Università della Svizzera italiana

Institute of Computational Science ICS

Numerical Computing **2020**

Student: Claudio Maggioni Discussed with: FULL NAME

Due date: Wednesday, 14 October 2020, 11:55 PM

Solution for Project 2

Submission instructions

(Please, notice that following instructions are mandatory: submissions that don't comply with, won't be considered)

- Assignments must be submitted to Moodle (i.e. in electronic format).
- Provide both executable package (single .class or .jar file) and sources (.java files). If you are using non-sdk libraries, please add them in the file. Sources must be organized in packages called: *ch.usi.inf.ncc12.assignment<assignmentNumber>.exercise<exerciseNumber>.<name>.<surname>* and the jar file must be called:

  assignment<AssignmentNumber>.<Name>.<Surname>.jar

  Projects exported directly from Eclipse would be much appreciated (Please, be sure that you are including also the sources in the jar file).
- The produced files (one pdf and one jar file) must be collected into a single archive file (.zip) named:

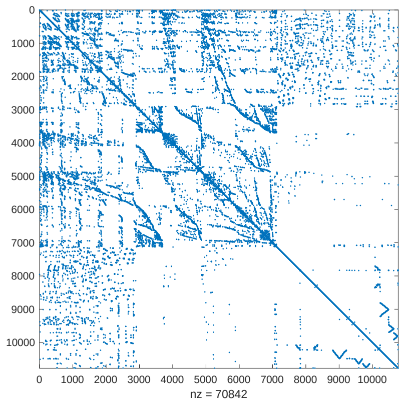
  assignment<AssignmentNumber>.<Name>.<Surname>.zip

The purpose of this assignment[1] is to learn the importance of sparse linear algebra algorithms to solve fundamental questions in social network analyses. We will use the coauthor graph from the Householder Meeting and the social network of friendships from Zachary's karate club [1]. These two graphs are one of the first examples where matrix methods were used in computational social network analyses.

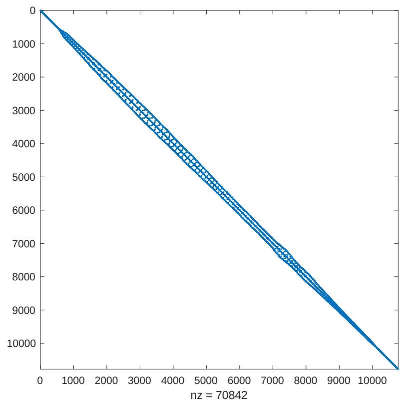
[1] This document is originally based on a blog from Cleve Moler, who wrote a fantastic blog post about the Lake Arrowhead graph, and John Gilbert, who initially created the coauthor graph from the 1993 Householder Meeting. You can find more information at http://blogs.mathworks.com/cleve/2013/06/10/lake-arrowhead-coauthor-graph/. Most of this assignment is derived from this archived work.

1. The Reverse Cuthill McKee Ordering [10 points]

The Reverse Cuthill McKee Ordering of matrix `A_SymPosDef` is computed with MATLAB's `sysrcm(...)` and the matrix is rearranged accordingly. Here are the spy plot of these matrices:



(a) Spy plot of `A_SymPosDef`



(b) Spy plot of `sysrcm(...)` rearranged version of `A_SymPosDef`

Figure 1. Spy plots of the two matrices

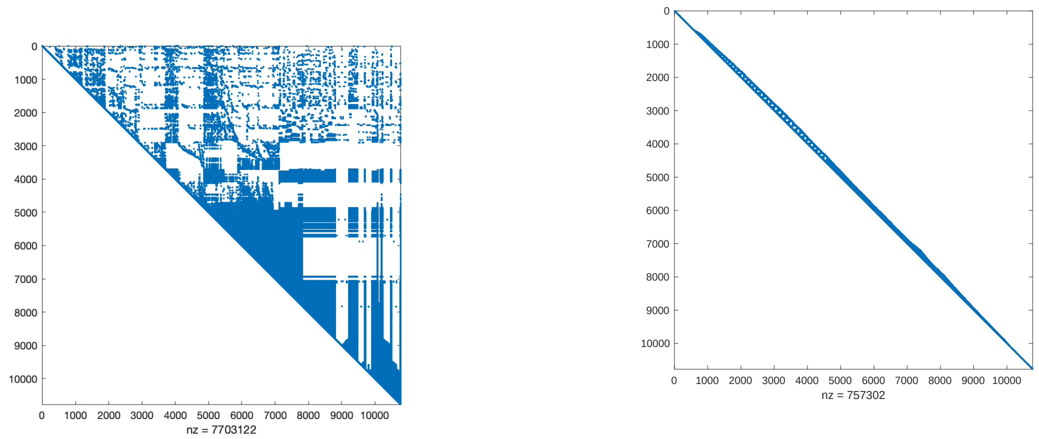
And the spy plots of the corresponding Cholesky factor are listed in figure 2.

