report

June 3, 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
```

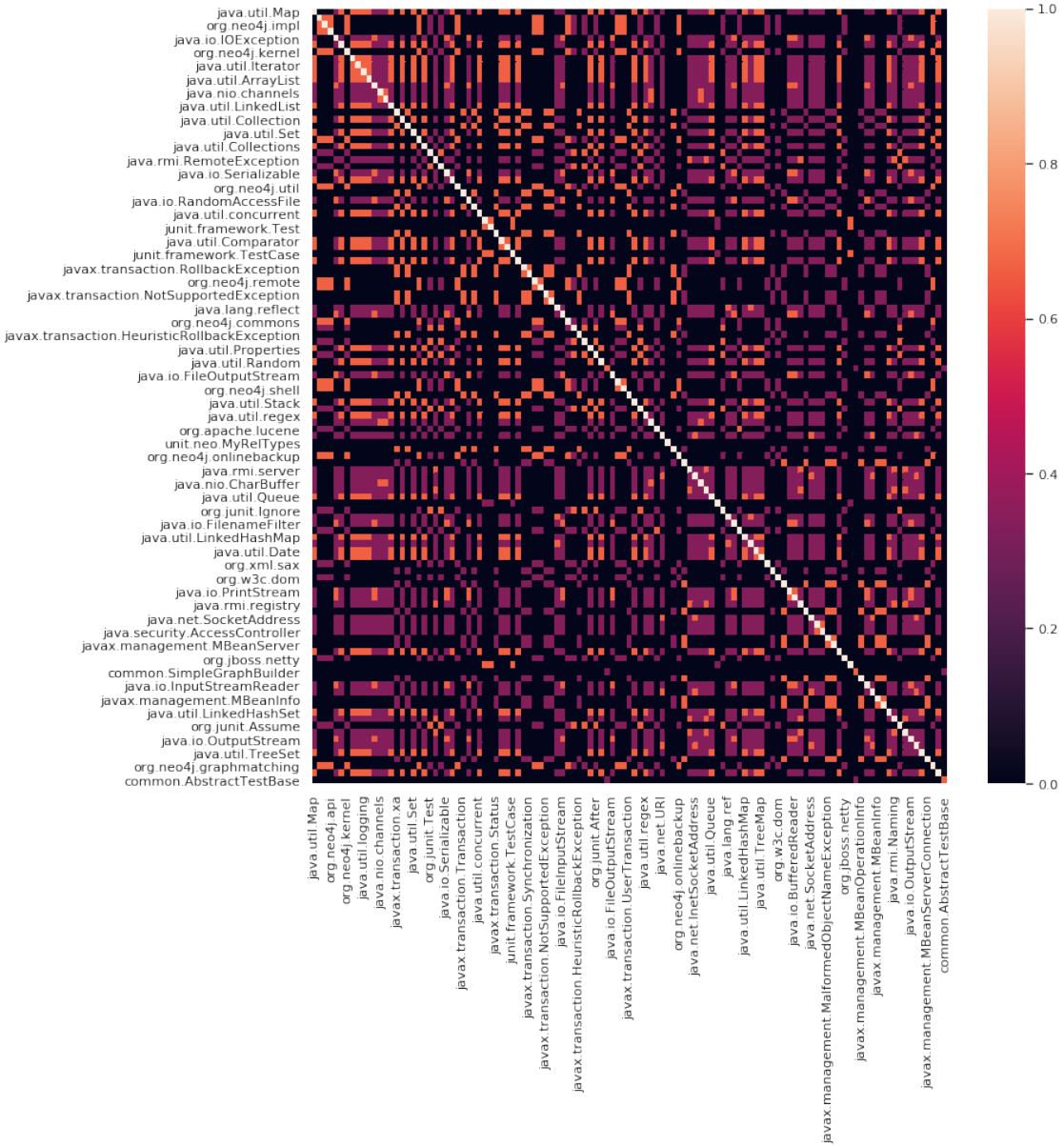
Henrik Larsson	398.000000	398	398	1
Johan Svensson	746.275904	40	1989	830
Mattias Persson	1235.443843	6	1992	739

Out[18]:

	first	last	count	coef
javax.transaction.InvalidTransactionException	1	1411	117	0.416418
java.lang.annotation	612	1823	18	0.774223
