report

May 29, 2019

1 Setup

```
Collecting coclust
Requirement already satisfied: scipy in /opt/conda/lib/python3.6/site-packages (from coclust)
Requirement already satisfied: scikit-learn in /opt/conda/lib/python3.6/site-packages (from co
Requirement already satisfied: numpy in /opt/conda/lib/python3.6/site-packages (from coclust)
Installing collected packages: coclust
Successfully installed coclust-0.2.1
```

2 Loading data

A fonction to get the name of a library according to its definition and the depth.

'java.util.ArrayList', depth=2 -> 'java.util'

Define the mapping of imports to an ids.
Some basic vector operators
Count for every commits the number of time a library was imported in all modified files.

```
Out[12]: {}
```

```
Total number of commit : 1993
Made by 8 unique contibutors
Represented by 1097 imports
mapped to 174 imports
```

Sum the commits by users.
Then drop the libraries imported once or less.
Finally sort the data in row and column by the number of modifications.

```
Out[15]:                          index
                                   mean  min   max count
          author
          Anders Nawroth   1403.568627  598  1819    51
          Emil Eifrem       273.608108    0   948    74
```

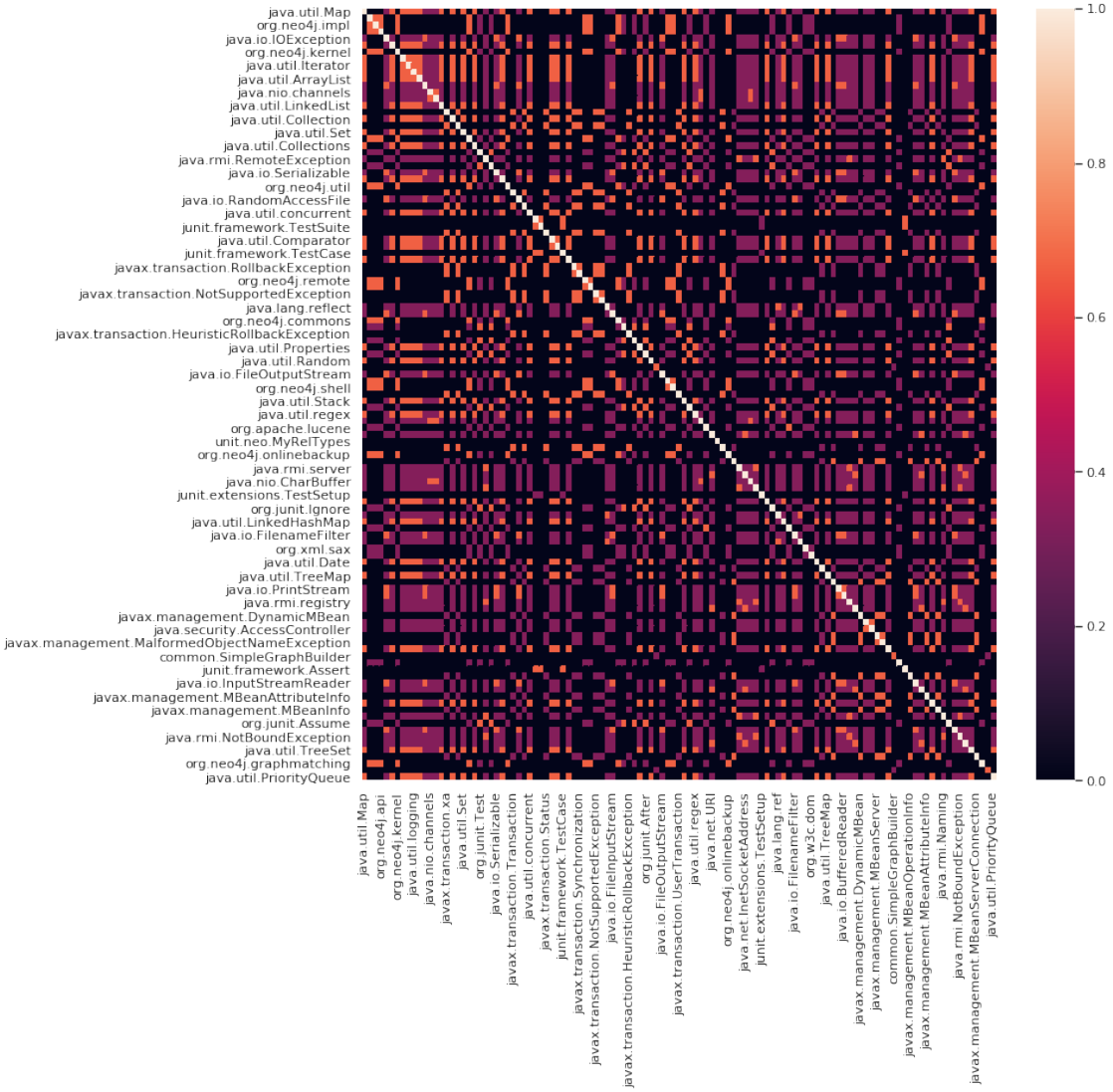
Henrik Larsson	398.000000	398	398	1
Johan Svensson	746.275904	40	1989	830
Mattias Persson	1235.443843	6	1992	739

Out[18]:

	first	last	count	coef
java.util.ConcurrentModificationException	1634	1636	3	0.584277
java.io.FileWriter	494	923	6	0.199760
javax.management.MBeanOperationInfo	1670	1823	36	0.774223
common.StandardGraphs	1310	1312	3	0.358765
jline.ConsoleReader	0	1192	1	0.299476