The number of nonzero elements in the Cholesky factor of the RCM optimized matrix are significantly lower (circa 0.1x) of the ones in the vanilla process. The respective nonzero counts can be found in figure 2.

2. Sparse Matrix Factorization [10 points]

2.1. Show that $A \in R^{nxn}$ has exactly $5n - 6$ nonzero elements.

The given description of A says that all the element at the edges of the matrix (rows and columns 1 and n) plus all the elements on the main diagonal are the only nonzero elements of A. Therefore, this cells can be counted as the 4 vertex cells in the matrix square plus 5 $n-2$-long segments, corresponding



(a) Spy plot of `chol(A_SymPosDef)`

(b) Spy plot of `chol(A_SymPosDef(sysrcm(A_SymPosDef), sysrcm(A_SymPosDef)))`

Figure 2. Spy plots of the two Cholesky factors

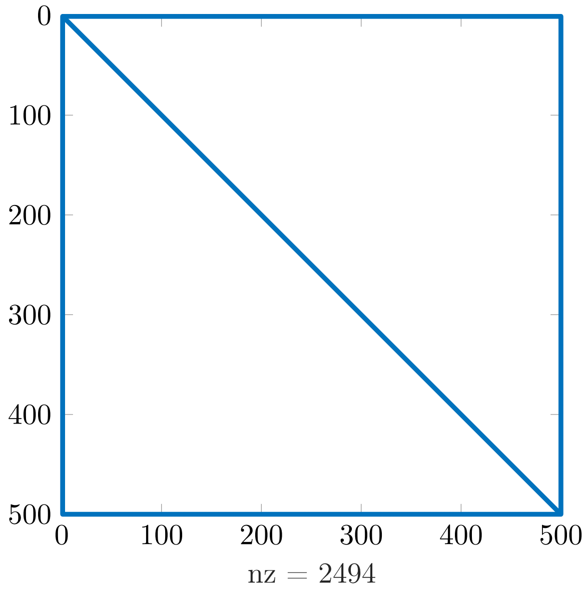
to all edges and the main diagonal. Therefore:

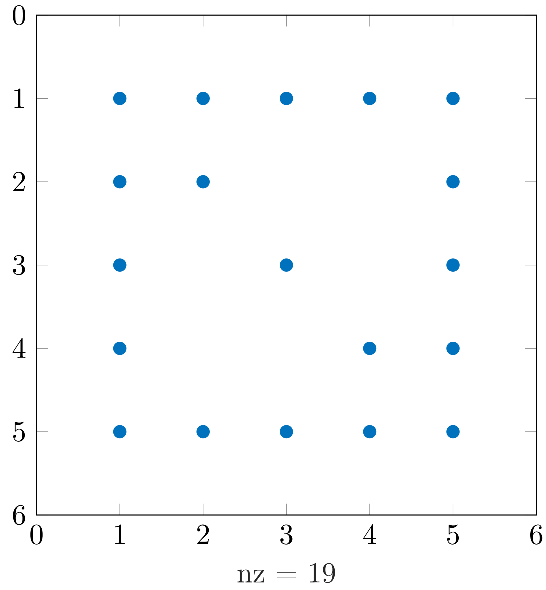
$$4 + 5\dot{(}n - 2) = 5n - 6$$

2.2. Write a short Matlab script to construct this matrix and visualize its non-zero structure(you can use, e.g., the command `spy()`).

The MATLAB script can be found in file `ex3.m`.