javax.transaction.Status	1	1910	666	0.882552
java.io.FilenameFilter	1	1859	66	0.817335
javax.transaction.Transaction	1	1984	858	0.986544

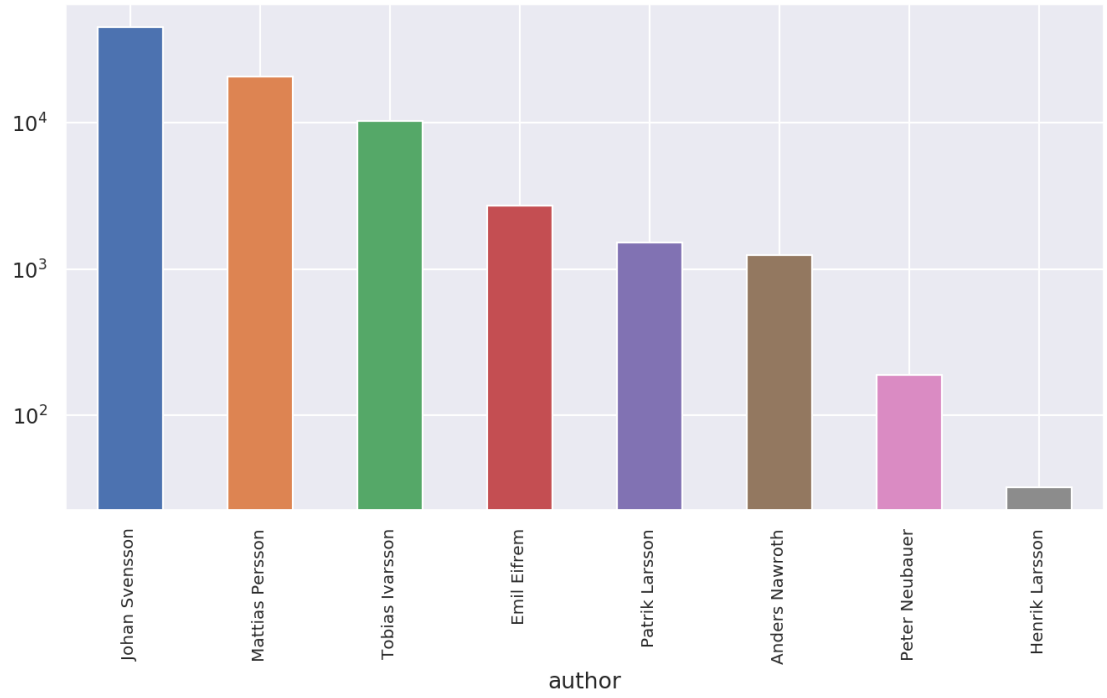
2.1 Distance between libraries

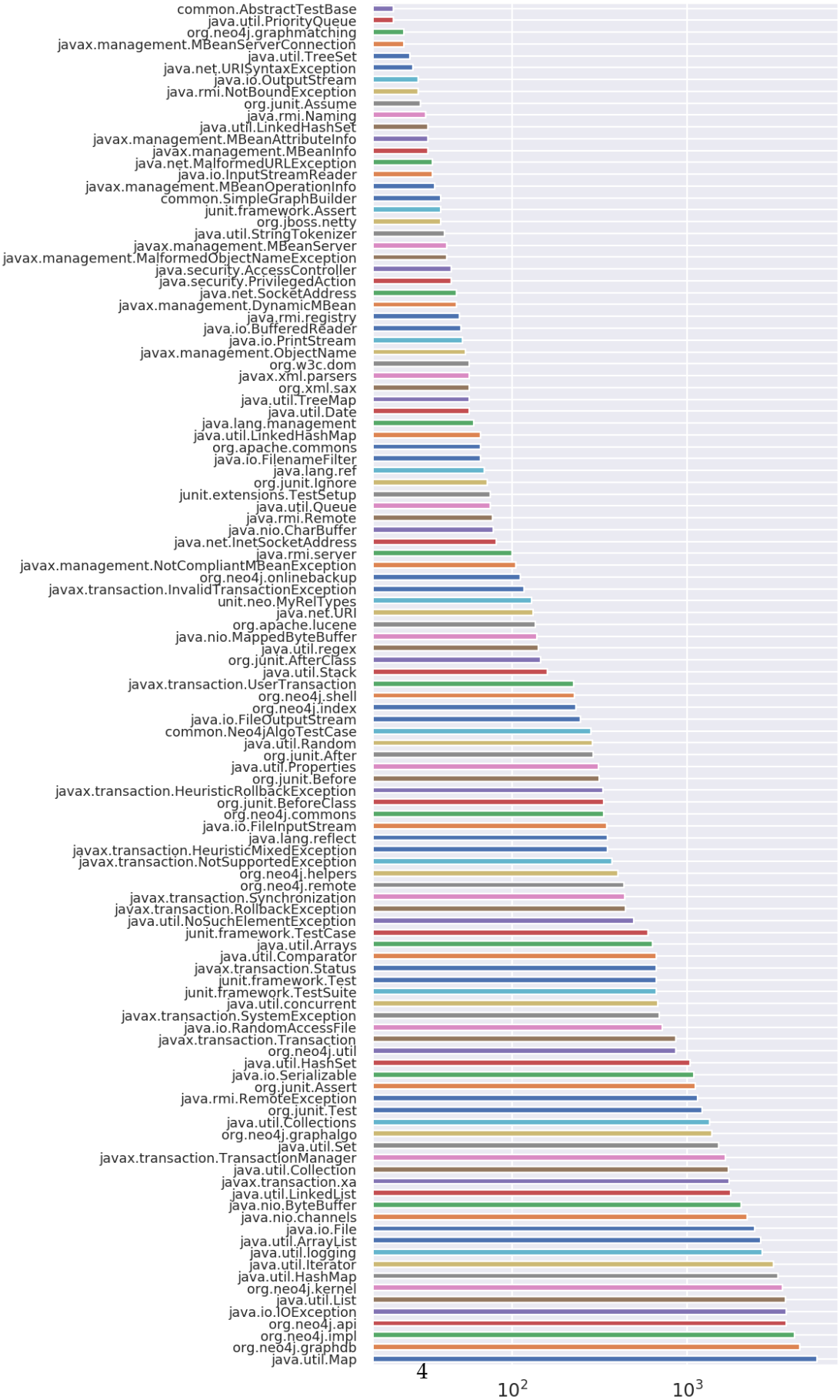
Out[19]: []

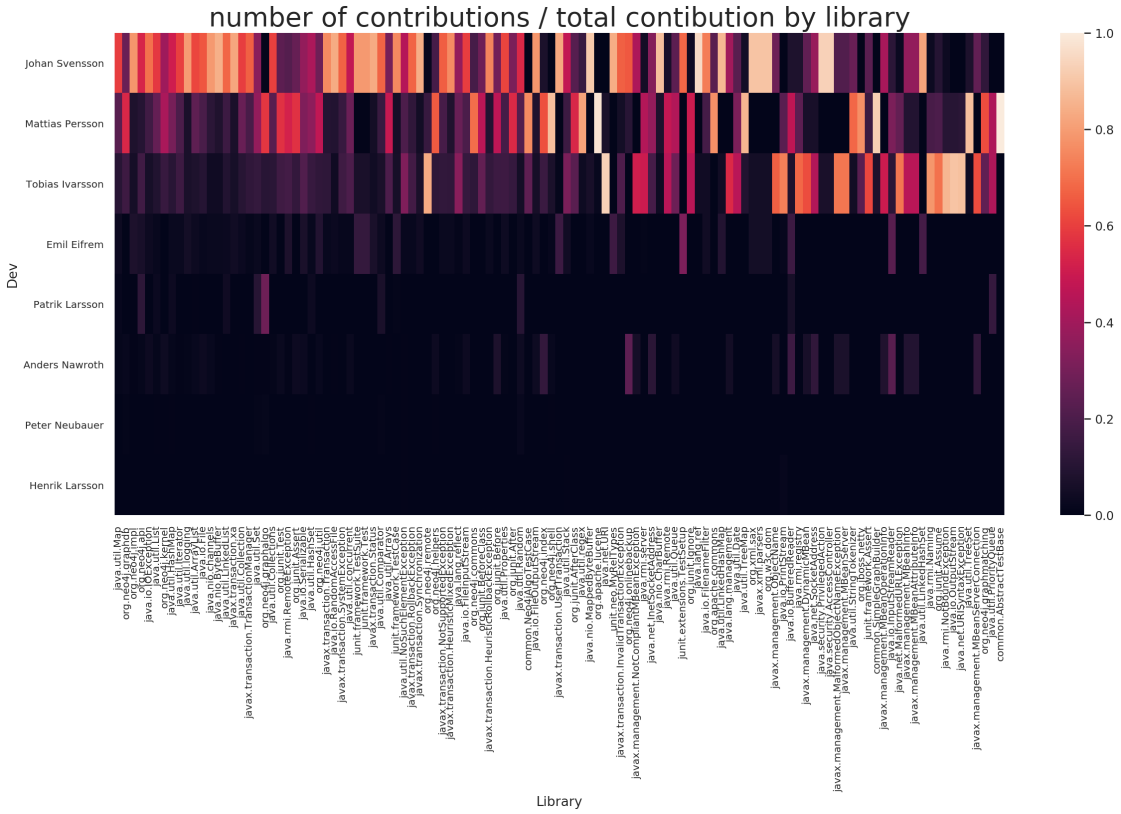


3 Analysis

3.1 Summary

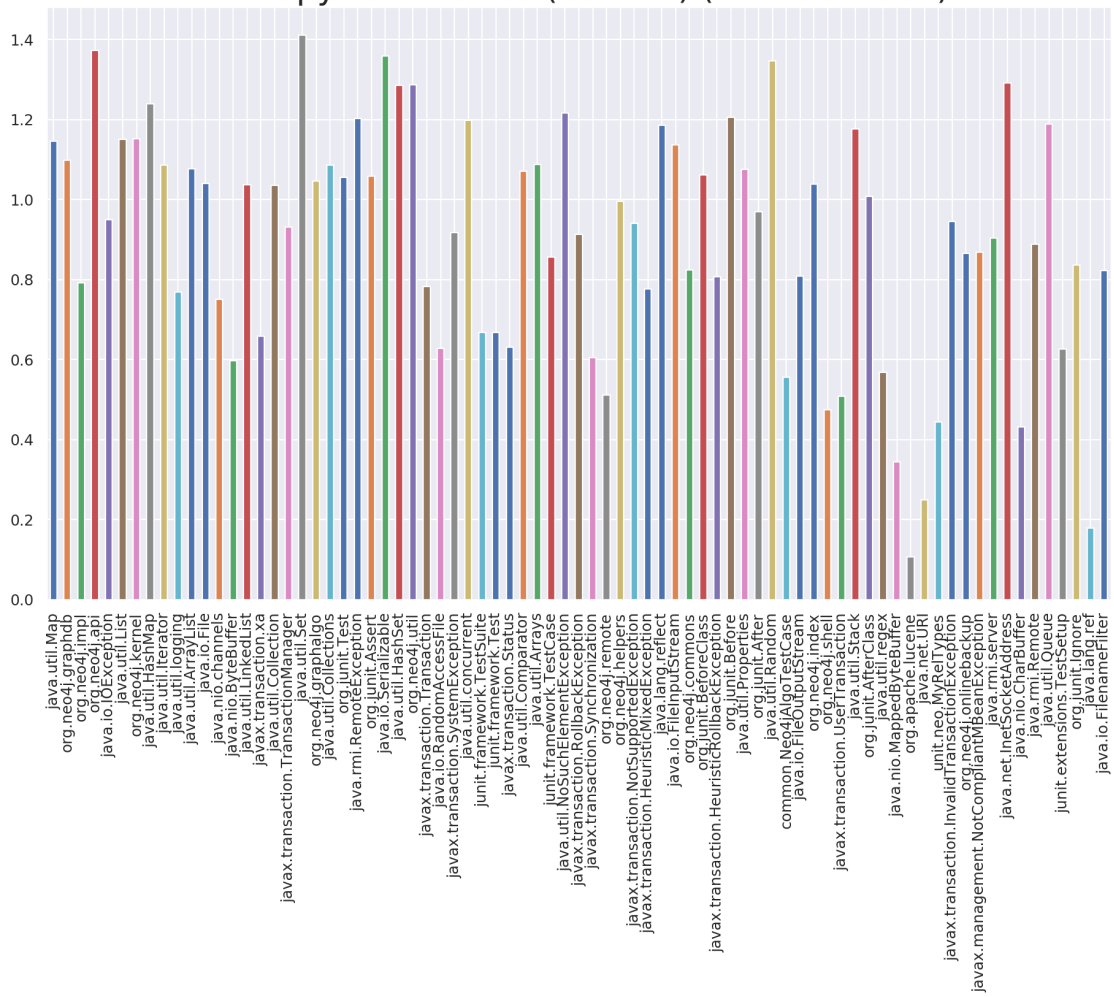



