CS3907/CS6444 Big Data and Analytics

Class project #1

R and Graph Analytics

Due Date: September 30, 2019 COB

### Description

**Data Set:** JDK Dependency

**Description:**

This is the software class dependency network of the JDK 1.6.0.7 framework. The network is directed. Nodes represent classes. An edge between them indicates that there exists a dependency between two classes. As there may be multiple references between classes the network has multiple edges.

### DataSet:

### subelj\_jdk.tar.bz2 (use 7-Zip to unpack)

### subelj.tar.bz2 (soure code to use to unextract from file/or write you own R code)

### konect-handbook.pdf

**Requirements:**

1. Install the igraph package from one of the CRAN mirrors. You may also use igraphdata package and rgraph (included in the SNA package) as well.

3. You will have to determine how to load the data into a data structure usable by the graph packages.

I have already given you some hints, such as read\_line and read\_table.

Note that the books in Outline:TechnicalBooks might be useful.

Here’s a procedure for creating a graph from a file. Note that your data is already loaded into R data structures. You will need to determine what type of data structures they are and modify this procedure as appropriate.

Creating a graph

Here is some help:

1. edges1<-read.table(<some file>)   assuming you have set working directory

2. convert to matrix:  em <-as.matrix(edges1)

3. extract vectors: v1 <- em[<$rows>:1], v2<-em[#rows:2] where you need to find the number of rows in the matrix. This gets the two vertices as separate vectors

4. relations<- data.frame(from=v1,to=v2)

5. g<-graph.data.frame(relations,directed=TRUE)  need to have installed igraph

6. plot(g)

This should work, Then you can apply various functions to it.

REMEMBER: Some functions work on matrices, some on data frames, and some on both.

4. Experiment with at least 10 of the functions that I have shown in the lecture notes/associated PPT file on Blackboard applied to the graph that you generate from the data set. Present the results in your project report.

This is a very large data set. You may have to simplify the graph somewhat in order to execute this project. If so, describe how you simplified the graph. You may use the *simplify* function, but you may have to do more than that.

If you have problems, ask a question at office hours or in class.

5. Explore other functions in the igraph package – at least 15 of them not shown in the lecture notes. Apply them to the graph you generate.You may have to do some programming in R. There are numerous books posted on the Blackboard.

6. Determine the (a) central node(s) in the graph, (b) longest path, (c) largest clique, (d) betweenness centrality and (e) power centrality.

a. Is there more than one node with the most degrees?

b. Are there multiple longest paths?

c. Are there multiple cliques?

d. Are there one or mulltiple with the highest ego?

e. What is the difference in betweenness centrality vs. power centrality for the cases you find? Consider comparing the nodes that are members of each set. Are there common nodes?

In each case what do you think the data tells you?

7. Find the 20 nodes with the largest networks, e.g., having the greatest diameters. Do any of these circles overlap?

a. Build a matrix of 20 nodes with their reachability to the 3rd level

b. Determine which of the 20 nodes share common nodes, if any, and, for each common node, list the nodes that share that common node.

8. Deliverables: You will deliver, by putting a zipfile in your group’s Blackboard file, with the following naming convention: Group-N-Project-1.zip, where N is your group number. Your deliverable should encompass the following items:

* A listing of all R functions that you have written
* Demonstrations of the igraph functions that you have explored per #4 and #5.
* Answers for #6 and #7.

Be clear about what you are doing with each function. Identify any problems you had and how to solve them.

Remember to name and save your workspace! In your Group area would be a good place so all members can get to it.

Include in your Word document the results required

(use a CTRL-ALT-PrintScreen) to grab the screen

You may use Irfanview 4.41 or later, [irfanview@gmx.net](mailto:irfanview@gmx.net). Paste in the screen image, and copy the image as JPEG to drop into your Word document.

7. Project #1 Value: 10 points

a. Overall description and documented R functions: 1.5 points

b. Item #4/#5 demonstrations – 2.5 pts

c. Item #6 demonstration – 2.5 points

d. Item #7 demonstration: 3.5 points

Note: To prepare for working with this data set:

>install.packages(“igraph”)

…. Lots of stuff omitted ….

So, you can see there is a lot of additional code required to support this package that RStudio downloads for you.