Here is a spy plot of the nonzero values of A, for $n = 5$:



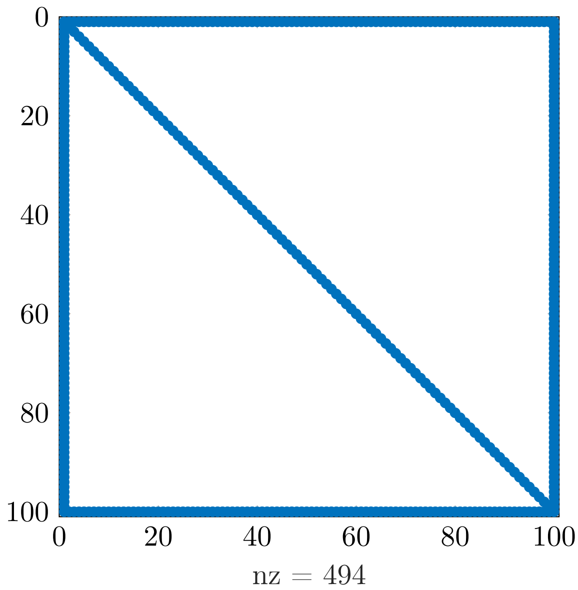


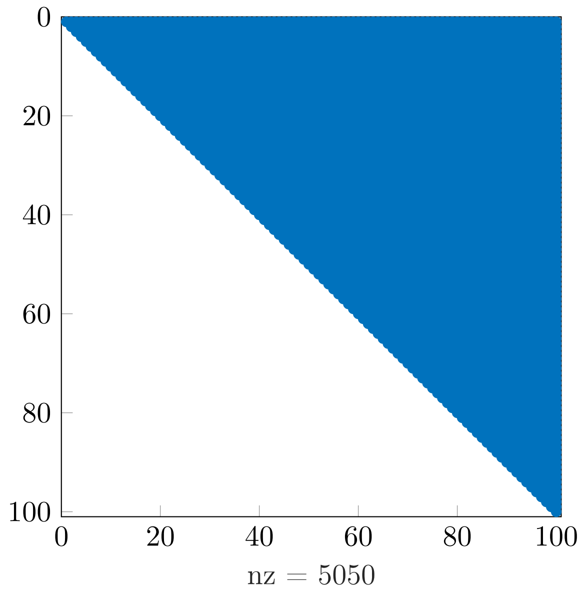
The matrix $A \in R^{nxn}$ looks like this (zero entries are represented as blanks):

$$A := \begin{bmatrix} n & 1 & 1 & \dots & 1 \\ 1 & n+1 & & & 1 \\ 1 & & n+2 & & 1 \\ \vdots & & & \ddots & \vdots \\ 1 & 1 & 1 & \dots & 2n-1 \end{bmatrix}$$

2.3. **Using again the `spy()` command, visualize side by side the original matrix A and the result of the Cholesky factorization (`chol()` in Matlab). Then explain why for n = 100000 using Matlab's `chol(...)` to solve $Ax = b$ for a given righthand-side vector would be problematic.**

Here is the plot of `spy(A)` (on the left) and `chol(spy(A))` (on the right) for $n = 100$.





Solving $Ax = b$ would be a costly operation since the a Cholesky decomposition of matrix A (performed using MATLAB's `chol(...)`) would drastically reduce the number of zero elements in the matrix in the very first iteration. This is due to the fact that the first row, by definition, is made of of only nonzero elements (namely 1s) and by subtracting the first row to every other row (as what would effectively happen in the first iteration of the Cholesky decomposition of A) the zero elements would become (negative) nonzero elements, thus making all columns but the first almost empty of 0s.

3. Degree Centrality [10 points]

Assuming that the degree of the Householder graph is the number of co-authors of each author and that an author is not co-author of himself, the degree centralities of all authors sorted in descending order are below.