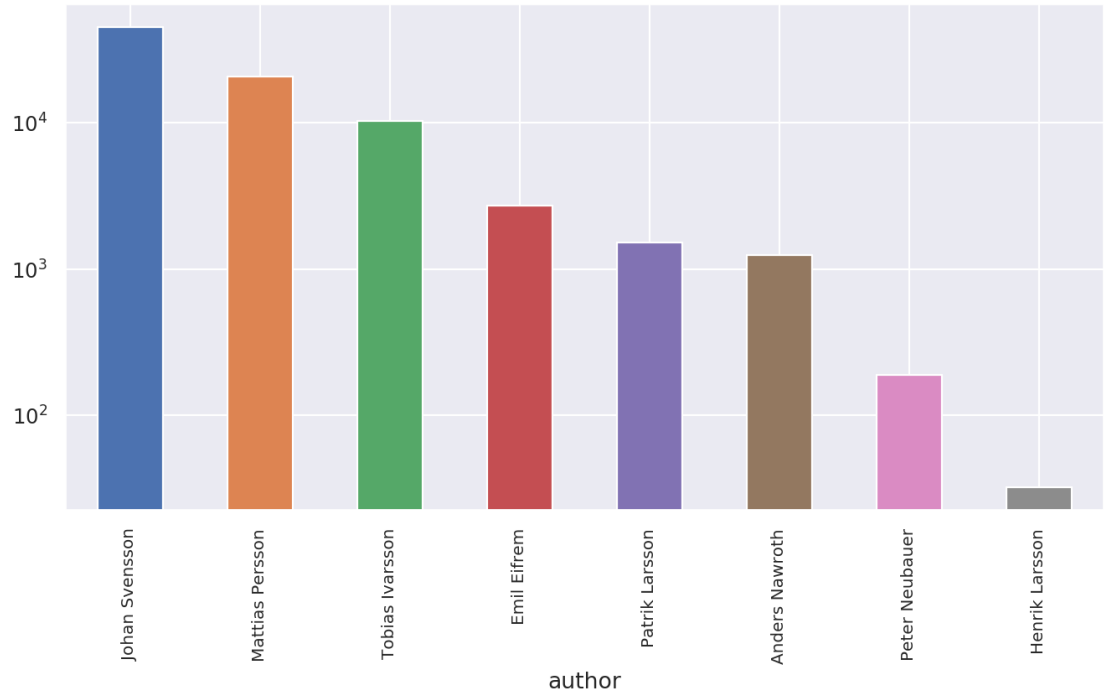
2.1 Distance between libraries

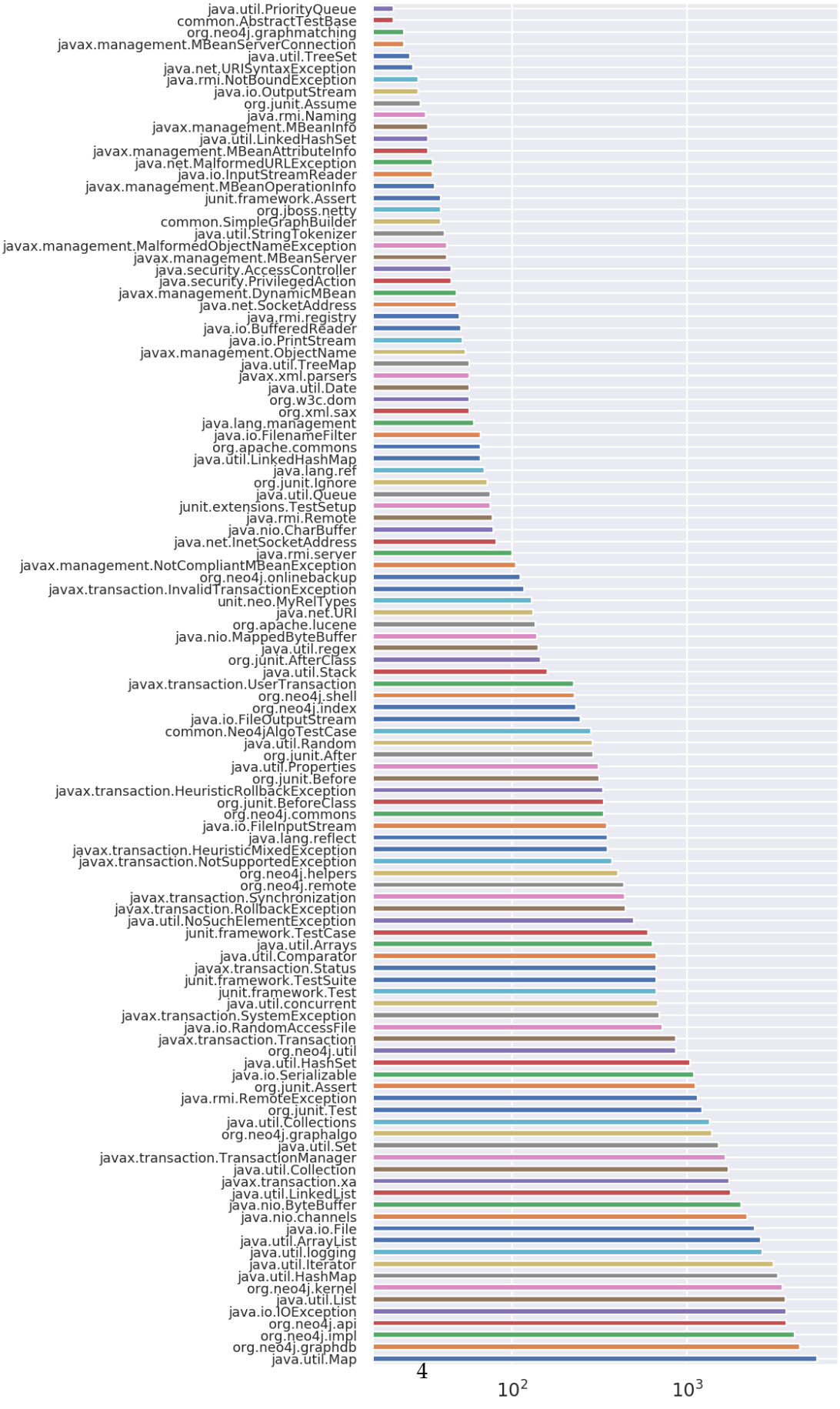
Out[19]: []