**Data Set Notes**:

[konect logo](http://konect.uni-koblenz.de/)

[KONECT](http://konect.uni-koblenz.de/)

* [Home](http://konect.uni-koblenz.de/)
* [Networks](http://konect.uni-koblenz.de/networks/)
* [Statistics](http://konect.uni-koblenz.de/statistics/)
* [Plots](http://konect.uni-koblenz.de/plots/)
* [Search](http://konect.uni-koblenz.de/search)
* [Downloads](http://konect.uni-koblenz.de/downloads/)
* [Software](http://konect.uni-koblenz.de/software)
* [Publications](http://konect.uni-koblenz.de/publications)
* [License](http://konect.uni-koblenz.de/license)
* [About](http://konect.uni-koblenz.de/about)
* [Help](http://konect.uni-koblenz.de/help/)

Top of Form

|  |  |  |
| --- | --- | --- |
| |  |  | | --- | --- | |  |  | |

Bottom of Form

Top of Form

|  |
| --- |
|  |

Bottom of Form

[KONECT](http://konect.uni-koblenz.de/) > [Networks](http://konect.uni-koblenz.de/networks/) > **JDK dependency**

# JDK dependency

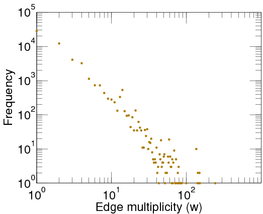
## About this network

This is the software class dependency network of the JDK 1.6.0.7 framework. The network is directed. Nodes represent classes. An edge between them indicates that there exists a dependency between two classes. As there may be multiple references between classes the network has multiple edges.

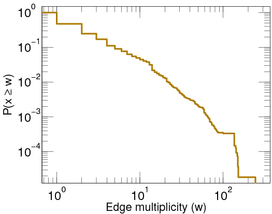
## Network info

|  |  |
| --- | --- |
| [Code](http://konect.uni-koblenz.de/help/codes) | DJ |
| [Category](http://konect.uni-koblenz.de/help/categories) | [⬤ Software](http://konect.uni-koblenz.de/help/categories) |
| Data source | <http://lovro.lpt.fri.uni-lj.si/support.jsp> |
| Vertex type | Class |
| Edge type | Dependency |
| [**Format**](http://konect.uni-koblenz.de/help/network_format) | [Directed: Edges are directedDirected](http://konect.uni-koblenz.de/help/network_format) |
| [**Edge weights**](http://konect.uni-koblenz.de/help/edge_weights) | [Multiple unweighted: Multiple edges are possibleMultiple unweighted](http://konect.uni-koblenz.de/help/edge_weights) |
| [Metadata](http://konect.uni-koblenz.de/help/metadata) | Entity metadata:  Nodes are annotated with metadata[Entity](http://konect.uni-koblenz.de/help/metadata) |
| [**Size**](http://konect.uni-koblenz.de/statistics/size) | 6,434 vertices (classes) |
| [**Volume**](http://konect.uni-koblenz.de/statistics/volume) | 150,985 edges (dependencies) |
| [**Unique volume**](http://konect.uni-koblenz.de/statistics/volume) | 53,892 edges (dependencies) |
| [**Average degree (overall)**](http://konect.uni-koblenz.de/statistics/avgdegree) | 46.933 edges / vertex |
| [**Fill**](http://konect.uni-koblenz.de/statistics/fill) | 0.0013021 edges / vertex2 |
| [**Maximum degree**](http://konect.uni-koblenz.de/statistics/maxdegree) | 32,530 edges |
| [**Reciprocity**](http://konect.uni-koblenz.de/statistics/reciprocity) | 0.868% |
| [**Size of LCC**](http://konect.uni-koblenz.de/statistics/coco) | 6,434 vertices (network is connected) |
| [**Size of LSCC**](http://konect.uni-koblenz.de/statistics/coco) | 77 vertices |
| [**Wedge count**](http://konect.uni-koblenz.de/statistics/twostars) | 52,676,393 |
| [**Claw count**](http://konect.uni-koblenz.de/statistics/threestars) | 92,489,410,361 |
| [**Triangle count**](http://konect.uni-koblenz.de/statistics/triangles) | 194,842 |
| [**Square count**](http://konect.uni-koblenz.de/statistics/squares) | 82,893,262 |
| [**4-tour count**](http://konect.uni-koblenz.de/statistics/tour4) | 873,958,984 |
| [**Power law exponent (estimated) with dmin**](http://konect.uni-koblenz.de/statistics/power) | 2.3610 (dmin = 6) |
| [**Gini coefficient**](http://konect.uni-koblenz.de/statistics/gini) | 75.1% |
| [**Relative edge distribution entropy**](http://konect.uni-koblenz.de/statistics/dentropyn) | 83.0% |
| [**Assortativity**](http://konect.uni-koblenz.de/plots/assortativity) | –0.22302 |
| [**Clustering coefficient**](http://konect.uni-koblenz.de/statistics/clusco) | 1.11% |
| [**Diameter**](http://konect.uni-koblenz.de/statistics/diameter) | 7 edges |
| [**90-percentile effective diameter**](http://konect.uni-koblenz.de/statistics/diameter) | 2.47 edges |
| [**Mean shortest path length**](http://konect.uni-koblenz.de/statistics/diameter) | 2.19 edges |
| [**Spectral norm**](http://konect.uni-koblenz.de/statistics/snorm) | 1062.1 |
| [**Algebraic connectivity**](http://konect.uni-koblenz.de/statistics/alcon) | 0.40122 |

