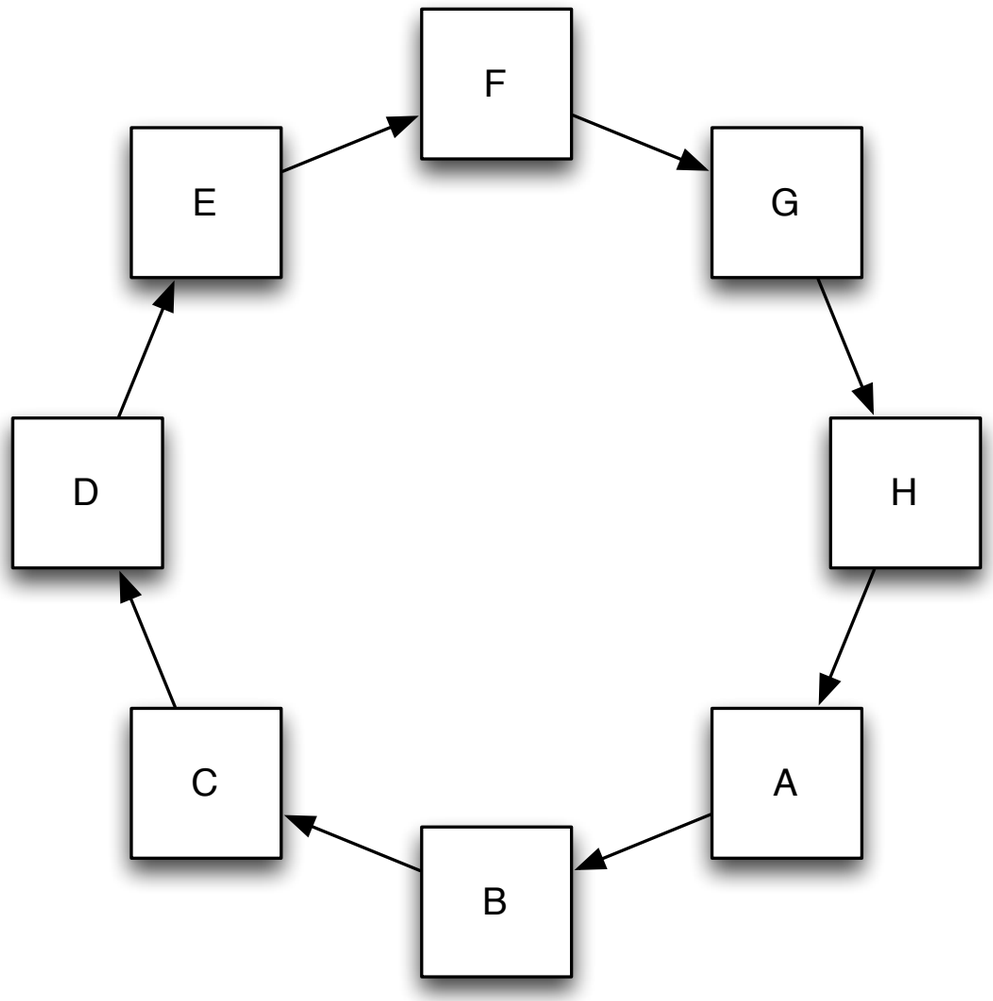
Link Analysis

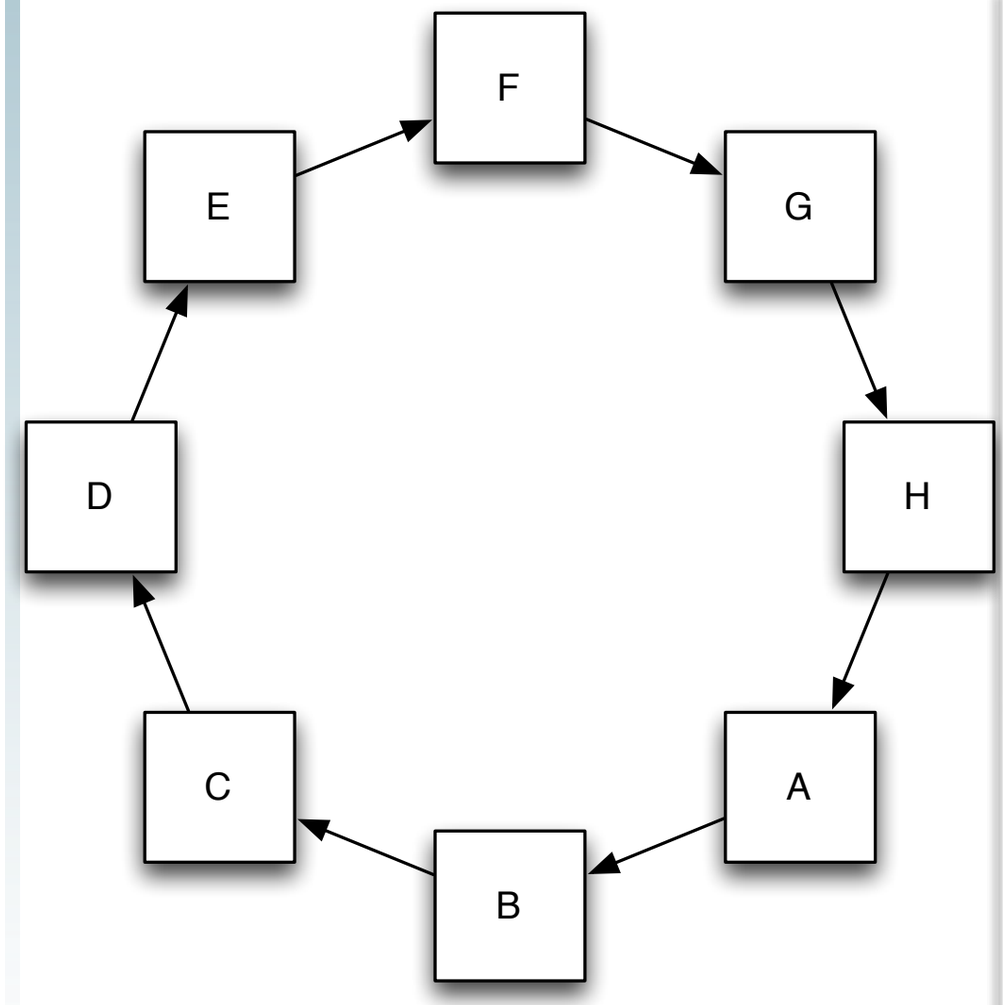
Introduction to Information Retrieval
INF 141/ CS 121
Donald J. Patterson