This output has been obtained by running `ex3.m`.

```
Author Centrality: Coauthors...

Golub 31: Wilkinson TChan Varah Overton Ernst VanLoan Saunders Bojanczyk
          Dubrulle George Nachtigal Kahan Varga Kagstrom Widlund
          OLeary Bjorck Eisenstat Zha VanDooren Tang Reichel Luk Fischer
          Gutknecht Heath Plemmons Berry Sameh Meyer Gill
Demmel 15: Edelman VanLoan Bai Schreiber Kahan Kagstrom Barlow
           NHigham Arioli Duff Hammarling Bunch Heath Greenbaum Gragg
Plemmons 13: Golub Nagy Harrod Pan Funderlic Bojanczyk George Barlow Heath
             Berry Sameh Meyer Nichols
Heath 12: Golub TChan Funderlic George Gilbert Eisenstat Ng Liu Laub Plemmons
```

Paige Demmel

Schreiber 12: TChan VanLoan Moler Gilbert Pothen NTrefethen Bjorstad NHigham Eisenstat Tang Elden Demmel

Hammarling 10: Wilkinson Kaufman Bai Bjorck VanHuffel VanDooren Duff Greenbaum Gill Demmel

VanDooren 10: Golub Boley Bojanczyk Kagstrom VanHuffel Luk Hammarling Laub Nichols Paige