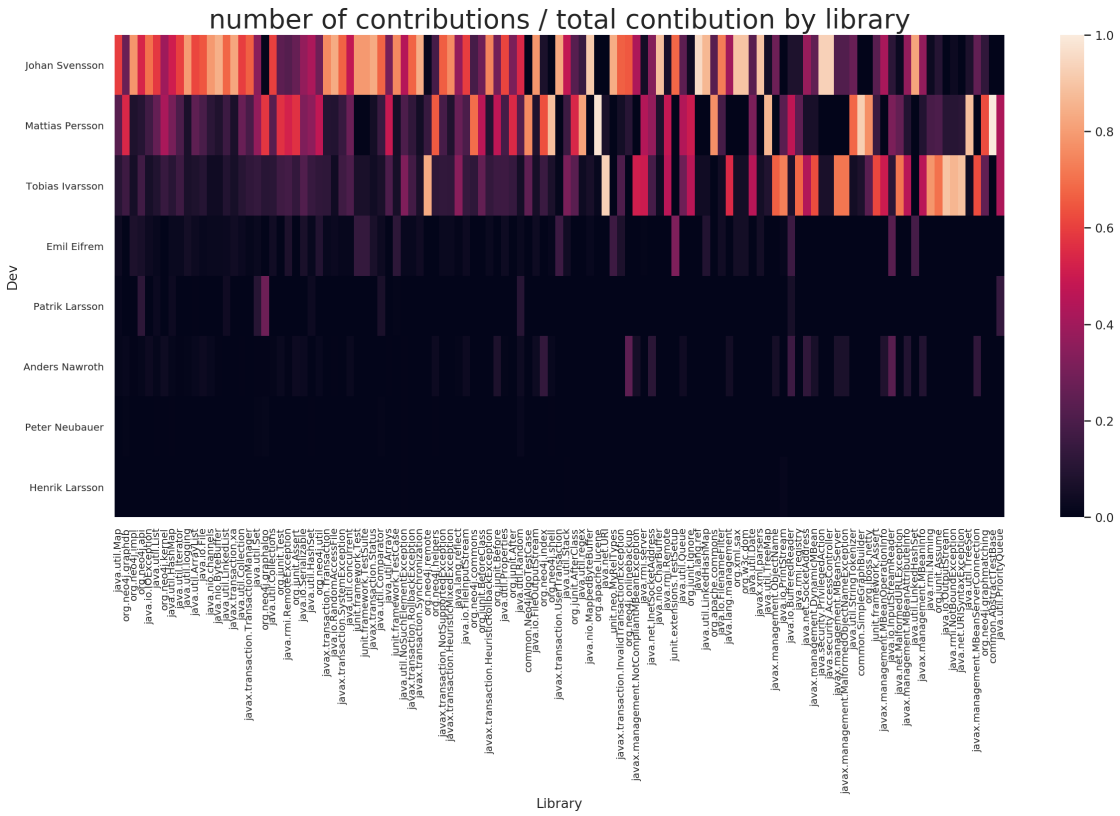


3 Analysis

3.1 Summary

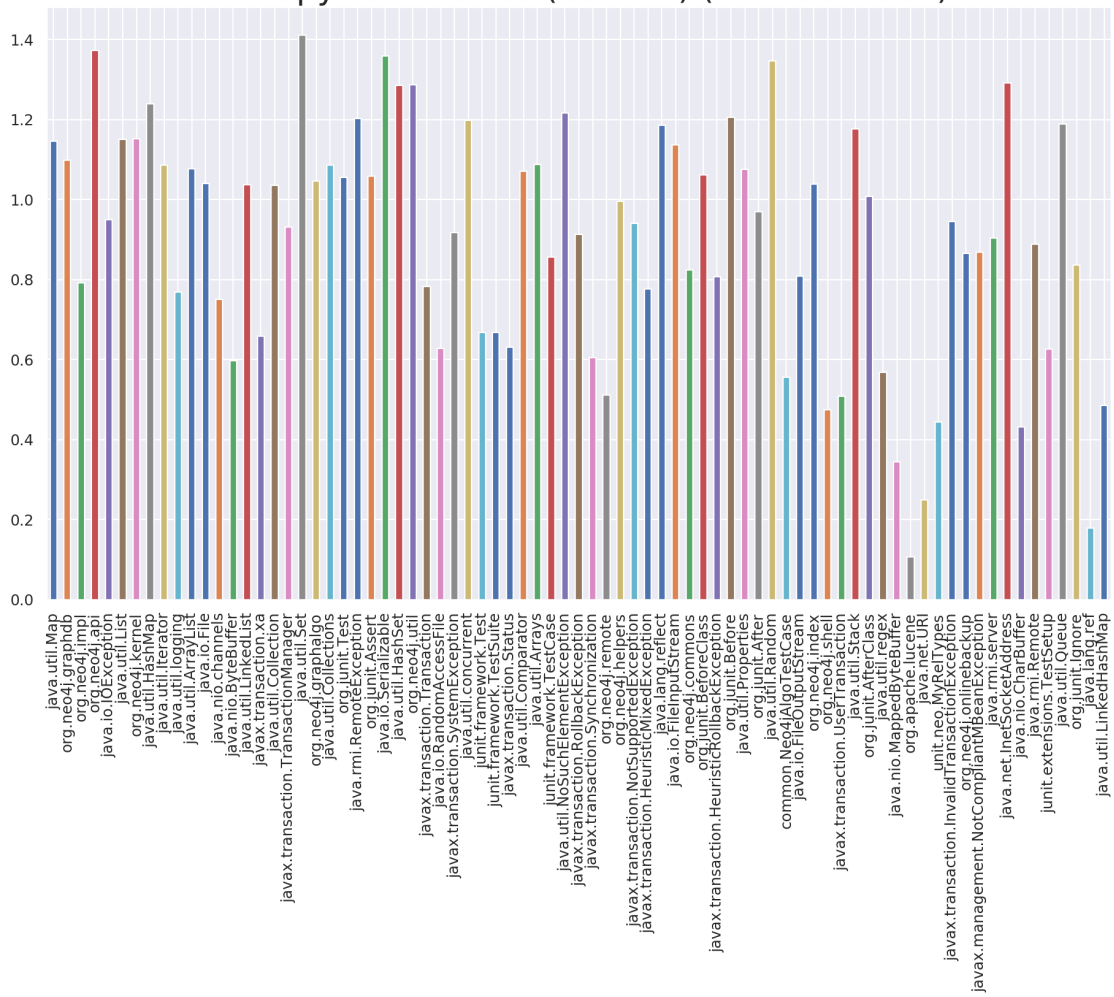



java.util.PriorityQueue
common.AbstractTestBase
org.neo4j.graphmatching
javax.management.MBeanServerConnection
java.util.TreeSet
java.net.URISyntaxException
java.rmi.NotBoundException
java.io.OutputStream
org.junit.Assume
java.rmi.Naming
javax.management.MBeanInfo
java.util.LinkedHashSet
javax.management.MBeanAttributeInfo
java.net.MalformedURLException
java.io.InputStreamReader
javax.management.MBeanOperationInfo
junit.framework.Assert
org.jboss.netty
common.SimpleGraphBuilder
java.util.StringTokenizer
javax.management.MalformedObjectNameException
javax.management.MBeanServer
java.security.AccessController
java.security.PrivilegedAction
javax.management.DynamicMBean
java.net.SocketAddress
java.rmi.registry
java.io.BufferedReader
java.io.PrintStream
javax.management.ObjectName
java.util.TreeMap
javax.xml.parsers
java.util.Date
org.w3c.dom
org.xml.sax
java.lang.management
java.io.FilenameFilter
org.apache.commons
java.util.LinkedHashMap
java.lang.ref
org.junit.Ignore
java.util.Queue
junit.extensions.TestSetup
java.rmi.Remote
java.nio.CharBuffer
java.net.InetSocketAddress
java.rmi.server
javax.management.NotCompliantMBeanException
org.neo4j.onlinebackup
javax.transaction.InvalidTransactionException
unit.neo.MyRelTypes
java.net.URI
org.apache.lucene
java.nio.MappedByteBuffer
java.util.regex
org.junit.AfterClass
java.util.Stack
javax.transaction.UserTransaction
org.neo4j.shell
org.neo4j.index
java.io.FileOutputStream
common.Neo4jAlgoTestCase
java.util.Random
org.junit.After
java.util.Properties
org.junit.Before
javax.transaction.HeuristicRollbackException
org.junit.BeforeClass
org.neo4j.commons
java.io.FileInputStream
java.lang.reflect
javax.transaction.HeuristicMixedException
javax.transaction.NotSupportedException
org.neo4j.helpers
org.neo4j.remote
javax.transaction.Synchronization
javax.transaction.RollbackException
java.util.NoSuchElementException
junit.framework.TestCase
java.util.Arrays
java.util.Comparator
javax.transaction.Status
junit.framework.TestSuite
junit.framework.Test
java.util.concurrent
javax.transaction.SystemException
java.io.RandomAccessFile
javax.transaction.Transaction
org.neo4j.util
java.util.HashSet
java.io.Serializable
org.junit.Assert
java.rmi.RemoteException
org.junit.Test
java.util.Collections
org.neo4j.graphalgo
java.util.Set
javax.transaction.TransactionManager
java.util.Collection
javax.transaction.xa
java.util.LinkedList
java.nio.ByteBuffer
java.nio.channels
java.io.File
java.util.ArrayList
java.util.logging
java.util.Iterator
java.util.HashMap
org.neo4j.kernel
java.util.List
java.io.IOException
org.neo4j.api
org.neo4j.impl
org.neo4j.graphdb
java.util.Map
4
10^2
10^3