## [Edge weight and multiplicity distribution](http://konect.uni-koblenz.de/plots/edge_weight_distribution)

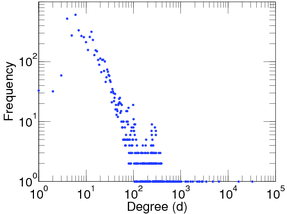
[](http://konect.uni-koblenz.de/img/weights.c.subelj_jdk.png)

Edge multiplicity distribution

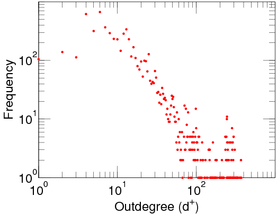
[](http://konect.uni-koblenz.de/img/weights.d.subelj_jdk.png)

Cumulative edge multiplicity distribution

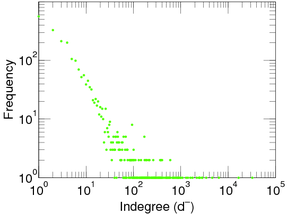
## [Degree distributions](http://konect.uni-koblenz.de/plots/degree_distribution)

[](http://konect.uni-koblenz.de/img/degree.a.subelj_jdk.png)

Degree distribution

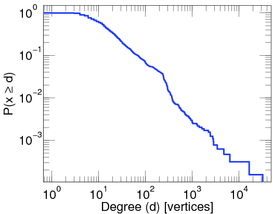
[](http://konect.uni-koblenz.de/img/degree.u.subelj_jdk.png)

Outdegree distribution

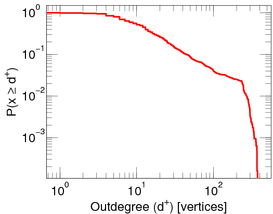
[](http://konect.uni-koblenz.de/img/degree.v.subelj_jdk.png)

Indegree distribution

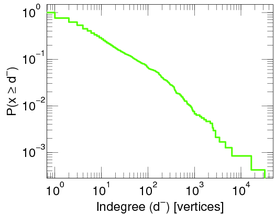
## [Cumulative degree distributions](http://konect.uni-koblenz.de/plots/bidd)

[](http://konect.uni-koblenz.de/img/bidd.ax.subelj_jdk.png)

Degree distribution

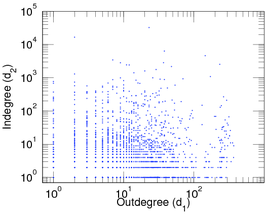
[](http://konect.uni-koblenz.de/img/bidd.ux.subelj_jdk.png)

Outdegree distribution

[](http://konect.uni-koblenz.de/img/bidd.vx.subelj_jdk.png)

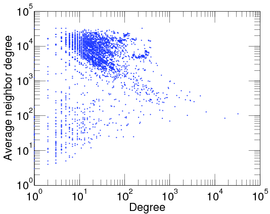
Indegree distribution

## Out/indegree comparison

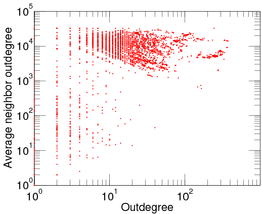
[](http://konect.uni-koblenz.de/img/outin.b.subelj_jdk.png)

