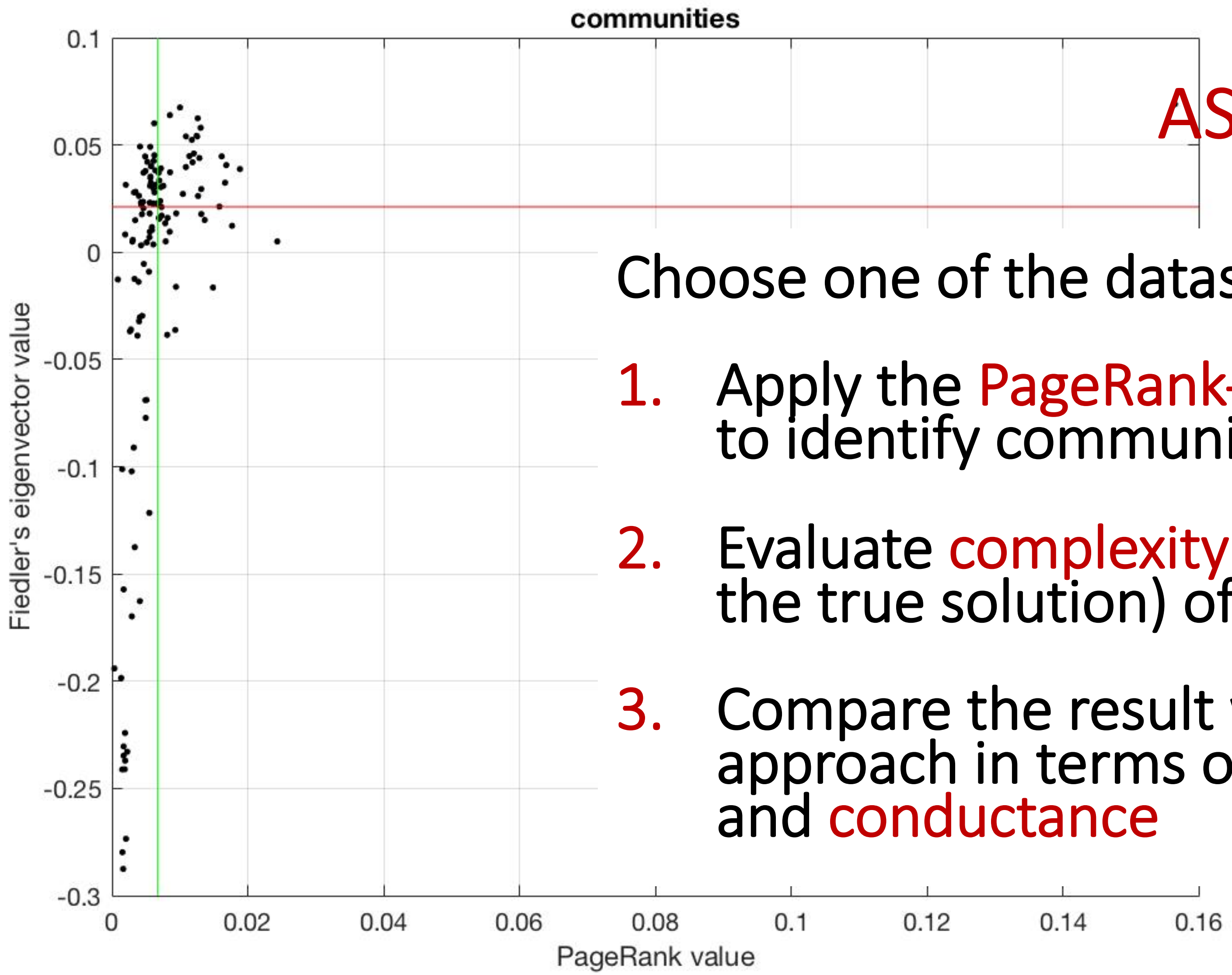
Network Science

Lab #6
Community detection
PageRank-Nibble approach

Timetable

- Lab 1 – Fri Oct 12
  Scale free properties
- Lab 2 – Fri Oct 19
  Albert-Baràbasi model
- Lab 3 – Fri Oct 26
  Assortativity
- Lab 4 – Fri Nov 16
  Ranking
- Lab 5 – Fri Nov 23
  Community detection – Spectral
- Lab 6 – Fri Nov 30
  Community detection – PageRank-Nibble
- ~~Lab 7 – Fri Dec 7~~
  ~~Gephi~~