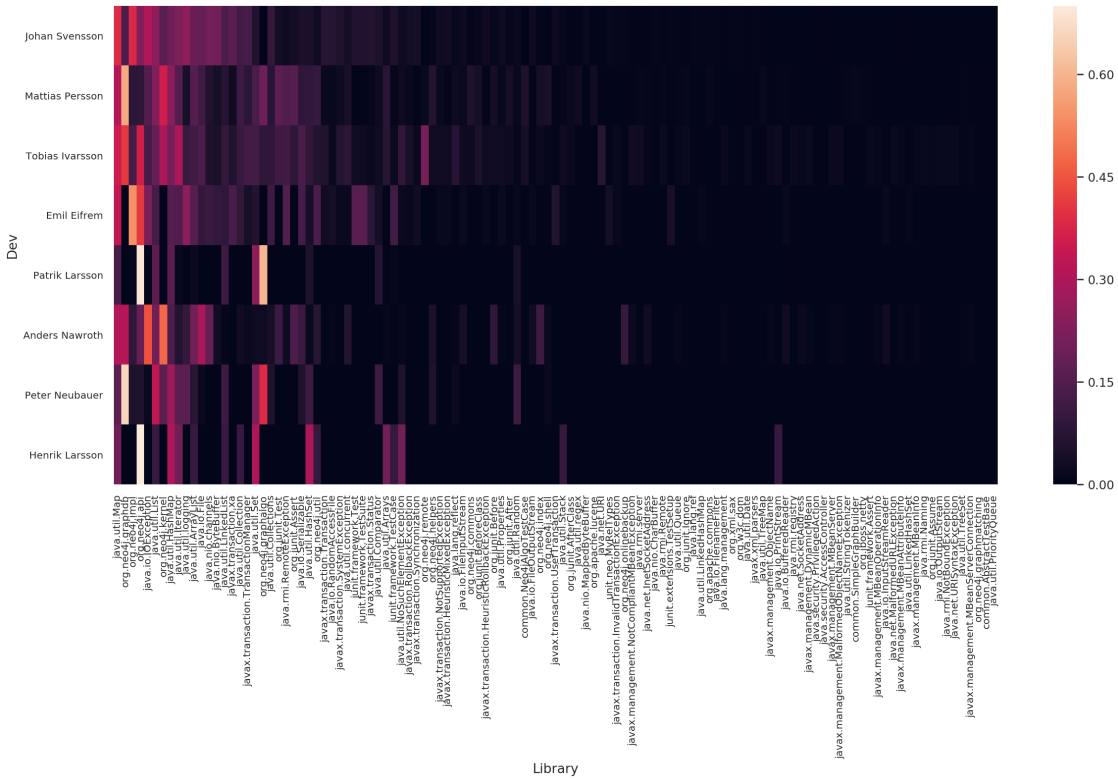
3.1.1 Entropy

Mostly indicate dispersion here.


Entropy for libraries (column) (75 most used)
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
java.util.Map
org.neo4j.graphdb
org.neo4j.impl
org.neo4j.api
java.io.IOException
java.util.List
org.neo4j.kernel
java.util.HashMap
java.util.Iterator
java.util.logging
java.util.ArrayList
java.io.File
java.nio.channels
java.nio.ByteBuffer
java.util.LinkedList
javax.transaction.xa
java.util.Collection
javax.transaction.TransactionManager
java.util.Set
org.neo4j.graphalgo
java.util.Collections
org.junit.Test
java.rmi.RemoteException
org.junit.Assert
java.io.Serializable
java.util.HashSet
org.neo4j.util
javax.transaction.Transaction
java.io.RandomAccessFile
javax.transaction.SystemException
java.util.concurrent
junit.framework.Test
junit.framework.TestSuite
javax.transaction.Status
java.util.Comparator
java.util.Arrays
junit.framework.TestCase
java.util.NoSuchElementException
javax.transaction.RollbackException
javax.transaction.Synchronization
org.neo4j.remote
org.neo4j.helpers
javax.transaction.NotSupportedException
javax.transaction.HeuristicMixedException
java.lang.reflect
java.io.FileInputStream
org.neo4j.commons
org.junit.BeforeClass
javax.transaction.HeuristicRollbackException
org.junit.Before
java.util.Properties
org.junit.After
java.util.Random
common.Neo4jAlgoTestCase
java.io.FileOutputStream
org.neo4j.index
org.neo4j.shell
javax.transaction.UserTransaction
java.util.Stack
org.junit.AfterClass
java.util.regex
java.nio.MappedByteBuffer
org.apache.lucene
java.net.URI
unit.neo.MyRelTypes
javax.transaction.InvalidTransactionException
org.neo4j.onlinebackup
javax.management.NotCompliantMBeanException
java.rmi.server
java.net.InetSocketAddress
java.nio.CharBuffer
java.rmi.Remote
junit.extensions.TestSetup
java.util.Queue
org.junit.Ignore
java.lang.ref
java.util.LinkedHashMap

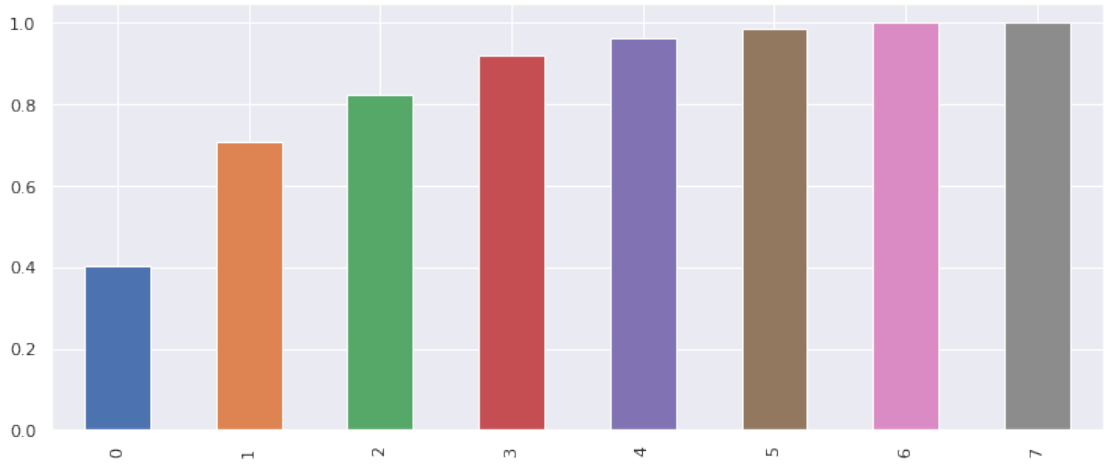
3.2 Nomalisation L2

This representation aims to represent the developpers contributions.