Out/indegree comparison

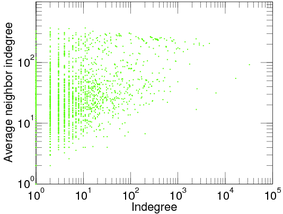
## [Assortativity plot](http://konect.uni-koblenz.de/plots/assortativity)

[](http://konect.uni-koblenz.de/img/assortativity.a.subelj_jdk.png)

Degree distribution

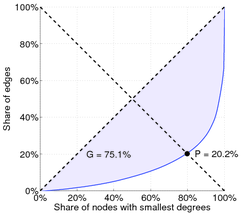
[](http://konect.uni-koblenz.de/img/assortativity.u.subelj_jdk.png)

Outdegree distribution

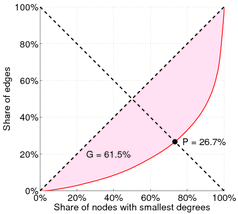
[](http://konect.uni-koblenz.de/img/assortativity.v.subelj_jdk.png)

Indegree distribution

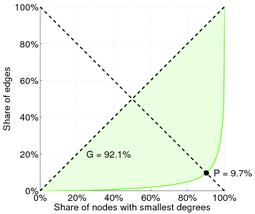
## [Lorenz curves](http://konect.uni-koblenz.de/plots/lorenz_curve)

[](http://konect.uni-koblenz.de/img/lorenz.a.subelj_jdk.png)

Degree distribution

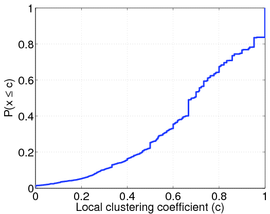
[](http://konect.uni-koblenz.de/img/lorenz.u.subelj_jdk.png)

Outdegree distribution

[](http://konect.uni-koblenz.de/img/lorenz.v.subelj_jdk.png)

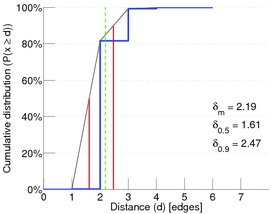
Indegree distribution

## [Clustering coefficient distribution](http://konect.uni-koblenz.de/plots/cluscod)

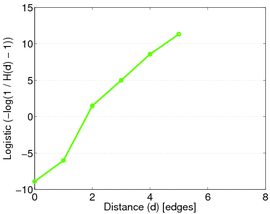
[](http://konect.uni-koblenz.de/img/cluscod.a.subelj_jdk.png)

Clustering coefficient distribution

## [Distance distribution](http://konect.uni-koblenz.de/plots/distance_distribution_plot)

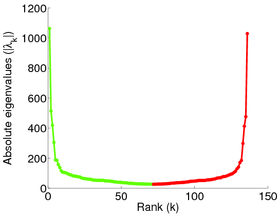
[](http://konect.uni-koblenz.de/img/hopdistr.a.subelj_jdk.png)

Distance distribution

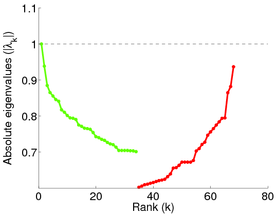
[](http://konect.uni-koblenz.de/img/hopdistr.b.subelj_jdk.png)

Distance distribution on a logistic scale

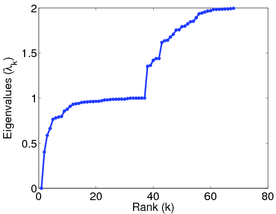
## [Top-*k* eigenvalues](http://konect.uni-koblenz.de/plots/spectrum)

[](http://konect.uni-koblenz.de/img/decomposition.b.sym.subelj_jdk.png)

Top-*k* eigenvalues of **A**

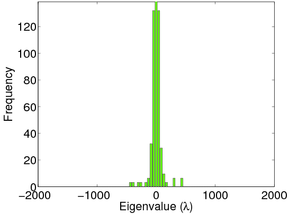
[](http://konect.uni-koblenz.de/img/decomposition.b.sym-n.subelj_jdk.png)