Content adapted from Hinrich Schütze
http://www.informationretrieval.org

Calculate the Page Rank of this graph with 70% chance of teleporting

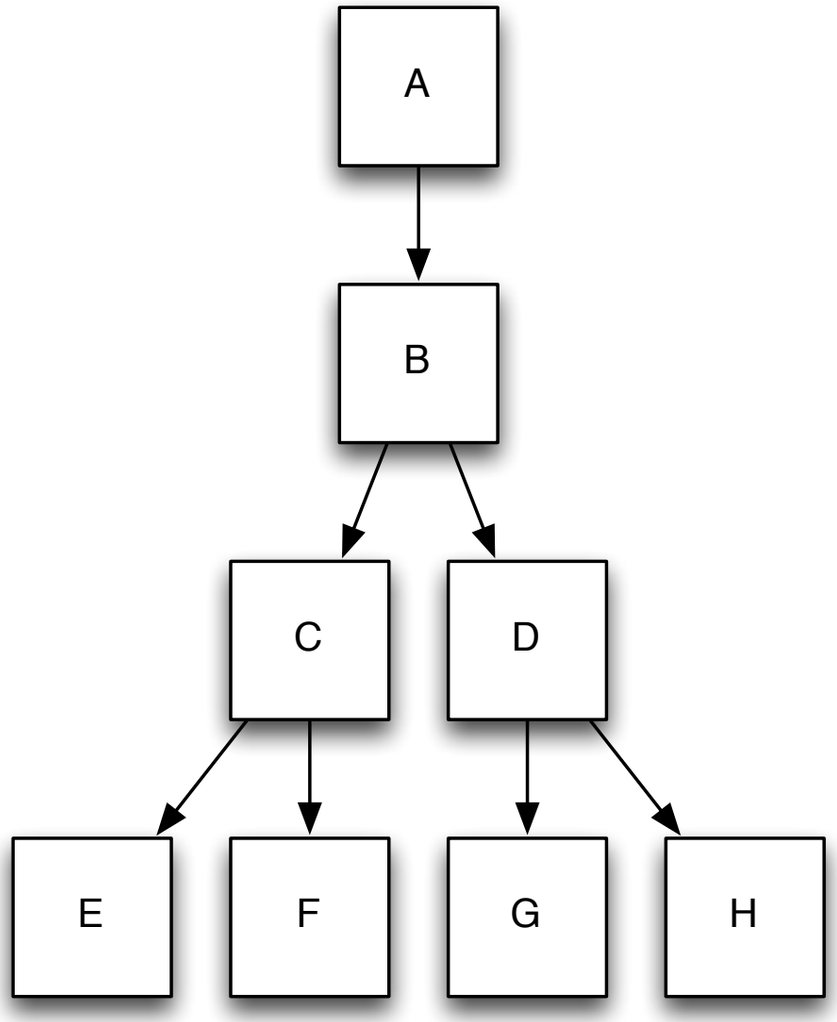


Calculate the Page Rank of this graph with 70% chance of teleporting



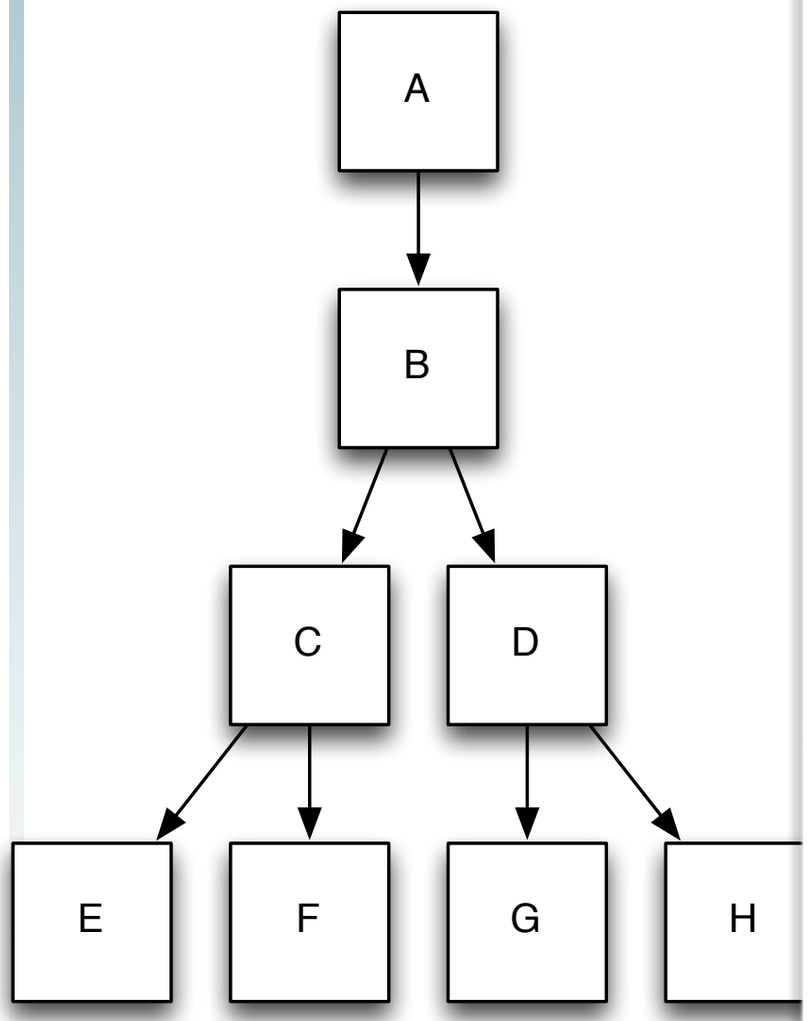

```
>> x

x =

     1     0     0     0     0     0     0     0

>> P

P =

    0.1000    0.3000    0.1000    0.1000    0.1000    0.1000    0.1000    0.1000
    0.1000    0.1000    0.3000    0.1000    0.1000    0.1000    0.1000    0.1000
    0.1000    0.1000    0.1000    0.3000    0.1000    0.1000    0.1000    0.1000
    0.1000    0.1000    0.1000    0.1000    0.3000    0.1000    0.1000    0.1000
    0.1000    0.1000    0.1000    0.1000    0.1000    0.3000    0.1000    0.1000
    0.1000    0.1000    0.1000    0.1000    0.1000    0.1000    0.3000    0.1000
    0.1000    0.1000    0.1000    0.1000    0.1000    0.1000    0.1000    0.3000
    0.3000    0.1000    0.1000    0.1000    0.1000    0.1000    0.1000    0.1000

>> x*P

ans =

    0.1000    0.3000    0.1000    0.1000    0.1000    0.1000    0.1000    0.1000

>> x*P*P*P*P*P*P*P*P*P*P*P

ans =

    0.1250    0.1250    0.1250    0.1250    0.1250    0.1250    0.1250    0.1250
```