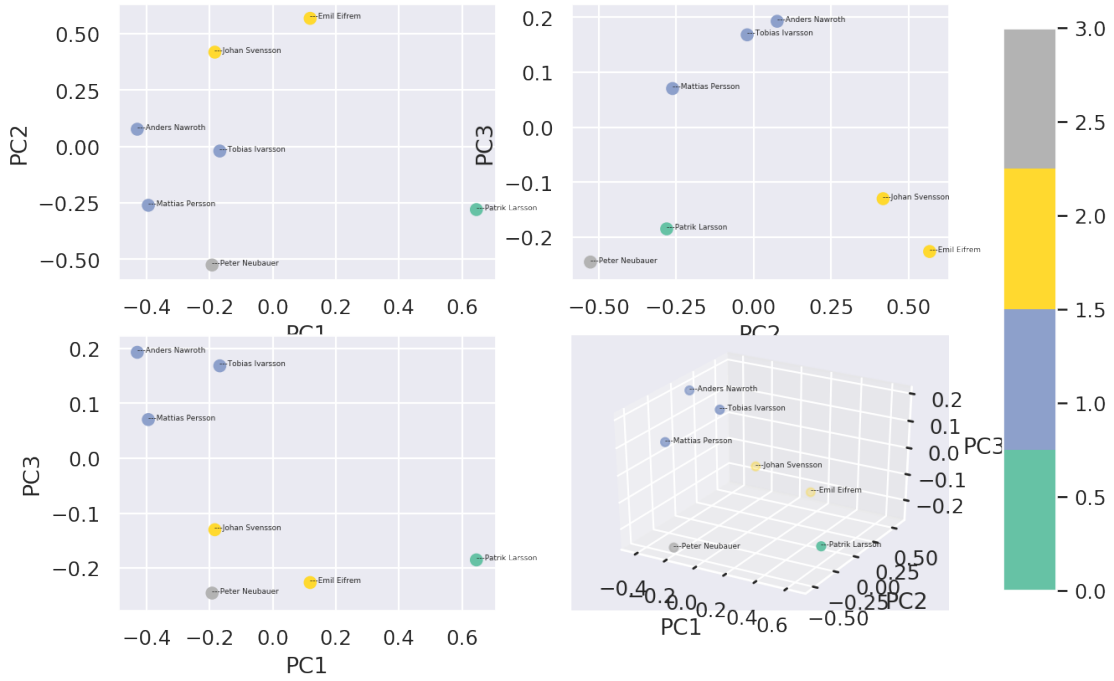
3.2.1 PCA

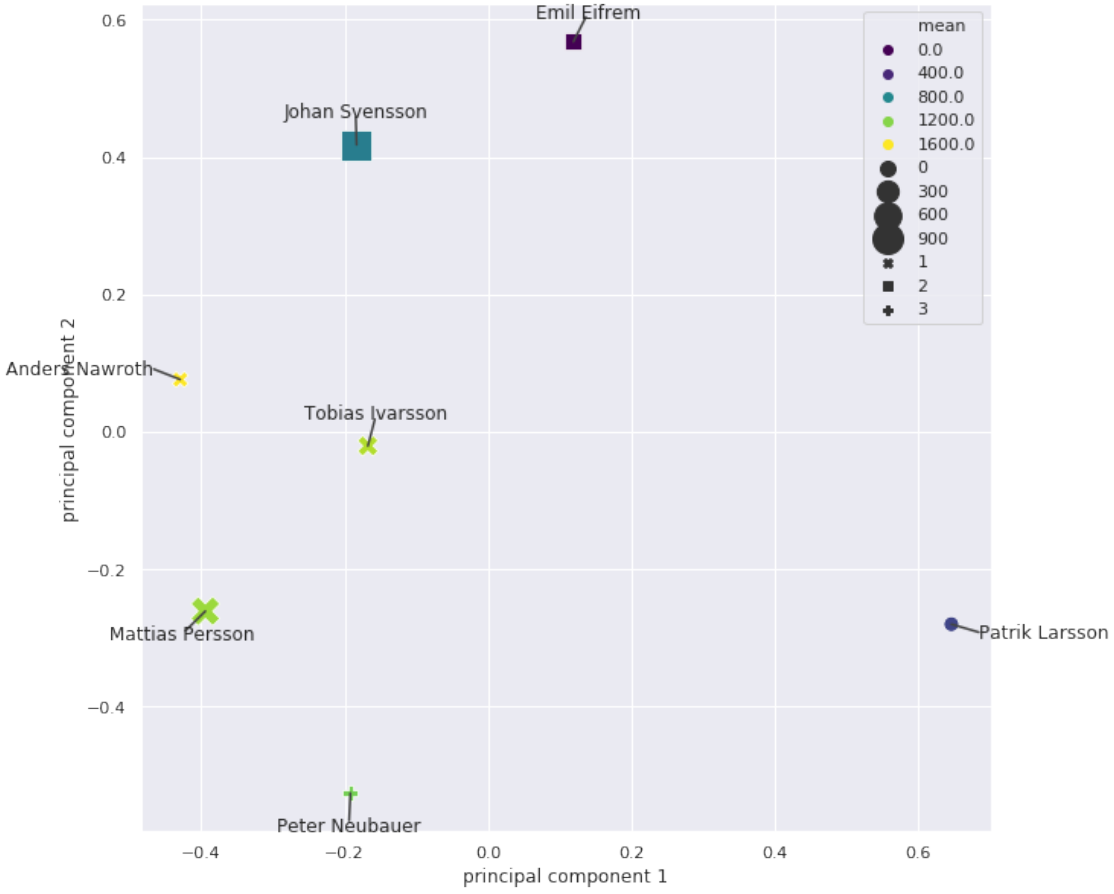
Cumulative sum of the explained variance



