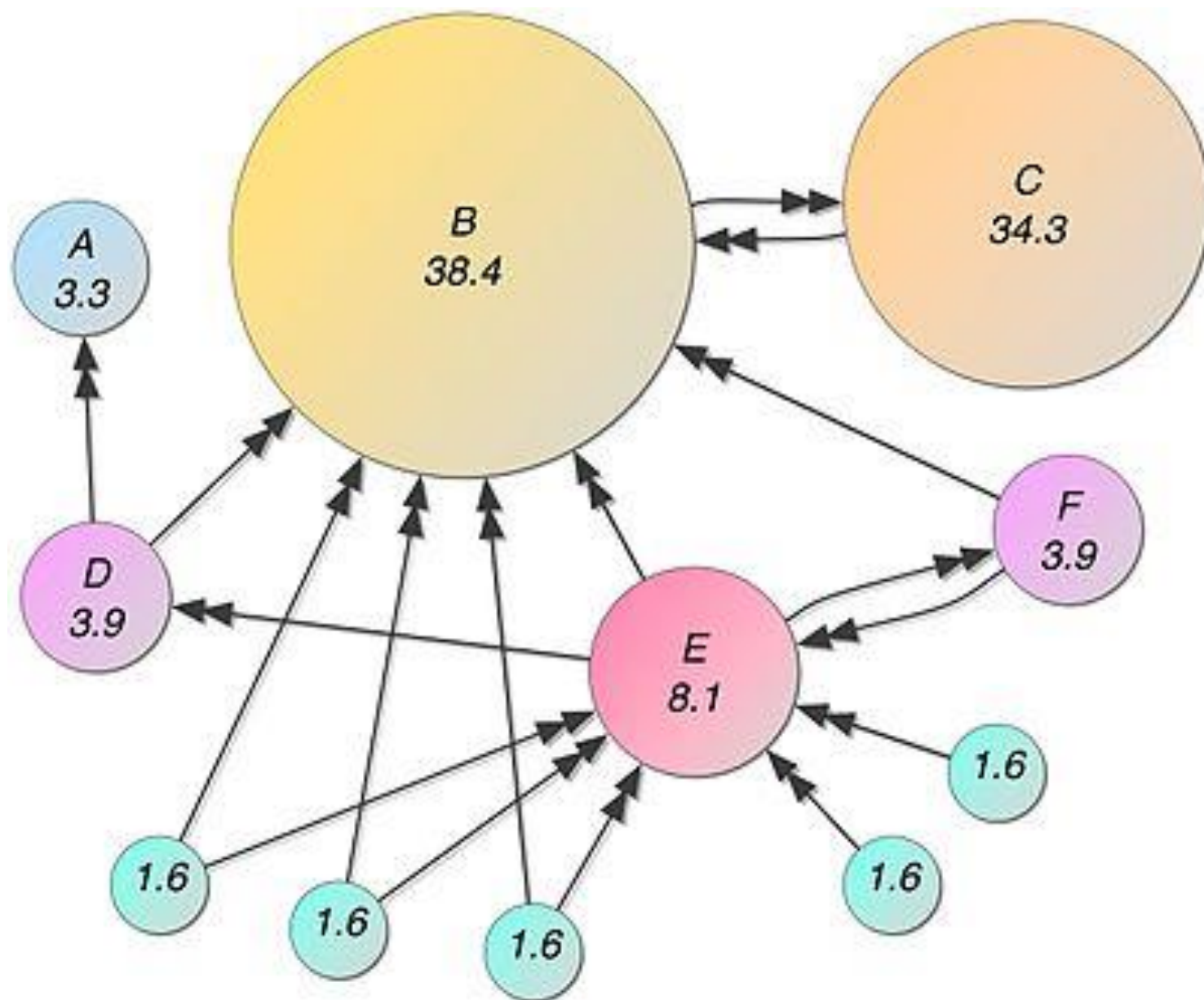


Practice 8

PageRank

Problem

- Use **PageRank** in **graphx**
 - Use predefined function in `org.apache.spark.graphx.GraphLoader`



Problem

➤ Use PageRank in graphx

- How to run a *scala* code

1. spark-shell (in terminal)

```
[cloudera@quickstart Desktop]$ spark-shell
```

2. :load ScriptName.scala (in spark-shell)

```
scala> :load pagerank.scala
```

Dataset for PageRank

➤ Citation Network

- Node = paper, Edge = citation (directed)

➤ Edge list

Start destination
1 3
2 5
6 3
...
...
n 623

❖ 'A->B' means 'A has been cited by B'

* UCI Machine Learning Repository :

<https://linqs.soe.ucsc.edu/data>

➤ You can download the pre-processed dataset on i-campus

Practice 8

1. Use predefined classes in *org.apache.spark.graphx : GraphLoader ()*
2. Perform the PageRank and report the ID and PageRank score of the top 10 pages.