Calculate the Page Rank of this graph with no teleporting, just deadend handling



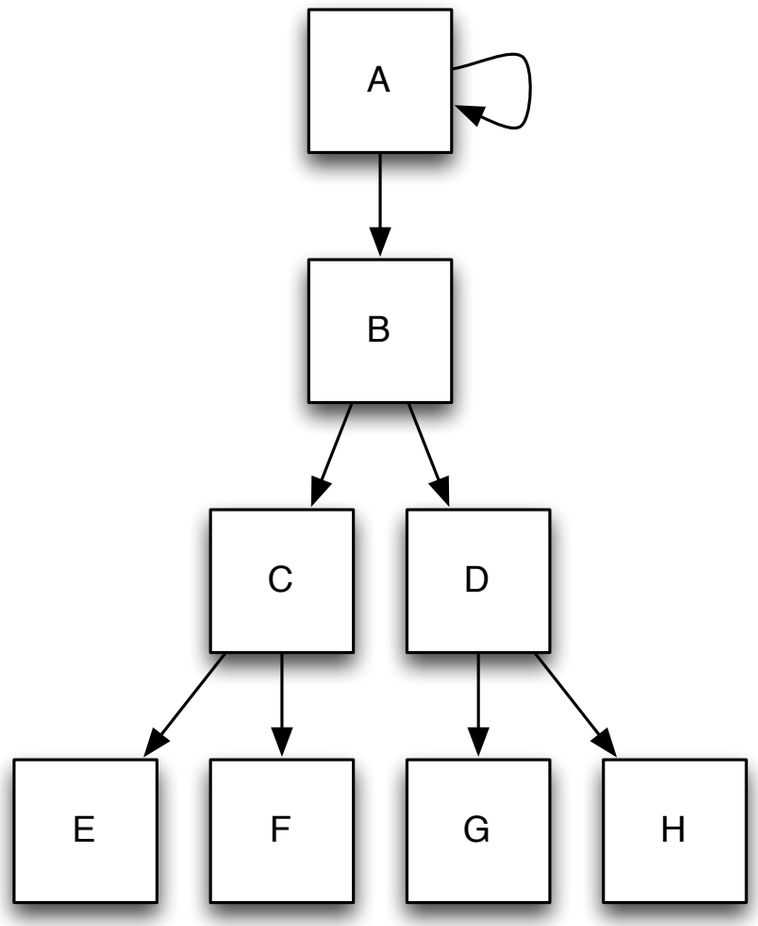
Link Analysis - Exercises

Calculate the P
teleporting, jus



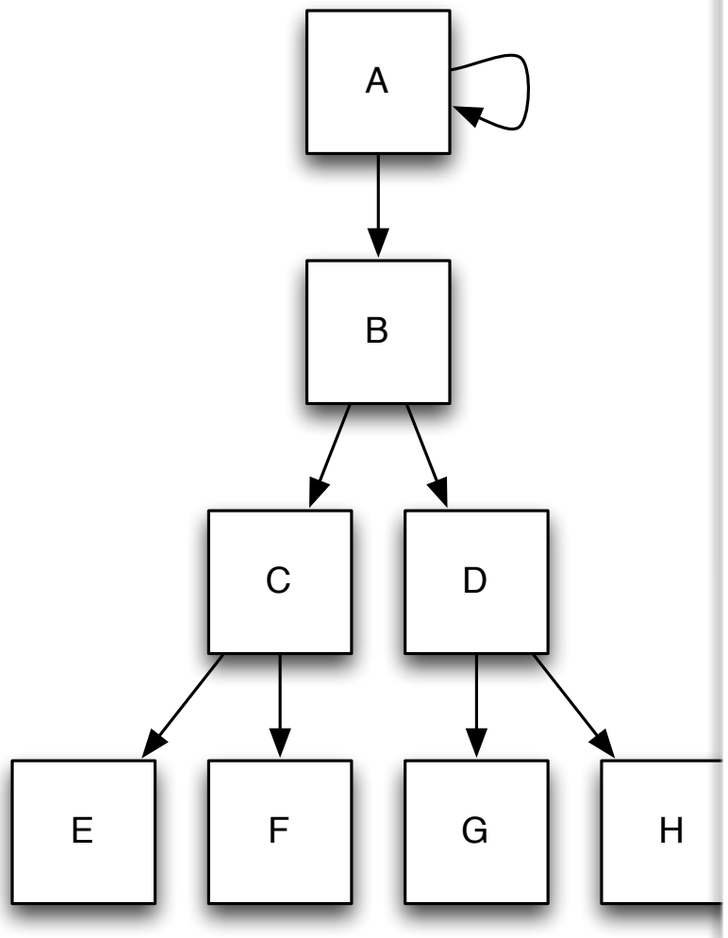

```
>> x

x =

     1     0     0     0     0     0     0     0

>> P

P =

         0    1.0000         0         0         0         0         0         0
         0         0    0.5000    0.5000         0         0         0         0
         0         0         0         0    0.5000    0.5000         0         0
         0         0         0         0         0         0    0.5000    0.5000
    0.1250    0.1250    0.1250    0.1250    0.1250    0.1250    0.1250    0.1250
    0.1250    0.1250    0.1250    0.1250    0.1250    0.1250    0.1250    0.1250
    0.1250    0.1250    0.1250    0.1250    0.1250    0.1250    0.1250    0.1250
    0.1250    0.1250    0.1250    0.1250    0.1250    0.1250    0.1250    0.1250

>> x*P

ans =

     0     1     0     0     0     0     0     0

>> x*P*P

ans =

         0         0    0.5000    0.5000         0         0         0         0

>> x*P*P*P

ans =

         0         0         0         0    0.2500    0.2500    0.2500    0.2500

>> x*P*P*P*P*P*P*P*P*P*P*P*P*P*P*P*P*P*P*P*P*P*P*P*P*P*P*P*P*P*P*P*P*P*P*P*P*P

ans =

    0.0667    0.1333    0.1333    0.1333    0.1333    0.1333    0.1333    0.1333
```