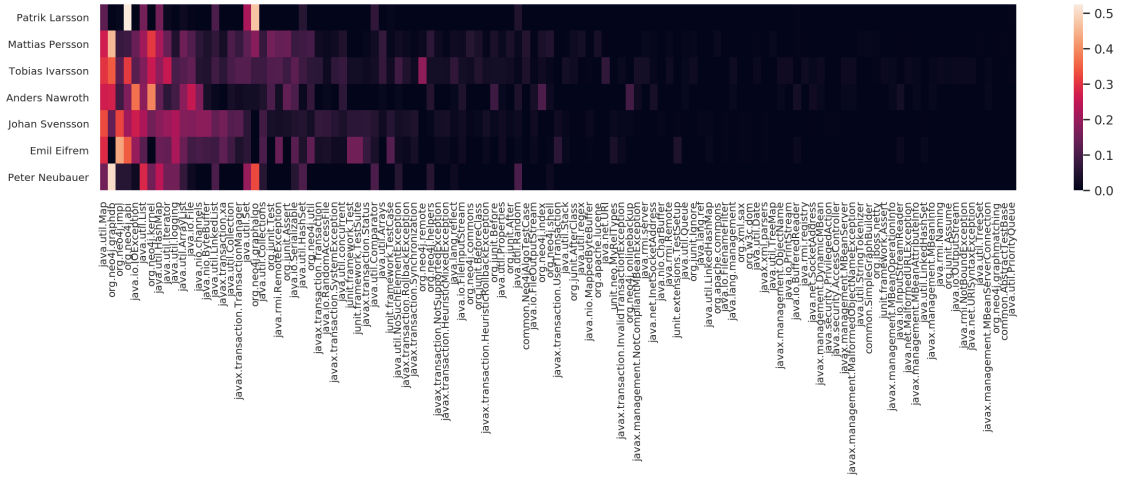

```
0.707367954956
```

```
/opt/conda/lib/python3.6/site-packages/matplotlib/figure.py:2117: UserWarning: This figure was
  warnings.warn("This figure was using constrained_layout==True, "
```

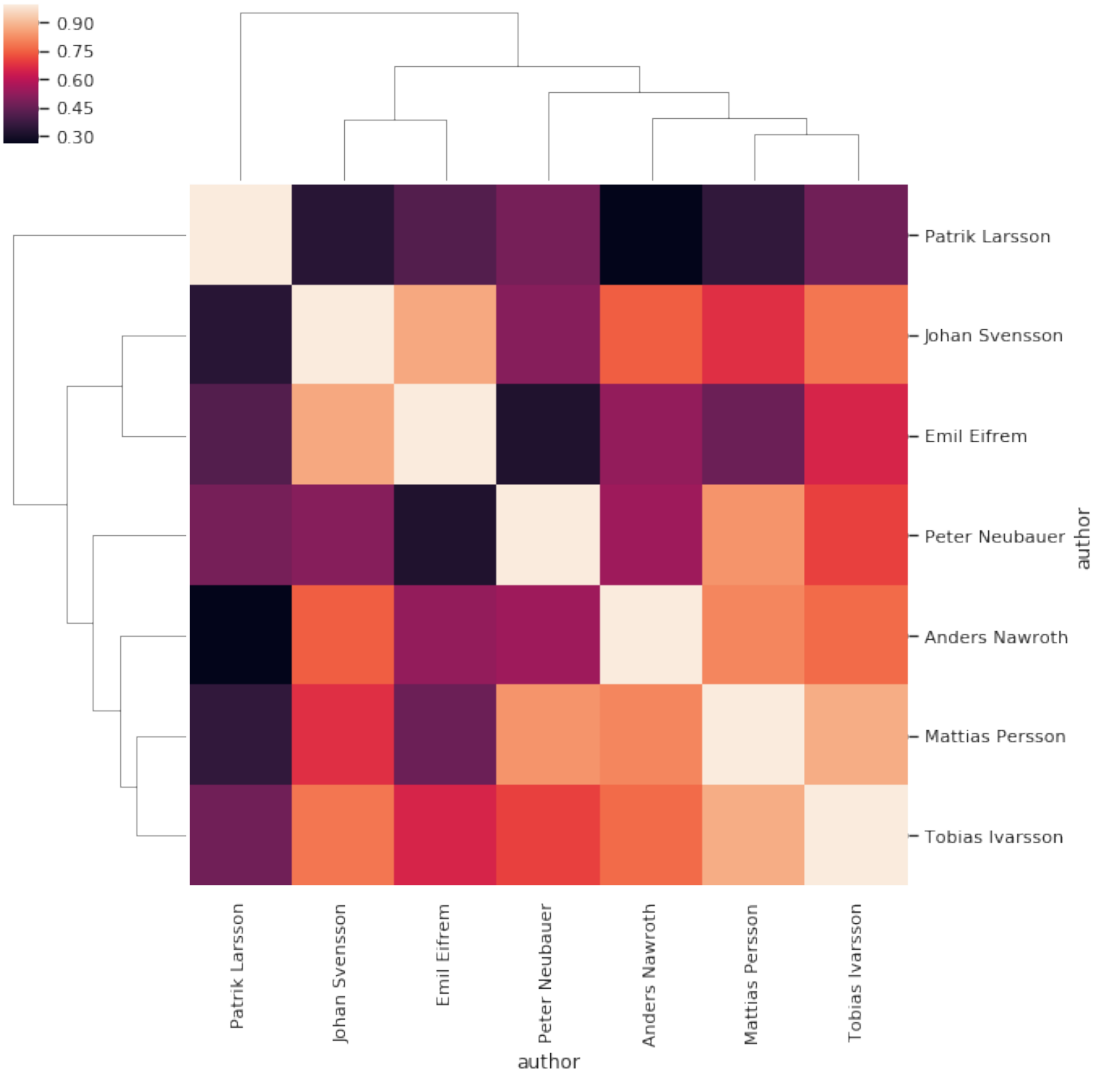

Emil Eifrem
Johan Svensson
Anders Nawroth
Tobias Ivarsson
Mattias Persson
Patrik Larsson
Peter Neubauer
PC1
PC2
PC3


Emil Eifrem
Johan Svensson
Anders Nawroth
Tobias Ivarsson
Mattias Persson
Patrik Larsson
Peter Neubauer
mean
0.0
400.0
800.0
1200.0
1600.0
0
300
600
900
1
2
3
principal component 1
principal component 2

