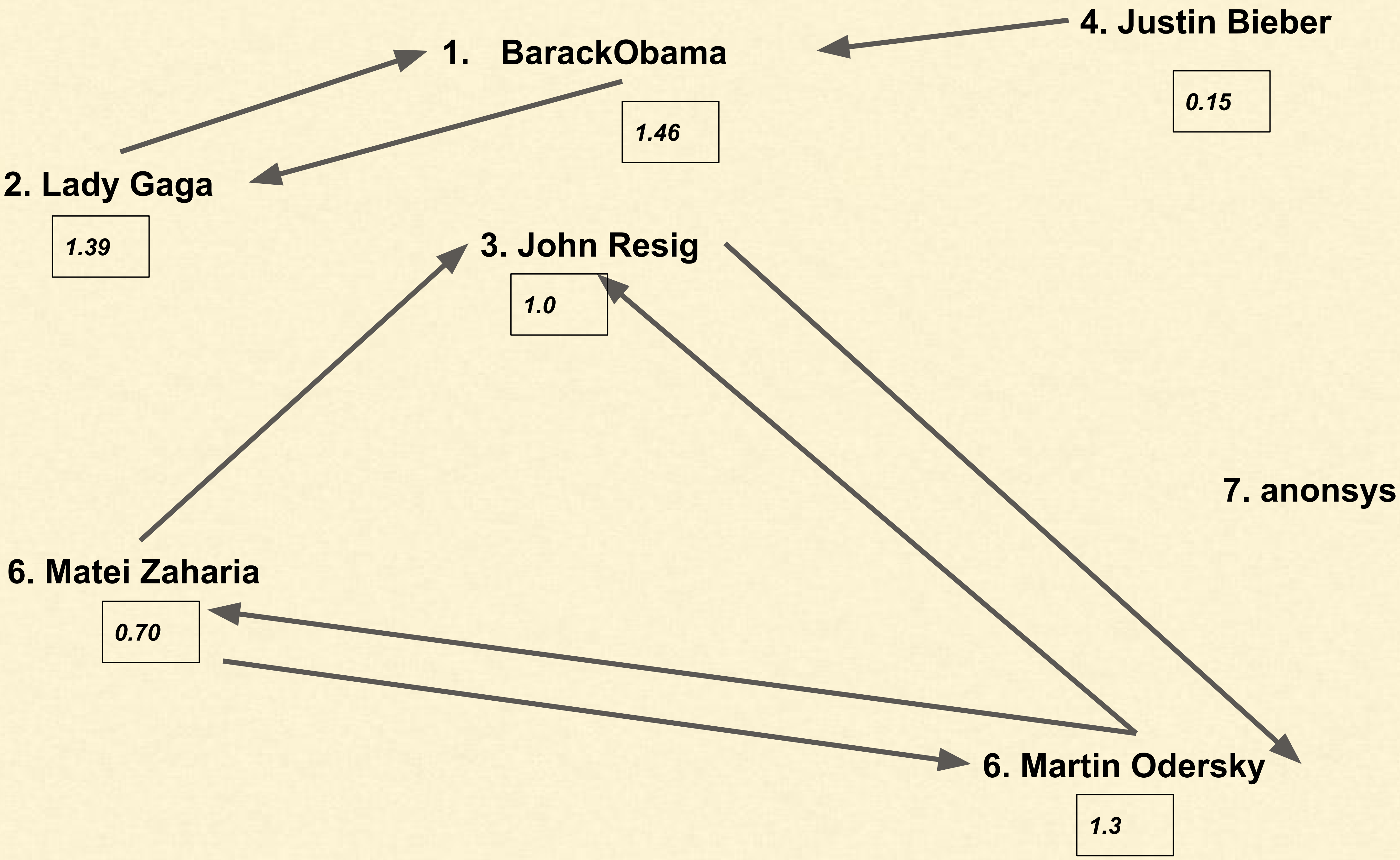
GraphX - Pagerank

```
import org.apache.spark.graphx.GraphLoader

// Load the edges as a graph
val graph = GraphLoader.edgeListFile(sc, "/data/spark/graphx/followers.txt")
// Run PageRank
val ranks = graph.pageRank(0.0001).vertices
// Join the ranks with the usernames
val users = sc.textFile("/data/spark/graphx/users.txt").map { line =>
  val fields = line.split(",")
  (fields(0).toLong, fields(1))
}
val ranksByUsername = users.join(ranks).map {
  case (id, (username, rank)) => (username, rank)
}
// Print the result
println(ranksByUsername.collect().mkString("\n"))
```

See more

GraphX - Pagerank



GraphX

Thank you!

reachus@cloudxlab.com