TChan 10: Golub Saied Ong Kuo Tong Schreiber Arioli Duff Heath Hansen

Gragg 9: Borges Kaufman Harrod Reichel Stewart BunseGerstner Ammar Warner Demmel

Moler 8: Wilkinson VanLoan Gilbert Schreiber Henrici Stewart Bunch Laub

VanLoan 8: Golub Moler Schreiber Kagstrom Luk Bunch Paige Demmel

Paige 7: Anjos VanLoan Saunders Bjorck VanDooren Laub Heath

Gutknecht 7: Golub Ashby Boley NTrefethen Nachtigal Varga Hochbruck

Luk 7: Golub Overton Boley VanLoan Bojanczyk Park VanDooren

Eisenstat 7: Golub Gu George Schreiber Liu Heath Ipsen

George 7: Golub Eisenstat Ng Liu Tang Heath Plemmons

Meyer 6: Golub Benzi Funderlic Stewart Ipsen Plemmons

Bunch 6: LeBorne Fierro VanLoan Moler Stewart Demmel

Stewart 6: Moler Bunch Gragg Meyer Gill Mathias

Reichel 6: Golub NTrefethen Nachtigal Fischer Gragg Ammar

Bjorck 6: Golub Park Duff Hammarling Elden Paige

NTrefethen 6: Schreiber Nachtigal Reichel Gutknecht Greenbaum ATrefethen

Nichols 5: Byers Barlow VanDooren Plemmons BunseGerstner

Greenbaum 5: Cullum Strakos NTrefethen Hammarling Demmel

Ipsen 5: Chandrasekaran Barlow Eisenstat Meyer Jessup

Laub 5: Kenney Moler VanDooren Heath Paige

Duff 5: TChan Bjorck Arioli Hammarling Demmel

Liu 5: George Gilbert Eisenstat Ng Heath

Park 5: Boley Bjorck VanHuffel Luk Elden

Zha 5: Golub Bai Barlow VanHuffel Hansen

Widlund 5: Golub Bjorstad OLeary Smith Szyld

Barlow 5: Zha Ipsen Plemmons Nichols Demmel

Kagstrom 5: Golub VanLoan VanDooren Ruhe Demmel

Varga 5: Golub Marek Young Gutknecht Starke

Gilbert 5: Moler Schreiber Ng Liu Heath

Gill 4: Golub Saunders Hammarling Stewart

Sameh 4: Golub Harrod Plemmons Berry

Berry 4: Golub Harrod Plemmons Sameh

BunseGerstner 4: He Byers Gragg Nichols

Hansen 4: TChan Fierro OLeary Zha

Ng 4: George Gilbert Liu Heath
Arioli 4: TChan MuntheKaas Duff Demmel
VanHuffel 4: Zha Park VanDooren Hammarling
Nachtigal 4: Golub NTrefethen Reichel Gutknecht
Bojanczyk 4: Golub VanDooren Luk Plemmons
Harrod 4: Plemmons Gragg Berry Sameh
Boley 4: Park VanDooren Luk Gutknecht
Wilkinson 4: Golub Dubrulle Moler Hammarling
Ammar 3: He Reichel Gragg
Elden 3: Schreiber Bjorck Park
Fischer 3: Golub Modersitzki Reichel
Tang 3: Golub George Schreiber
NHigham 3: Schreiber Pothen Demmel
OLeary 3: Golub Widlund Hansen
Bjorstad 3: Schreiber Widlund Boman
Kahan 3: Golub Davis Demmel
Bai 3: Zha Hammarling Demmel
Saunders 3: Golub Paige Gill
Funderlic 3: Heath Plemmons Meyer
Kaufman 3: Hammarling Gragg Warner
Starke 2: Varga Hochbruck
Hochbruck 2: Gutknecht Starke
Jessup 2: Crevelli Ipsen
Warner 2: Kaufman Gragg
Ruhe 2: Wold Kagstrom
Szyld 2: Marek Widlund
Young 2: Kincaid Varga
Pothen 2: Schreiber NHigham
Tong 2: TChan Kuo
Kuo 2: TChan Tong
Marek 2: Varga Szyld
Dubrulle 2: Golub Wilkinson
Fierro 2: Bunch Hansen
Byers 2: BunseGerstner Nichols
Overton 2: Golub Luk
He 2: BunseGerstner Ammar
Mathias 1: Stewart
Davis 1: Kahan
ATrefethen 1: NTrefethen
Henrici 1: Moler

```
Smith 1: Widlund
MuntheKaas 1: Arioli
Boman 1: Bjorstad
Chandrasekaran 1: Ipsen
Wold 1: Ruhe
Ong 1: TChan
Saied 1: TChan
Strakos 1: Greenbaum
Cullum 1: Greenbaum
Edelman 1: Demmel
Pan 1: Plemmons
Nagy 1: Plemmons
Gu 1: Eisenstat
Benzi 1: Meyer
Anjos 1: Paige
Crevelli 1: Jessup
Kincaid 1: Young
Borges 1: Gragg
Ernst 1: Golub
Modersitzki 1: Fischer
LeBorne 1: Bunch
Ashby 1: Gutknecht
Kenney 1: Laub
Varah 1: Golub
```

4. The Connectivity of the Coauthors [10 points]

The author indexes of the common authors between the author at index i and the author at index j can be computed by listing the indexes of the nonzero elements in the Schur product (or element-wise product) between $A_{:,i}$ and $A_{:,j}$ (respectively the i-th and j-th column vector of A). Therefore the set C of common coauthor's indexes can be defined as:

$$C = \{i \in N_0 \mid (A_{:,i} \odot A_{:,j})_i = 1\}$$

The results below were computing by using the script `ex4.m`.

The common Co-authors between Golub and Moler are Wilkinson and Van Loan.

The common Co-authors between Golub and Saunders are Golub, Saunders and Gill.

The common Co-authors between TChan and Demmel are Schreiber, Arioli, Duff and Heath.

5. PageRank of the Coauthor Graph [10 points]

The PageRank values for all authors were computing by using the scripts `ex5.m` and `pagerank.m`, a basically identical version of `pagerank.m` from Mini Project 1. The output is shown below.

	page-rank	in	out	author
1	0.0511	32	32	Golub
104	0.0261	16	16	Demmel
86	0.0229	14	14	Plemmons
44	0.0212	13	13	Schreiber
3	0.0201	11	11	TChan
81	0.0198	13	13	Heath
90	0.0181	10	10	Gragg
74	0.0177	11	11	Hammarling
66	0.0171	11	11	VanDooren
42	0.0152	9	9	Moler
79	0.0151	8	8	Gutknecht
32	0.0142	9	9	VanLoan
59	0.0135	8	8	Eisenstat
98	0.0133	8	8	Paige
46	0.0130	7	7	NTrefethen
49	0.0129	6	6	Varga
96	0.0128	7	7	Meyer
77	0.0128	7	7	Stewart
73	0.0127	8	8	Luk
78	0.0127	7	7	Bunch
53	0.0127	6	6	Widlund
72	0.0125	7	7	Reichel
41	0.0125	8	8	George
82	0.0124	6	6	Ipsen
83	0.0122	6	6	Greenbaum
58	0.0113	7	7	Bjorck
97	0.0107	6	6	Nichols
51	0.0106	6	6	Kagstrom
80	0.0106	6	6	Laub
52	0.0104	6	6	Barlow
60	0.0103	6	6	Zha
69	0.0102	6	6	Duff
62	0.0100	6	6	Park
89	0.0099	5	5	BunseGerstner
63	0.0098	5	5	Arioli