Calculate the Page Rank of this graph with no teleporting, just deadened handling



Link Analysis - Exercises

Calculate the I
teleporting, ju

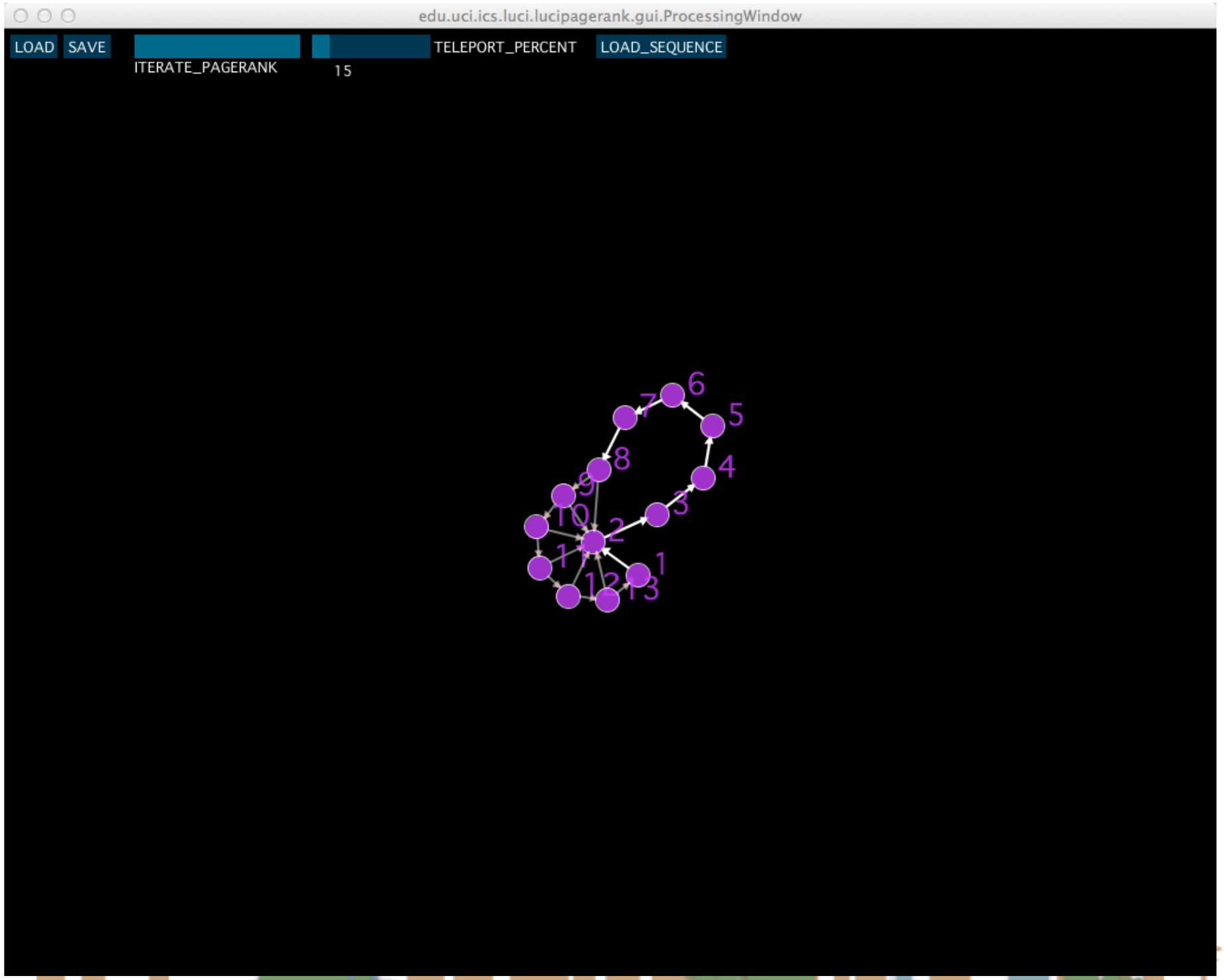



```
>> x

x =

     1     0     0     0     0     0     0     0

>> P

P =

    0.5000    0.5000         0         0         0         0         0         0
         0         0    0.5000    0.5000         0         0         0         0
         0         0         0         0    0.5000    0.5000         0         0
         0         0         0         0         0         0    0.5000    0.5000
    0.1250    0.1250    0.1250    0.1250    0.1250    0.1250    0.1250    0.1250
    0.1250    0.1250    0.1250    0.1250    0.1250    0.1250    0.1250    0.1250
    0.1250    0.1250    0.1250    0.1250    0.1250    0.1250    0.1250    0.1250
    0.1250    0.1250    0.1250    0.1250    0.1250    0.1250    0.1250    0.1250

>> x*P

ans =

    0.5000    0.5000         0         0         0         0         0         0

>> x*P*P

ans =

    0.2500    0.2500    0.2500    0.2500         0         0         0         0

>> x*P*P*P*P*P*P*P*P*P*P*P*P*P*P*P*P*P*P*P*P*P*P*P*P*P*P*P*P*P*P*P

ans =

    0.1250    0.1250    0.1250    0.1250    0.1250    0.1250    0.1250    0.1250
```