Lab 6 – Community detection



ASSIGNMENT

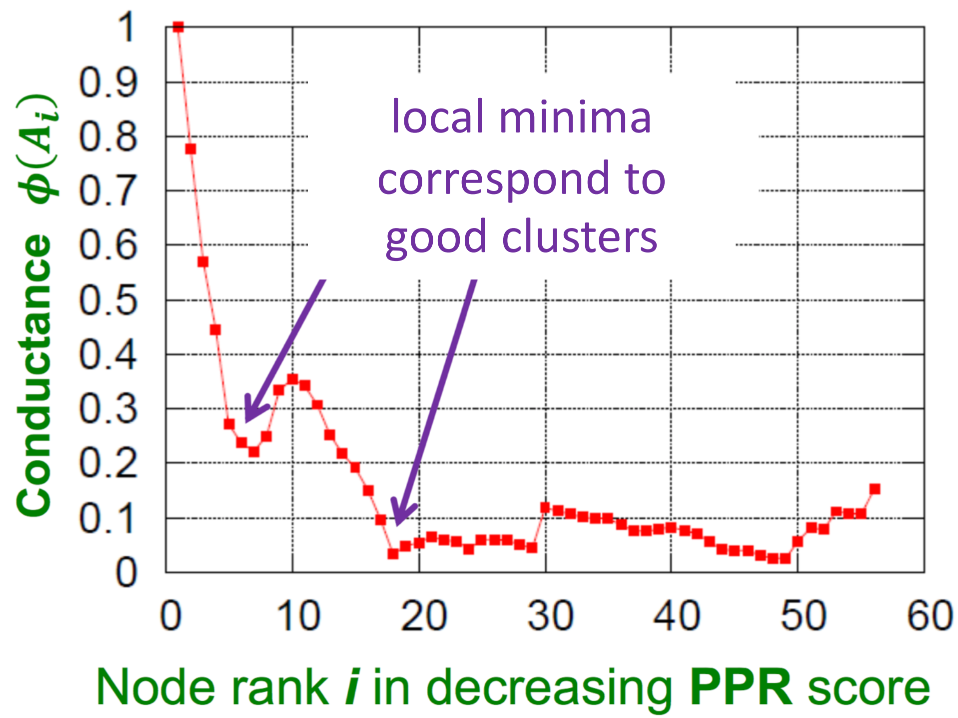
Choose one of the datasets, then :

1. Apply the PageRank-nibble algorithm to identify communities
2. Evaluate complexity and precision (vs. the true solution) of the result
3. Compare the result with the spectral approach in terms of communities and conductance

The PageRank-Nibble approach

Algorithm

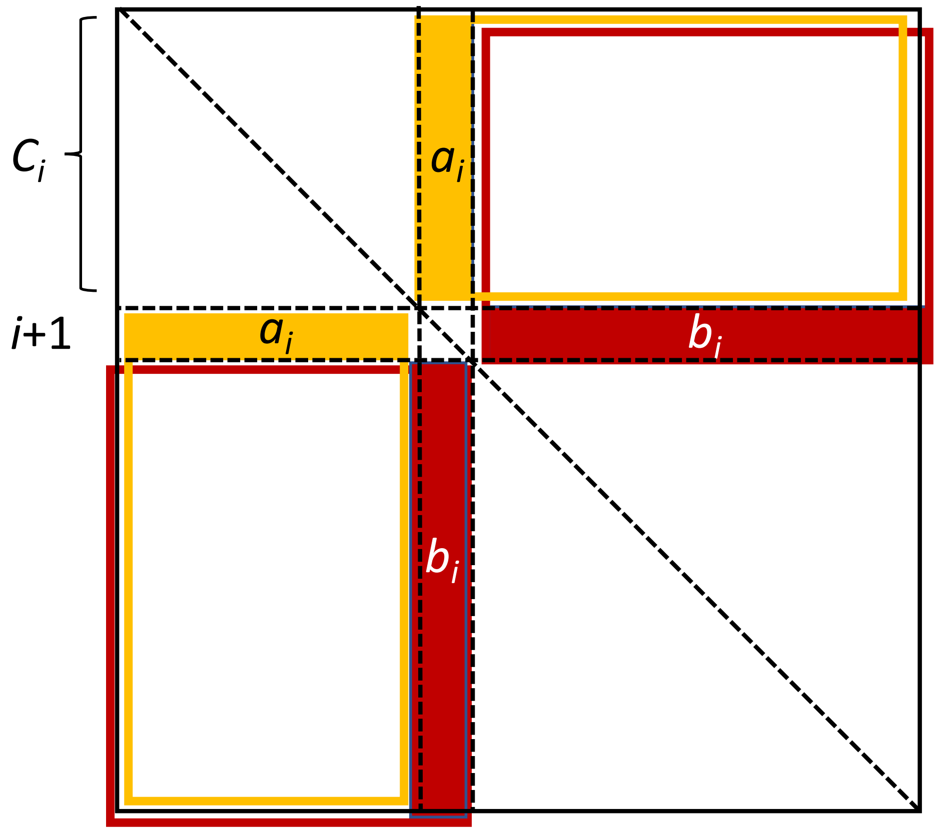
- Run approximate PageRank with teleport set $\{i\}$
- Order nodes by ranking value (in decreasing order)
- Sweep over the nodes to find a good cluster



Goodness of fit = Conductance

✓ $\phi(S) = \mathrm{cut}(S, S^c) / \min(\mathrm{assoc}(S), \mathrm{assoc}(S^c))$