43	0.0097	6	6	Gilbert
67	0.0096	6	6	Liu
87	0.0096	5	5	Hansen
47	0.0090	5	5	Nachtigal
54	0.0090	4	4	Bjorstad
2	0.0088	5	5	Wilkinson
23	0.0088	5	5	Harrod
99	0.0087	5	5	Gill
92	0.0086	5	5	Sameh
91	0.0086	5	5	Berry
15	0.0086	5	5	Boley
76	0.0085	4	4	Fischer
50	0.0085	3	3	Young
61	0.0084	5	5	VanHuffel
100	0.0084	3	3	Jessup
48	0.0083	4	4	Kahan
35	0.0083	5	5	Bojanczyk
65	0.0082	5	5	Ng
93	0.0082	4	4	Ammar
55	0.0079	4	4	OLeary
84	0.0079	3	3	Ruhe
19	0.0078	4	4	Kaufman
56	0.0076	4	4	NHigham
37	0.0075	3	3	Marek
75	0.0075	3	3	Szyld
103	0.0074	3	3	Starke
34	0.0072	4	4	Saunders
25	0.0072	4	4	Funderlic
39	0.0072	4	4	Bai
102	0.0072	3	3	Hochbruck
88	0.0071	4	4	Elden
71	0.0070	4	4	Tang
38	0.0069	3	3	Kuo
40	0.0069	3	3	Tong
4	0.0068	3	3	He
13	0.0067	2	2	Kincaid
14	0.0067	2	2	Crevelli
94	0.0065	3	3	Warner
17	0.0065	3	3	Byers
21	0.0064	3	3	Fierro

31	0.0064	2	2	Wold
45	0.0062	3	3	Pothen
36	0.0060	3	3	Dubrulle
57	0.0058	2	2	Boman
10	0.0058	3	3	Overton
9	0.0057	2	2	Modersitzki
68	0.0056	2	2	Smith
95	0.0056	2	2	Davis
33	0.0056	2	2	Chandrasekaran
27	0.0055	2	2	Cullum
28	0.0055	2	2	Strakos
64	0.0054	2	2	MuntheKaas
7	0.0053	2	2	Ashby
85	0.0053	2	2	ATrefethen
29	0.0052	2	2	Saied
30	0.0052	2	2	Ong
18	0.0052	2	2	Benzi
101	0.0052	2	2	Mathias
8	0.0052	2	2	LeBorne
12	0.0052	2	2	Borges
6	0.0051	2	2	Kenney
70	0.0050	2	2	Henrici

6. Zachary's karate club: social network of friendships between 34 members [50 points]

6.1. Write a Matlab code that ranks the five nodes with the largest degree centrality? What are their degrees?

Results found here can be computed using the file `ex6.m`.

Please find the top 5 nodes by degree centrality, with their degree and their neighbours listed below:

```
Node       Degree: Neighbours...
             34 16: 9, 10, 14, 15, 16, 19, 20, 21, 23, 24, 27, 28, 29, 30, 31, 32, 33,
              1 15: 2, 3, 4, 5, 6, 7, 8, 9, 11, 12, 13, 14, 18, 20, 22, 32,
             33 11: 3, 9, 15, 16, 19, 21, 23, 24, 30, 31, 32, 34,
              3  9: 1, 2, 4, 8, 9, 10, 14, 28, 29, 33,
              2  8: 1, 3, 4, 8, 14, 18, 20, 22, 31,
```

6.2. Rank the five nodes with the largest eigenvector centrality. What are their (properly normalized) eigenvector centralities?

Results found here can be computed using the file `ex6.m`.

Please find the top 5 nodes by eigenvector centrality (page-rank column) listed below:

	page-rank	in	out	author
34	0.1009	17	17	34
1	0.0970	16	16	1
33	0.0717	12	12	33
3	0.0571	10	10	3
2	0.0529	9	9	2

6.3. Are the rankings in (a) and (b) identical? Give a brief verbal explanation of the similarities and differences.

The rankings found are identical, even though if we normalize the degree centrality to the greatest eigenvector centrality we find slighly different values ($[0.1009, 0.0946, 0.0694, 0.0568, 0.0505]$) w.r.t the actual eigenvector centrality.