Draw a graph with 10 nodes

1) such that 1 node clearly has the highest PageRank

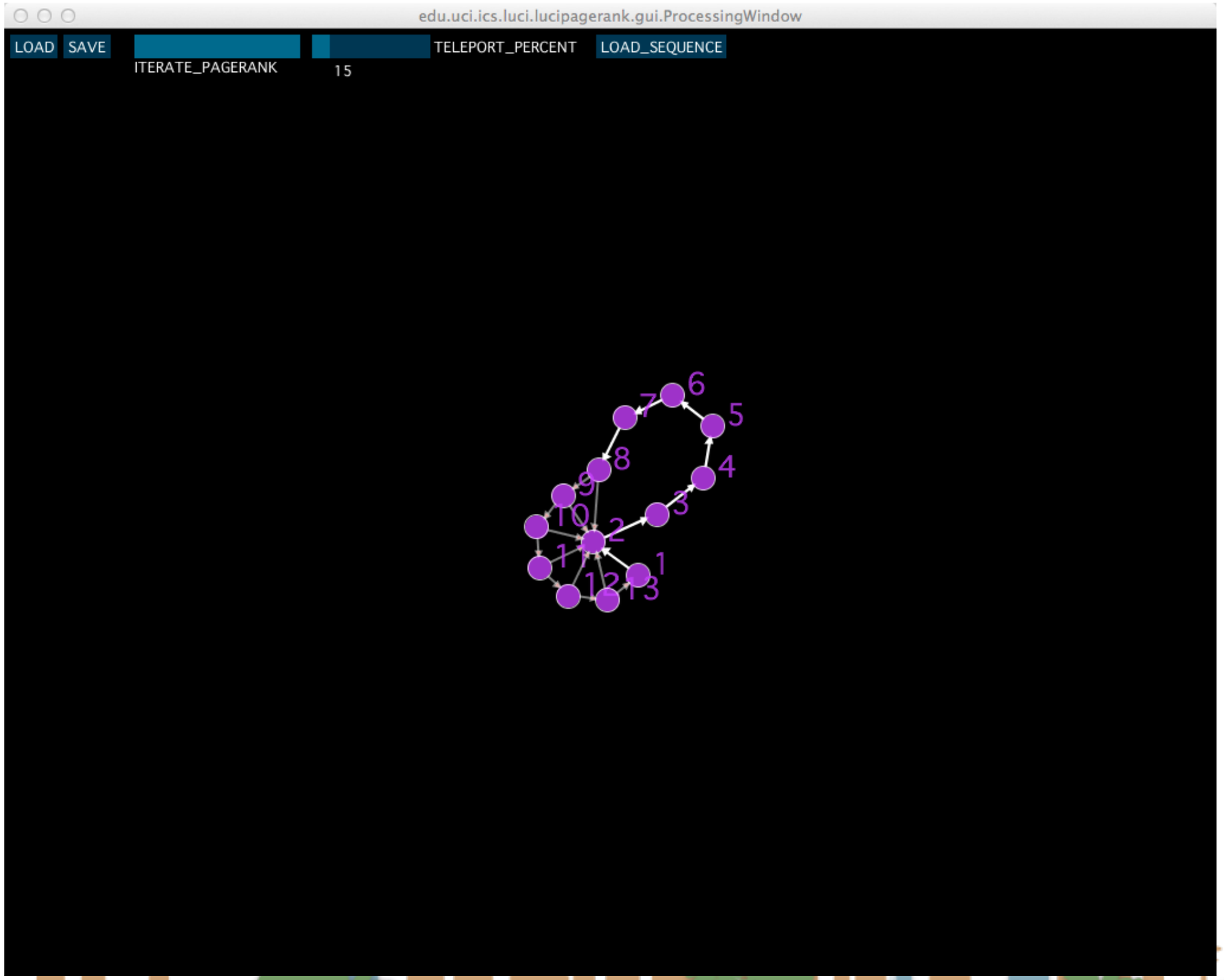
Link Analysis - Exercises



Draw a graph with 10 nodes

2) such that 4 nodes have very high and equal PageRank