Computing the sweep



Algorithm

- Let $C_i = \{1,2,\ldots,i\}$
- Node degree is $d_i = a_i + b_i$
- Association update

  $\text{assoc}(C_{i+1}) = \text{assoc}(C_i) + d_i$
- Cut update

  $\text{cut}(C_{i+1}) = \text{cut}(C_i) - a_i + b_i$

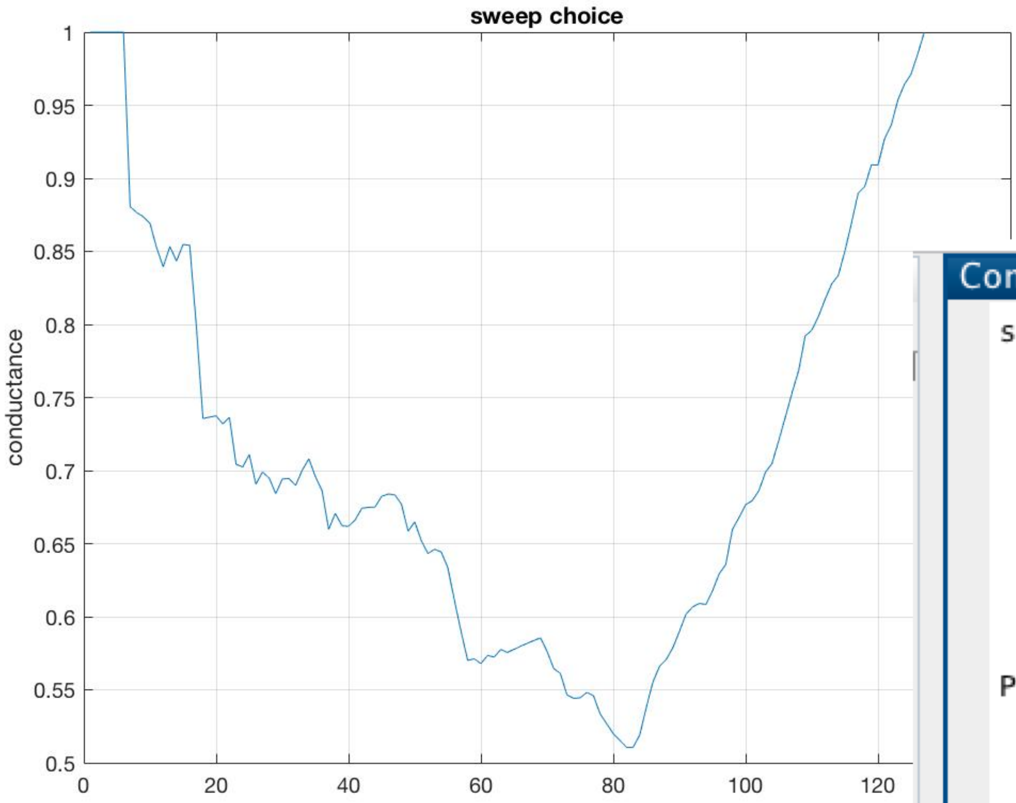
Approximate SimRank

- Set $\boldsymbol{u} = \boldsymbol{0}$, $\boldsymbol{v} = \boldsymbol{q}$
- Choose a starting index i
- If $v_i > \varepsilon\, d_i / D$ apply the Push operation
  - define vector $\boldsymbol{\delta}$ with $\delta_i = v_i$ and all the rest to 0
  - $\boldsymbol{u}^+ = \boldsymbol{u} + (1-c)\,\boldsymbol{\delta}$
  - $\boldsymbol{v}^+ = \boldsymbol{v} - \boldsymbol{\delta} + c\,\boldsymbol{M}\,\boldsymbol{\delta}$
- Otherwise step index i by 1
- Exit if none of the entries of $\boldsymbol{v}$ is available for Push

True SimRank

- Solve
  - $\boldsymbol{u} = c\,\boldsymbol{M}\,\boldsymbol{u} + (1\text{-}c)\,\boldsymbol{q}$, or
  - $(\boldsymbol{I} - c\,\boldsymbol{M})\,\boldsymbol{u} = (1\text{-}c)\,\boldsymbol{q}$

Lab 6 – Community detection



Command Window

```
spectral approach
   Minimum conductance: 0.40834
   Cheeger's upper bound: 0.93678
   # of links: 2075
   Cut value: 842
   Assoc value: 2062
   Community size #1: 68
   Community size #2: 60

PageRank-nibble approach
   complexity/D: 23.7289
   epsilon: 0.001
   prec: 0.00046518
   Minimum conductance: 0.51059
   # of links: 2075
   Cut value: 1037
   Assoc value: 2031
   Community size #1: 82
   Community size #2: 46
fx >>
```

Lab 6 – MatLab hints

1. spdiags(ones(N,1),0,N,N) : sparse identity matrix

2. eigs: extracts (ordered) eigenvalues

3. A\b: solves the linear system Ax = b

4. cumsum: evaluates a cumulative sum
5. triu: extracts the upper triangular matrix
6. tril: extracts the lower triangular matrix