3.2.2 Ordered by clusters

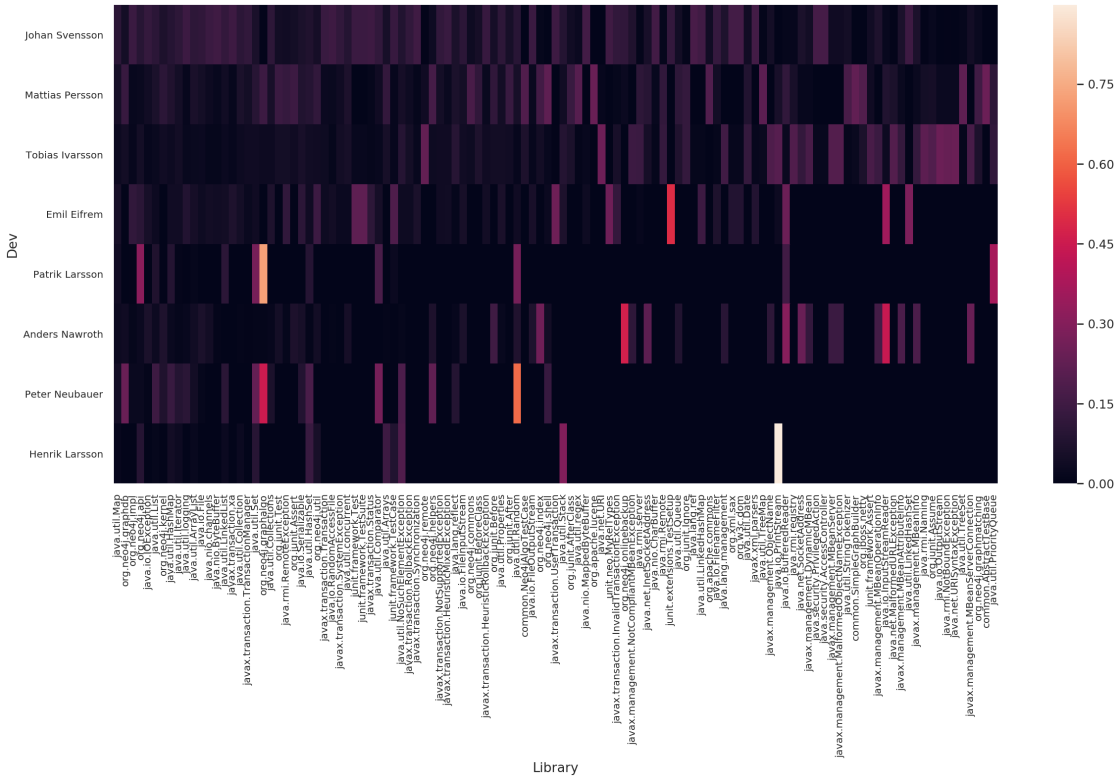


3.2.3 Scalar product (similarity of work)

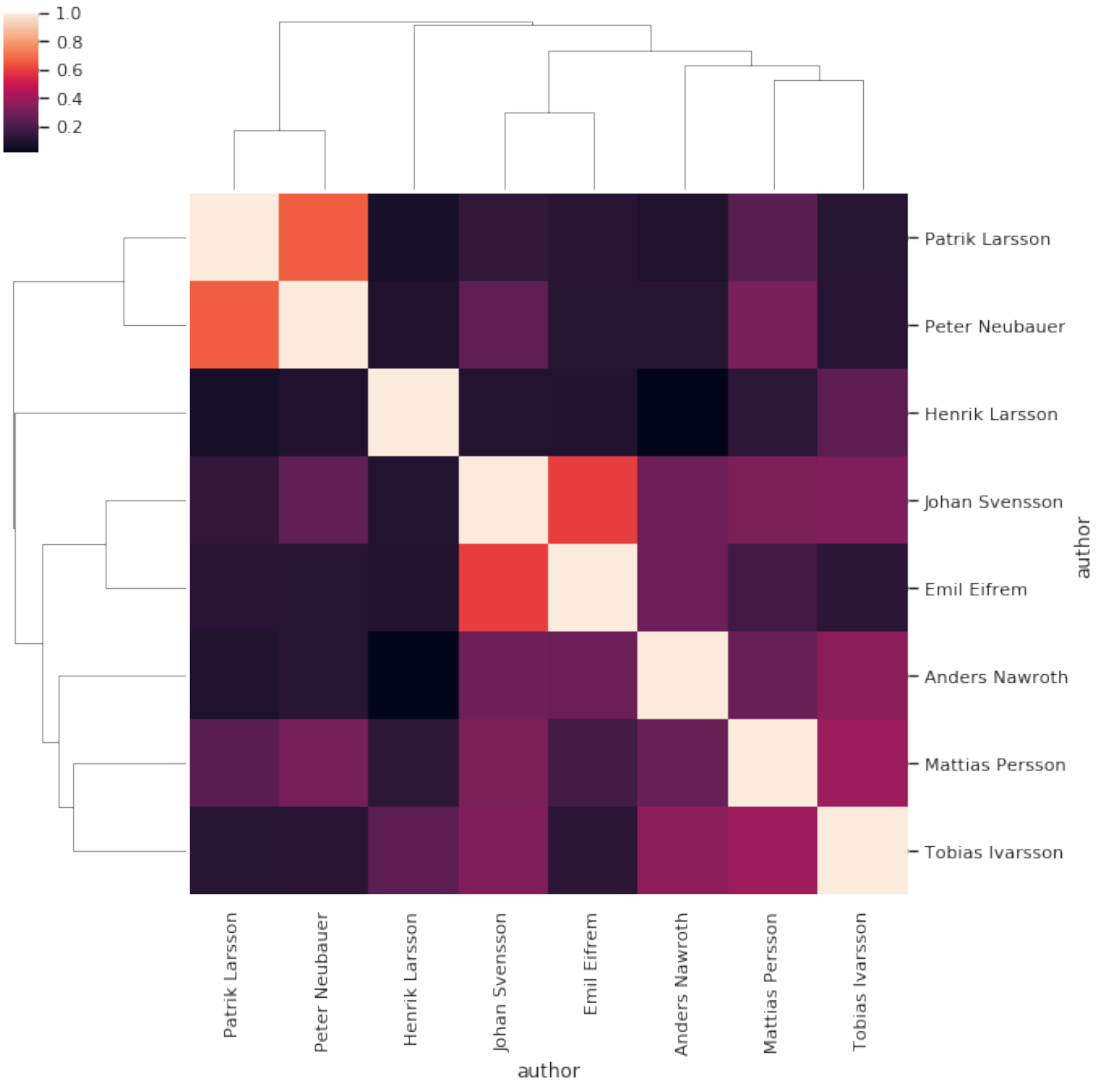


3.3 Normalisation expertise

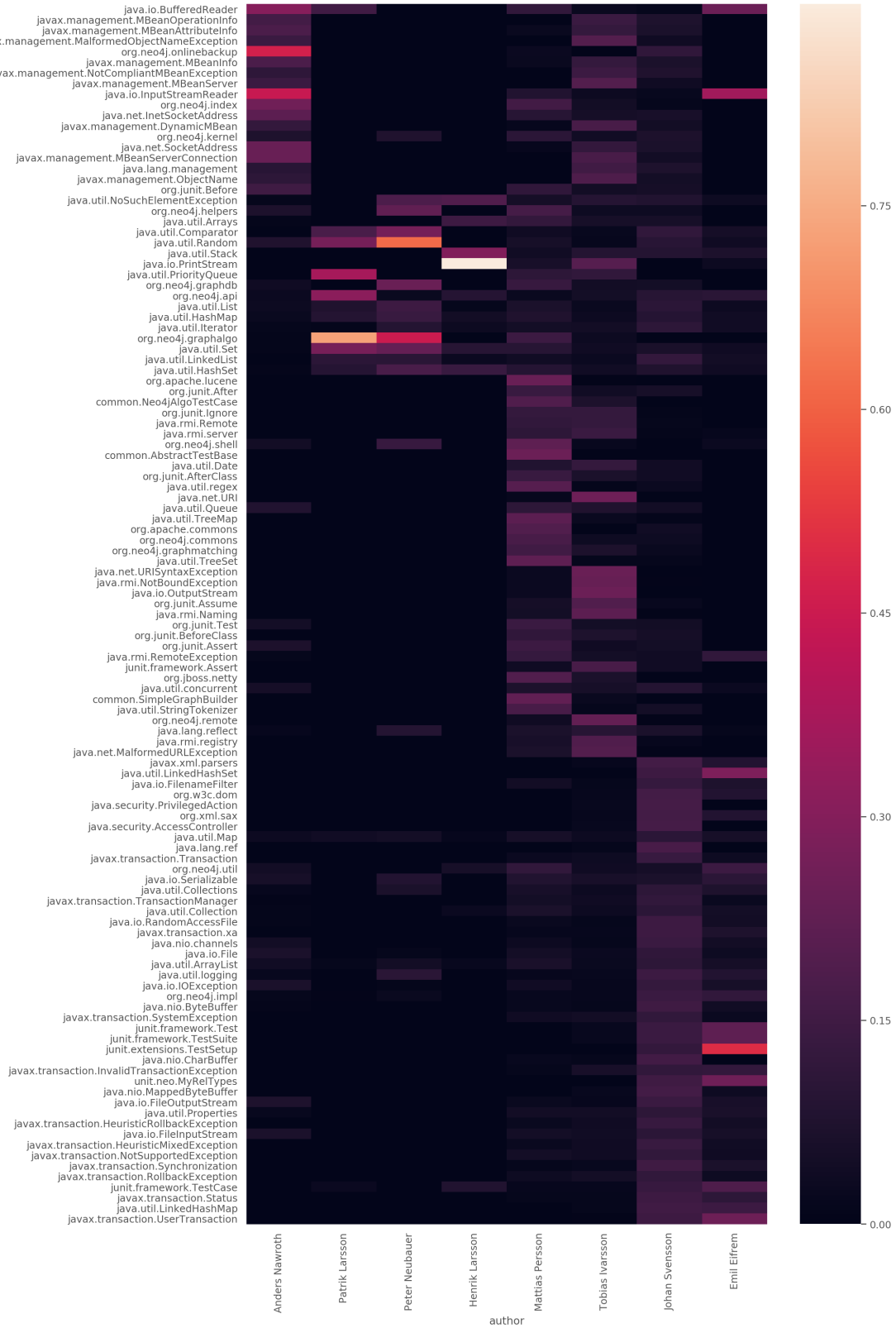
First we divide by the total number of imports per library and then we normalize (L2) on the devs. The goal is to have a vector that would represent it's expertise.


Dev
Johan Svensson
Mattias Persson
Tobias Ivarsson
Emil Eifrem
Patrik Larsson
Anders Nawroth
Peter Neubauer
Henrik Larsson
0.75
0.60
0.45
0.30
0.15
0.00
Library

3.3.1 Scalar product (similarity of expertise)



3.3.2 Co-Clust


java.io.BufferedReader
javax.management.MBeanOperationInfo
javax.management.MBeanAttributeInfo
javax.management.MalformedObjectNameException
org.neo4j.onlinebackup
javax.management.MBeanInfo
javax.management.NotCompliantMBeanException
javax.management.MBeanServer
java.io.InputStreamReader
org.neo4j.index
java.net.InetSocketAddress
javax.management.DynamicMBean
org.neo4j.kernel
java.net.SocketAddress
javax.management.MBeanServerConnection
java.lang.management
javax.management.ObjectName
org.junit.Before
java.util.NoSuchElementException
org.neo4j.helpers
java.util.Arrays
java.util.Comparator
java.util.Random
java.util.Stack
java.io.PrintStream
java.util.PriorityQueue
org.neo4j.graphdb
org.neo4j.api
java.util.List
java.util.HashMap
java.util.Iterator
org.neo4j.graphalgo
java.util.Set
java.util.LinkedList
java.util.HashSet
org.apache.lucene
org.junit.After
common.Neo4jAlgoTestCase
org.junit.Ignore
java.rmi.Remote
java.rmi.server
org.neo4j.shell
common.AbstractTestBase
java.util.Date
org.junit.AfterClass
java.util.regex
java.net.URI
java.util.Queue
java.util.TreeMap