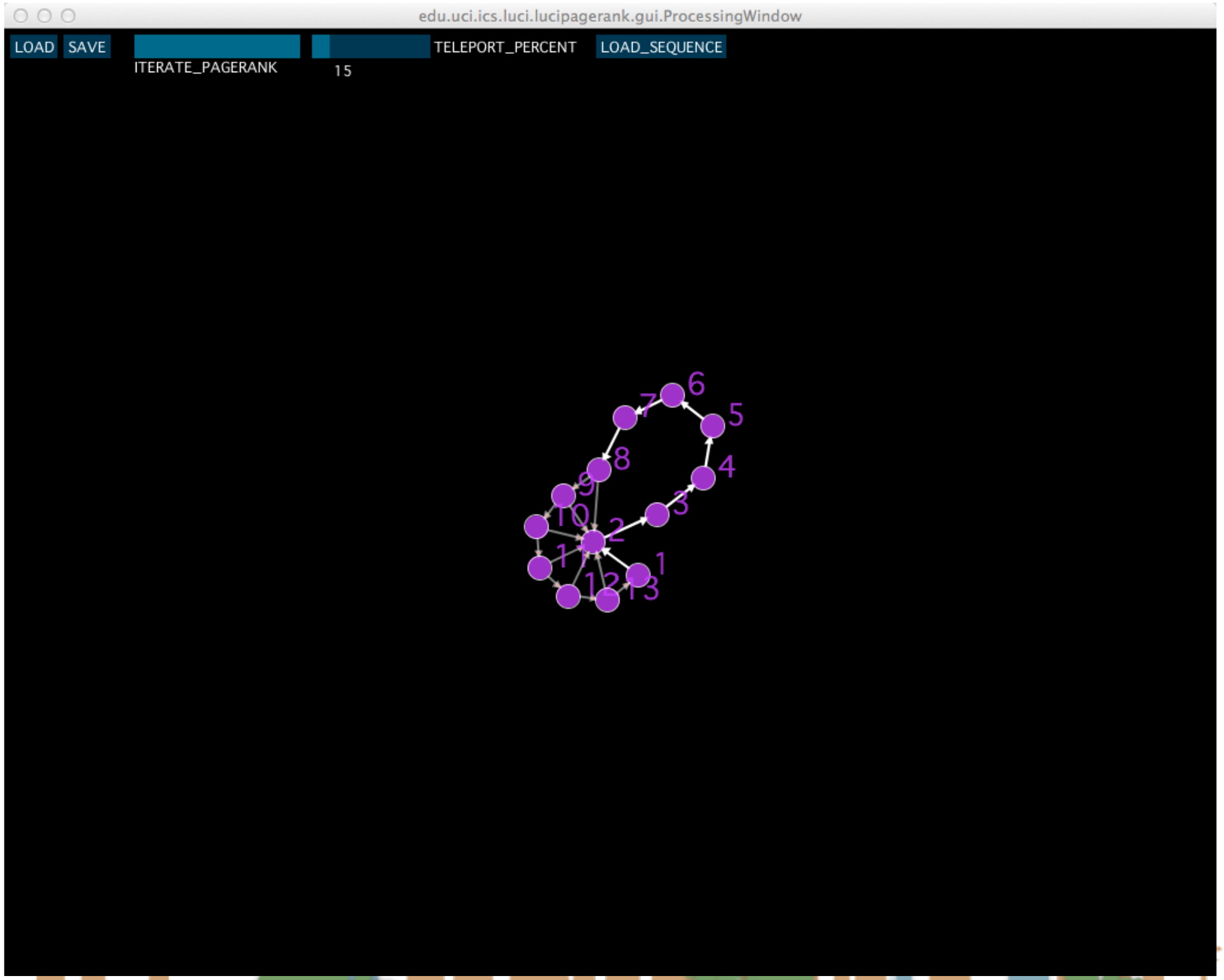
Link Analysis - Exercises



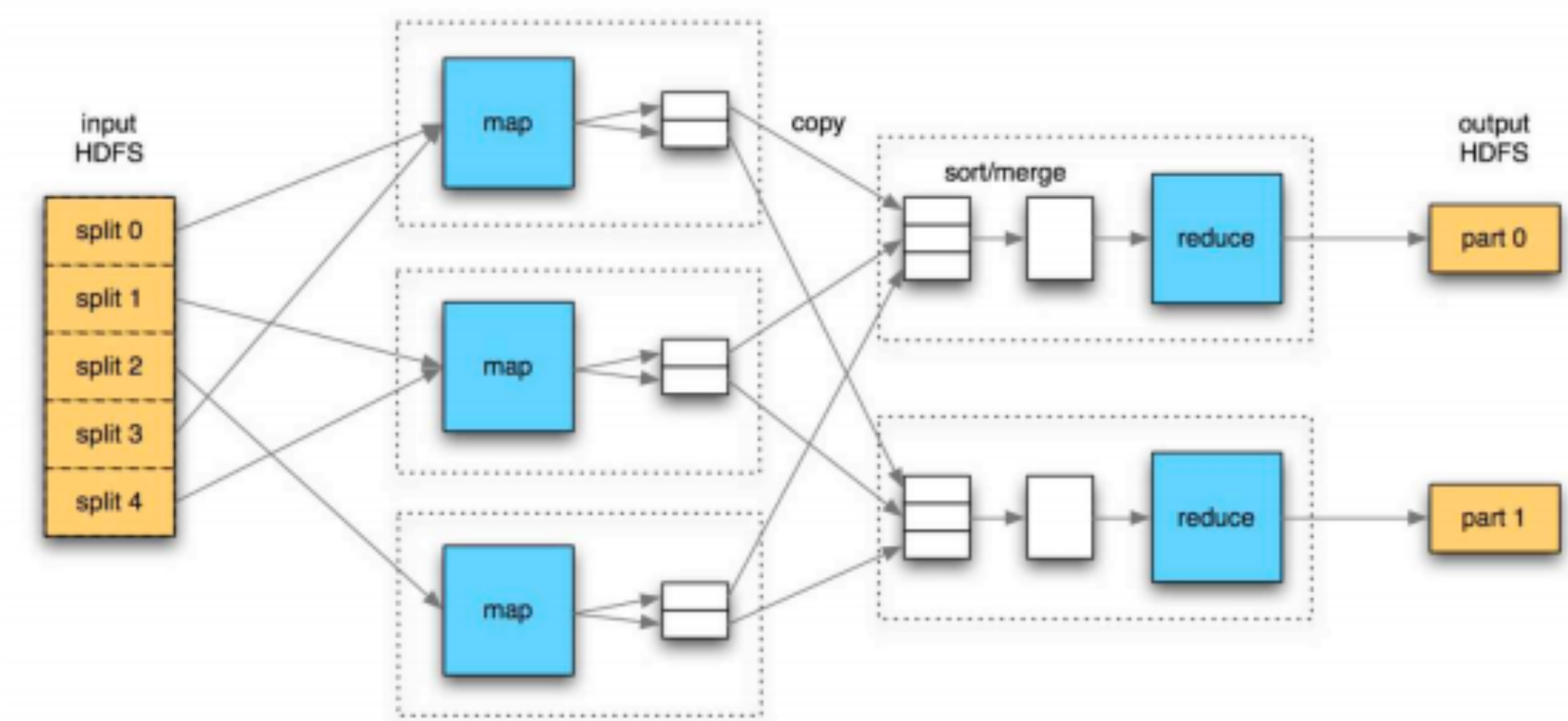
Draw a graph with 10 nodes

3) such that no node has the same PageRank

Link Analysis - Exercises



How could PageRank be calculated in Hadoop?



Link Analysis - Exercises

input: node_a:[P(node_a), [node_b,node_c]]

map out: [node_b, P(node_a)/2]
[node_c, P(node_a)/2]
[node_a,[node_b,node_c]]

reduce in:
node_x: [P(in1),…,P(in3)….[node_y,node_z]]

reduce out:
node_x: [sum(P(in1)…P(in3)),[node_y,node_z]]