common.AbstractTestBase
java.util.PriorityQueue
org.neo4j.graphmatching
javax.management.MBeanServerConnection
java.util.TreeSet
java.net.URISyntaxException
java.io.OutputStream
java.rmi.NotBoundException
org.junit.Assume
java.rmi.Naming
java.util.LinkedHashSet
javax.management.MBeanAttributeInfo
javax.management.MBeanInfo
java.net.MalformedURLException
java.io.InputStreamReader
javax.management.MBeanOperationInfo
common.SimpleGraphBuilder
junit.framework.Assert
org.jboss.netty
java.util.StringTokenizer
javax.management.MBeanServer
javax.management.MalformedObjectNameException
java.security.AccessController
java.security.PrivilegedAction
java.net.SocketAddress
javax.management.DynamicMBean
java.rmi.registry
java.io.BufferedReader
java.io.PrintStream
javax.management.ObjectName
org.w3c.dom
javax.xml.parsers
org.xml.sax
java.util.TreeMap
java.util.Date
java.lang.management
java.util.LinkedHashMap
org.apache.commons
java.io.FilenameFilter
java.lang.ref
org.junit.Ignore
junit.extensions.TestSetup
java.util.Queue
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.Test
junit.framework.TestSuite
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



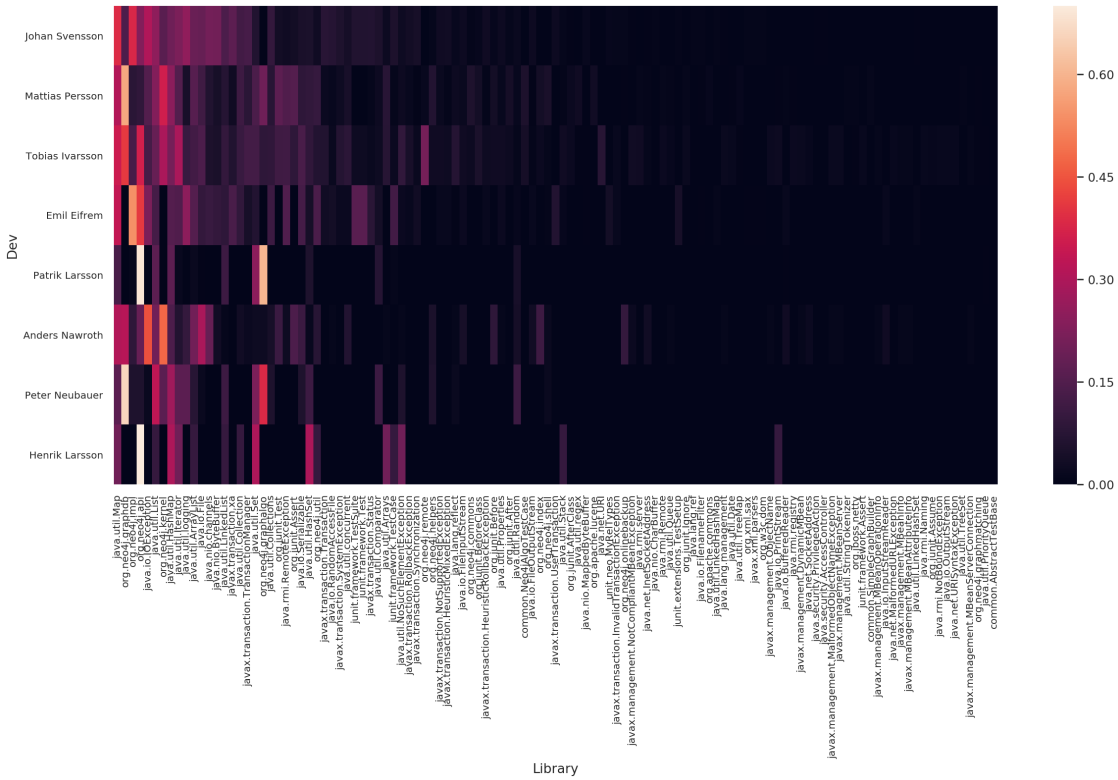
3.1.1 Entropy

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.TestSuite
junit.framework.Test
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
java.util.Queue
junit.extensions.TestSetup
org.junit.Ignore
java.lang.Error
java.io.FilenameFilter

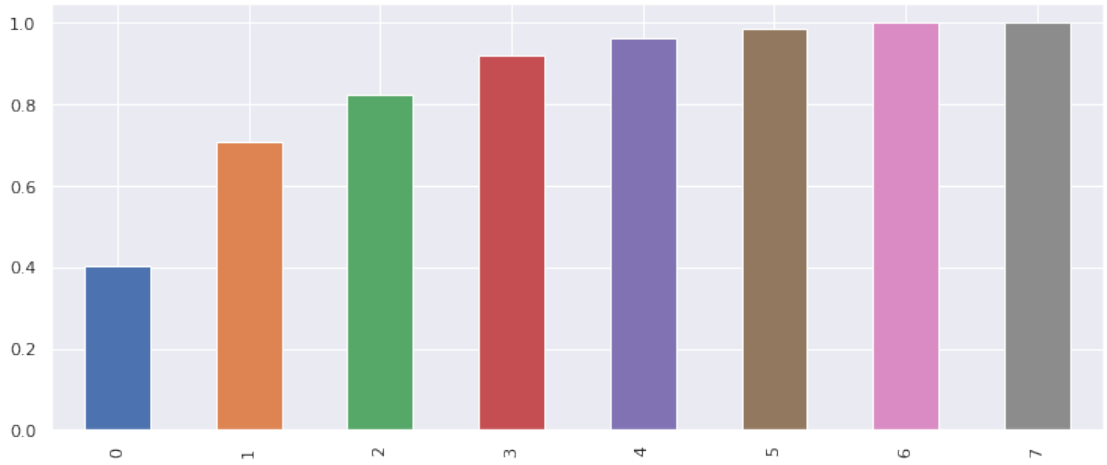