Top-*k* eigenvalues of **N**

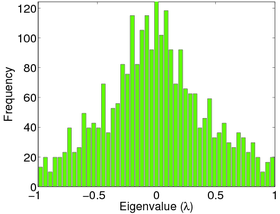
[](http://konect.uni-koblenz.de/img/decomposition.a.lap.subelj_jdk.png)

Top-*k* eigenvalues of **L**

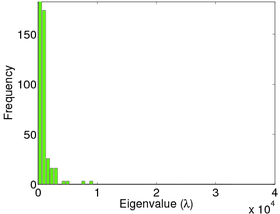
## [Spectral plots](http://konect.uni-koblenz.de/plots/distr)

[](http://konect.uni-koblenz.de/img/distr.a.sym.subelj_jdk.png)

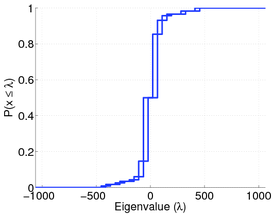
Spectral distribution of the eigenvalues of **A**

[](http://konect.uni-koblenz.de/img/distr.a.sym-n.subelj_jdk.png)

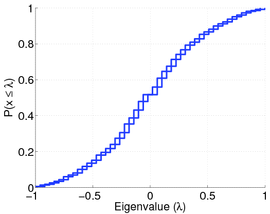
Spectral distribution of the eigenvalues of **N**

[](http://konect.uni-koblenz.de/img/distr.a.lap.subelj_jdk.png)

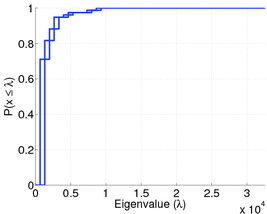
Spectral distribution of the eigenvalues of **L**

[](http://konect.uni-koblenz.de/img/distr.f.sym.subelj_jdk.png)

Cumulative spectral distribution of **A**

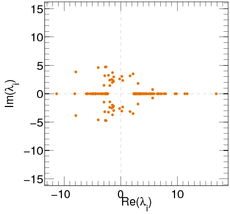
[](http://konect.uni-koblenz.de/img/distr.f.sym-n.subelj_jdk.png)

Cumulative spectral distribution of **N**

[](http://konect.uni-koblenz.de/img/distr.f.lap.subelj_jdk.png)

Cumulative spectral distribution of **L**

## [Complex eigenvalues of the asymmetric adjacency matrix](http://konect.uni-koblenz.de/plots/complex.diag)

[](http://konect.uni-koblenz.de/img/decomposition.complex.diag.subelj_jdk.png)

Complex eigenvalues of the asymmetric adjacency matrix

## Downloads

|  |  |
| --- | --- |
| TSV file: | [[download](http://konect.uni-koblenz.de/downloads/tsv/subelj_jdk.tar.bz2)subelj\_jdk.tar.bz2](http://konect.uni-koblenz.de/downloads/tsv/subelj_jdk.tar.bz2) (256.84 KiB) |
| Extraction code: | [[download](http://konect.uni-koblenz.de/downloads/extraction/subelj.tar.bz2)subelj.tar.bz2](http://konect.uni-koblenz.de/downloads/extraction/subelj.tar.bz2) (15.63 KiB) |

## References

|  |  |
| --- | --- |
| [1] | Jdk dependency network dataset -- KONECT, April 2017. [ [http](http://konect.uni-koblenz.de/networks/subelj_jdk) ] |

[BibTeX](http://konect.uni-koblenz.de/cites/cite.subelj_jdk.bib)

Updated on Thu Apr 27 12:12:40 2017  |  [Privacy](http://konect.uni-koblenz.de/piwik/index.php?module=CoreAdminHome&action=optOut&language=en)

by [Jérôme Kunegis](https://west.uni-koblenz.de/de/ueber-uns/team/dr-j%C3%A9r%C3%B4me-kunegis)

[University of Koblenz-Landau Logo](http://www.uni-koblenz.de/)[WeST Institute Logo](http://west.uni-koblenz.de/)[ROBUST Project Logo](http://www.robust-project.eu/)[Gplv3Creative Commons License](http://konect.uni-koblenz.de/license)