org.apache.commons
org.neo4j.commons
org.neo4j.graphmatching
java.util.TreeSet
java.net.URISyntaxException
java.rmi.NotBoundException
java.io.OutputStream
org.junit.Assume
java.rmi.Naming
org.junit.Test
org.junit.BeforeClass
org.junit.Assert
java.rmi.RemoteException
junit.framework.Assert
org.jboss.netty
java.util.concurrent
common.SimpleGraphBuilder
java.util.StringTokenizer
org.neo4j.remote
java.lang.reflect
java.rmi.registry
java.net.MalformedURLException
javax.xml.parsers
java.util.LinkedHashSet
java.io.FilenameFilter
org.w3c.dom
java.security.PrivilegedAction
org.xml.sax
java.security.AccessController
java.util.Map
java.lang.ref
javax.transaction.Transaction
org.neo4j.util
java.io.Serializable
java.util.Collections
javax.transaction.TransactionManager
java.util.Collection
java.io.RandomAccessFile
javax.transaction.xa
java.nio.channels
java.io.File
java.util.ArrayList
java.util.logging
java.io.IOException
org.neo4j.impl
java.nio.ByteBuffer
javax.transaction.SystemException
junit.framework.Test
junit.framework.TestSuite
junit.extensions.TestSetup
java.nio.CharBuffer
javax.transaction.InvalidTransactionException
unit.neo.MyRelTypes
java.nio.MappedByteBuffer
java.io.FileOutputStream
java.util.Properties
javax.transaction.HeuristicRollbackException
java.io.FileInputStream
javax.transaction.HeuristicMixedException
javax.transaction.NotSupportedException
javax.transaction.Synchronization
javax.transaction.RollbackException
junit.framework.TestCase
javax.transaction.Status
java.util.LinkedHashMap
javax.transaction.UserTransaction
0.75
0.60
0.45
0.30
0.15
0.00
Anders Nawroth
Patrik Larsson
Peter Neubauer
Henrik Larsson
Mattias Persson
Tobias Ivarsson
Johan Svensson
Emil Eifrem
author