3.2 Nomalisation L2

This representation aims to represent the developpers contributions.



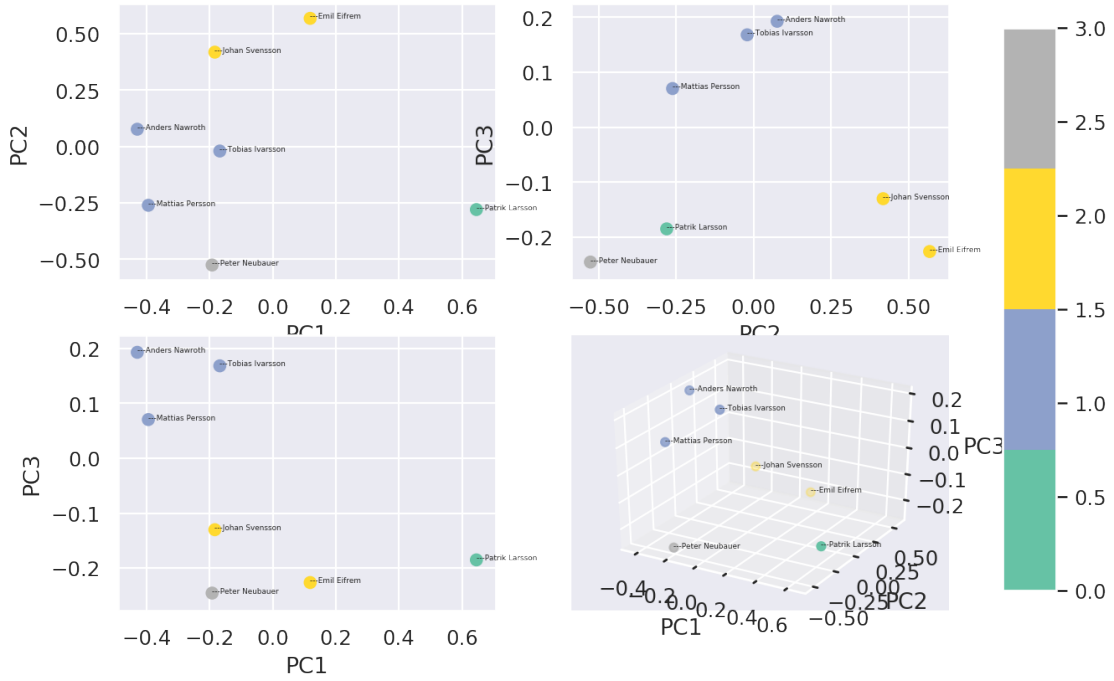
3.2.1 PCA

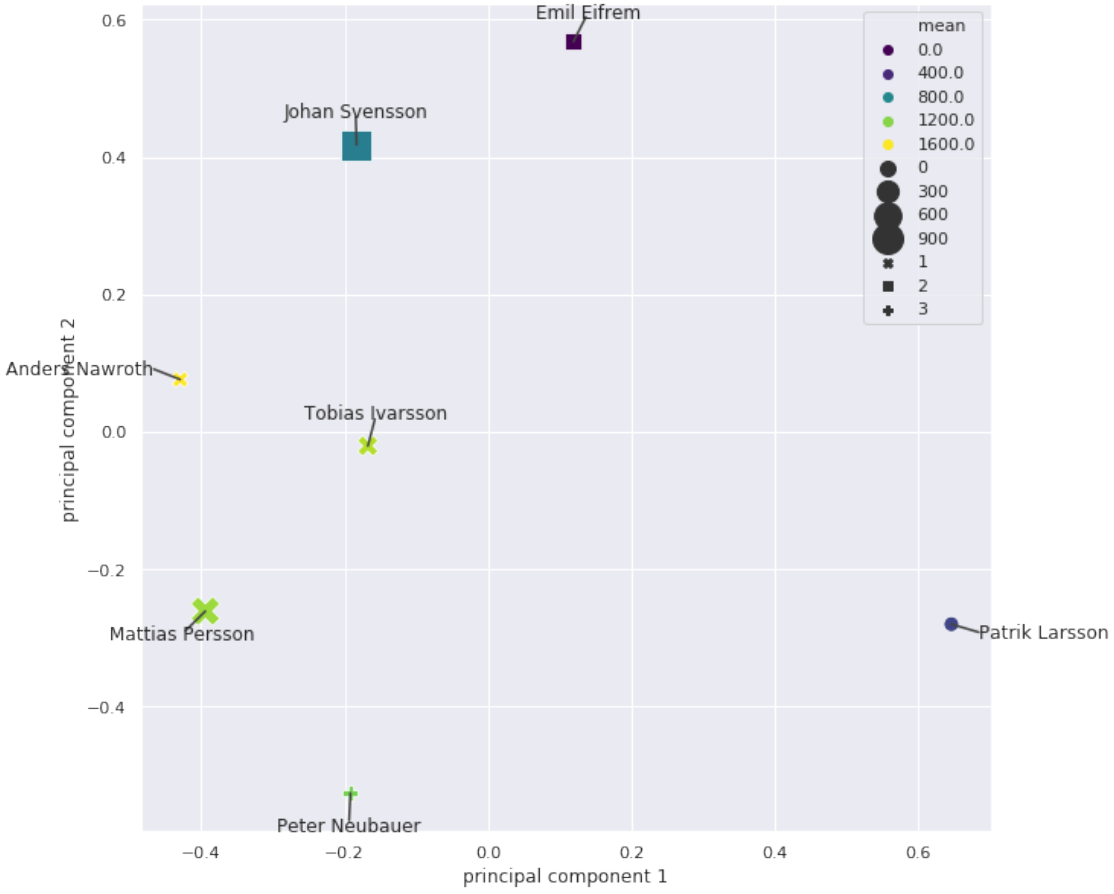
Cumulative sum of the explained variance



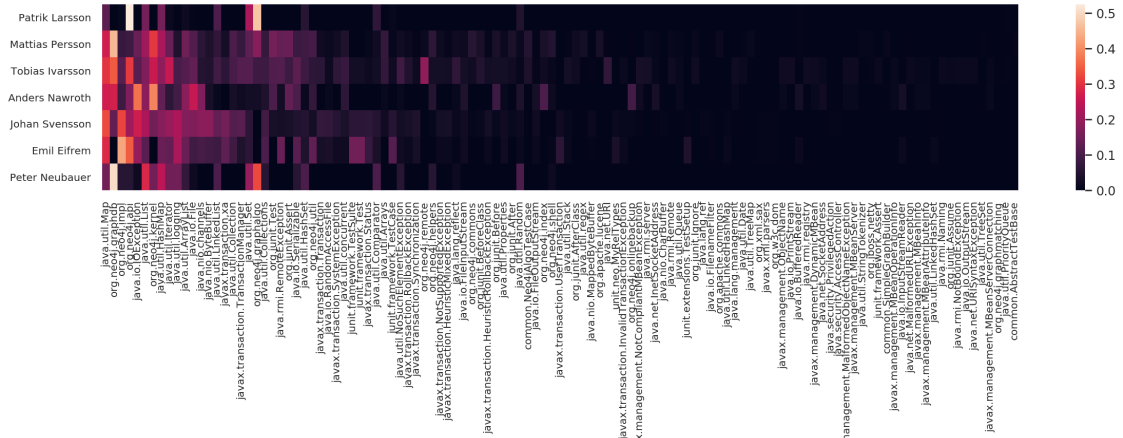

```
0.707367954956
```

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