The identical rankings may be explained by the fact that by computing the eigenvector centrality we are effectively applying PageRank to a symmetrical matrix, i.e. to a graph with bidirectional links. Since the links are bidirectional, we effectively make all the nodes in the graph of the same "importance" to the eyes of PageRank, thus avoiding a case where a node has high PageRank thank to connections with few, but very "important" nodes. Therefore PageRank is simply reduced to a priotarization of nodes with many edges, i.e. the degree centrality ranking.

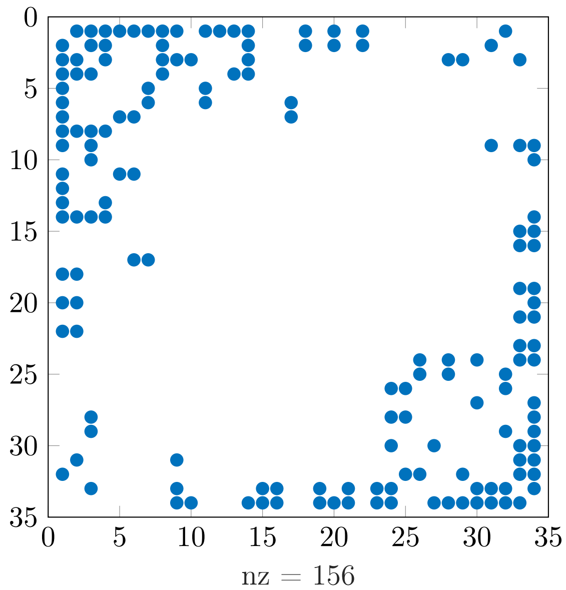
6.4. Use spectral graph partitioning to find a near-optimal split of the network into two groups of 16 and 18 nodes, respectively. List the nodes in the two groups. How does spectral bisection compare to the real split observed by Zachary?

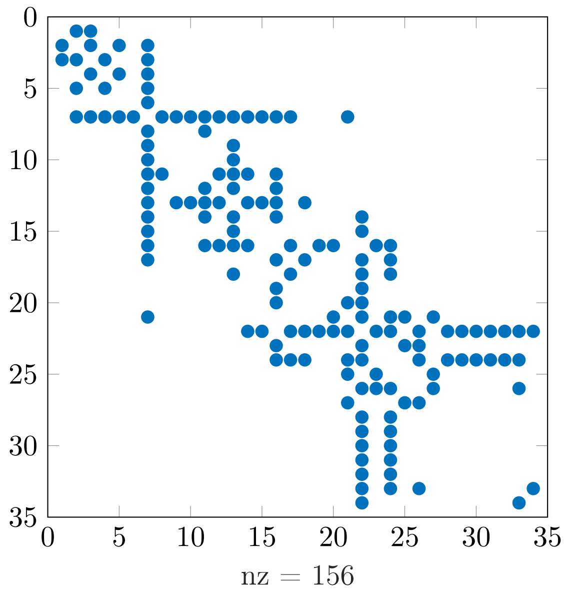
The spectral bisection of the matrix a in two groups of 16 and 18 members respectively is identical to the real split observed by Zachary. To compute the split, the script `ex6.m` was used.

Here are the (sorted) two groups found:

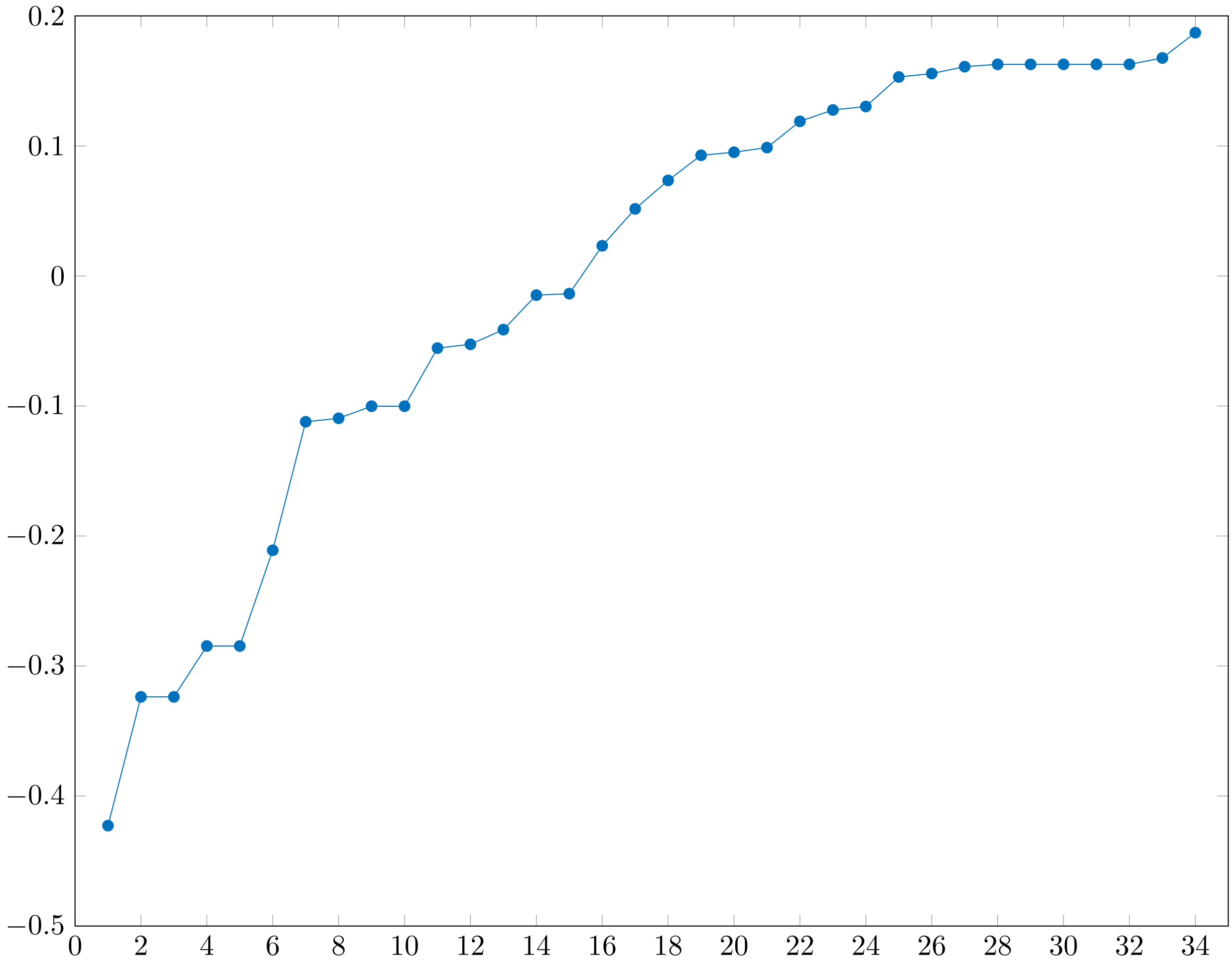
$$G_1 = [1, 2, 3, 4, 5, 6, 7, 8, 11, 12, 13, 14, 17, 18, 20, 22]$$
$$G_2 = [9, 10, 15, 16, 19, 21, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34]$$

Here are the spy plots of the original matrix A (to the left) and the spectral bisected permutated matrix (to the right):





Here is a plot of the sorted elements of the second eigenvector λ_2:



and here are the actual (sorted) values of λ_2:

$$sort(\lambda_2) = [-0.4228, -0.3237, -0.3237, -0.2846, -0.2846, -0.2110, -0.1121, -0.1095, -0.1002, -0.1002, -0.0555, -0.0526, -0.0413, -0.0147, -0.0136, 0.0232, 0.0516, 0.0735, 0.0928, 0.0952, 0.0988, 0.1189, 0.1277, 0.1303, 0.1530, 0.1557, 0.1610, 0.1628, 0.1628, 0.1628, 0.1628, 0.1628, 0.1677, 0.1871]^T$$

As it can be seen above, there are only 15 negative values out the 16 we would need to obtain a perfect 16/18 partition. We therefore add the index corresponding to the smallest positive value in λ_2 in the set of indexes of group 1. This seems to be a good approximation since indeed we get the same partitioning as the original Zachary's one.

References

[1] The social network of a karate club at a US university, M. E. J. Newman and M. Girvan, Phys. Rev. E 69,026113 (2004) pp. 219-229.