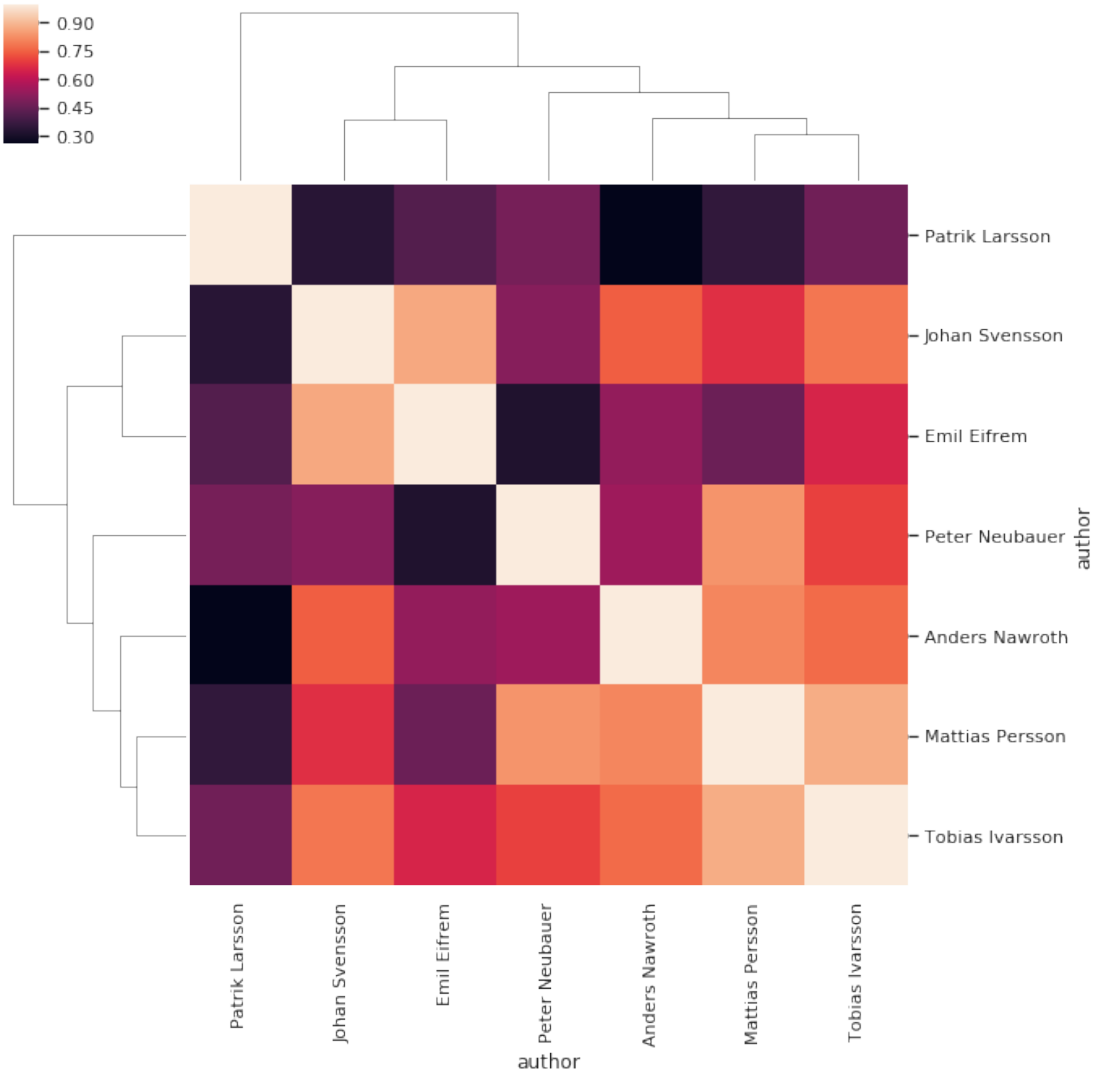

Emil Eifrem
Johan Svensson
Anders Nawroth
Tobias Ivarsson
Mattias Persson
Patrik Larsson
Peter Neubauer
PC1
PC2
PC3


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

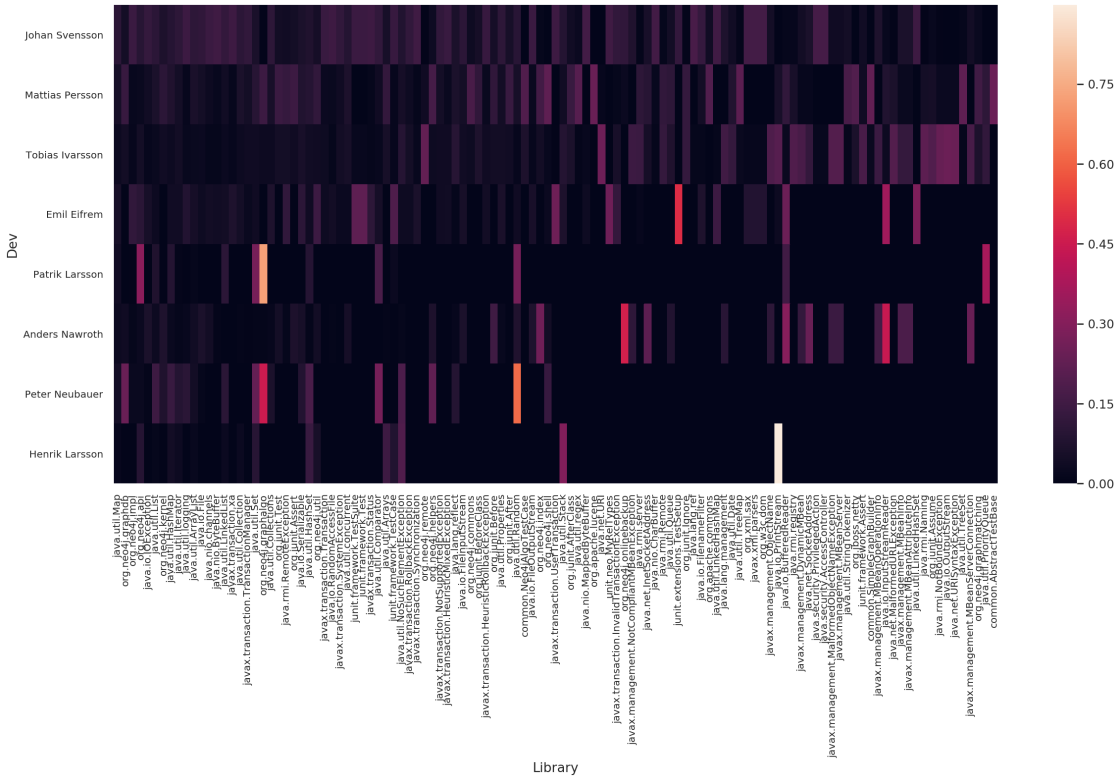


3.2.3 Scalar product (similarity of work)

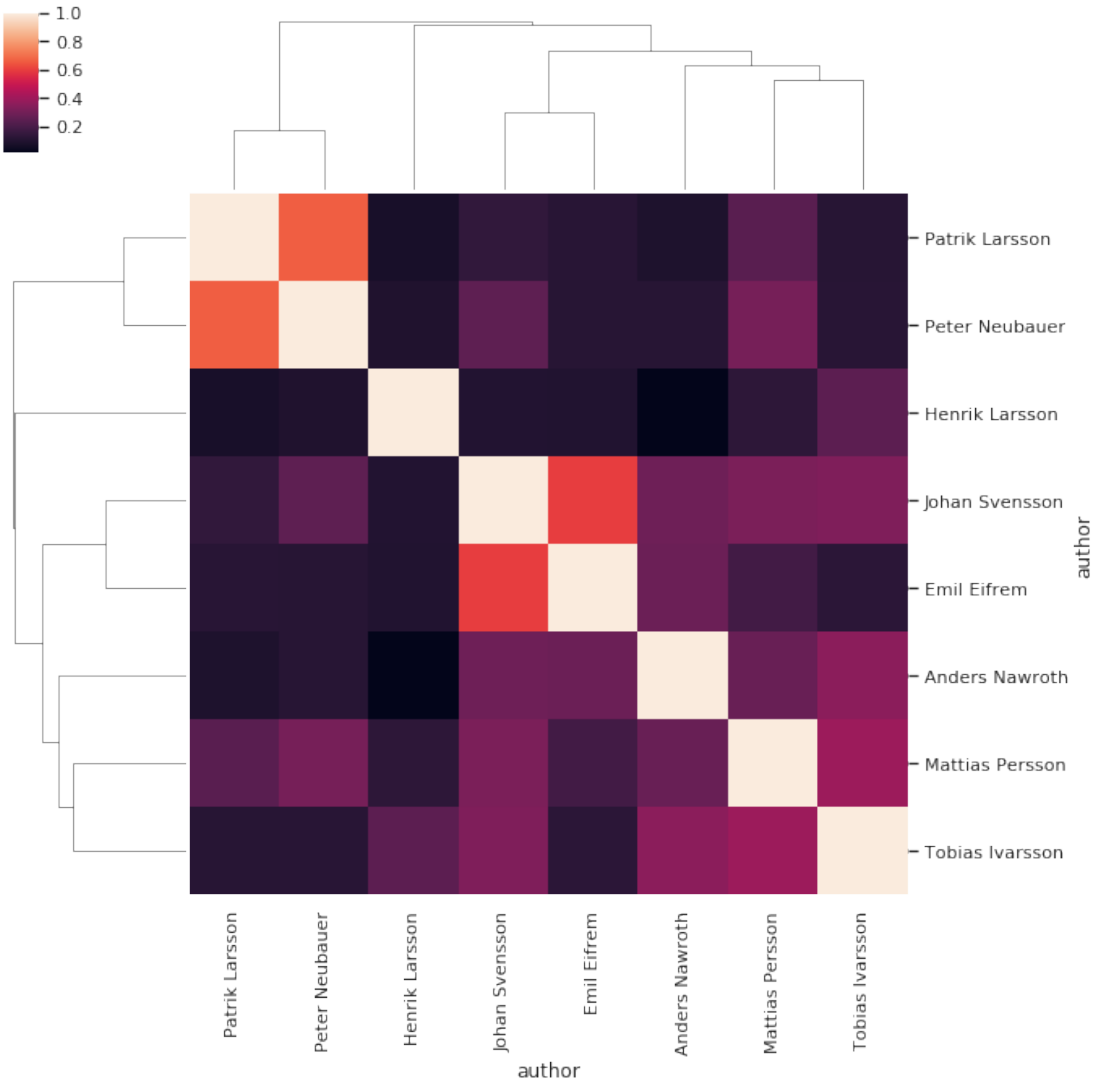


3.3 Normalisation expertise

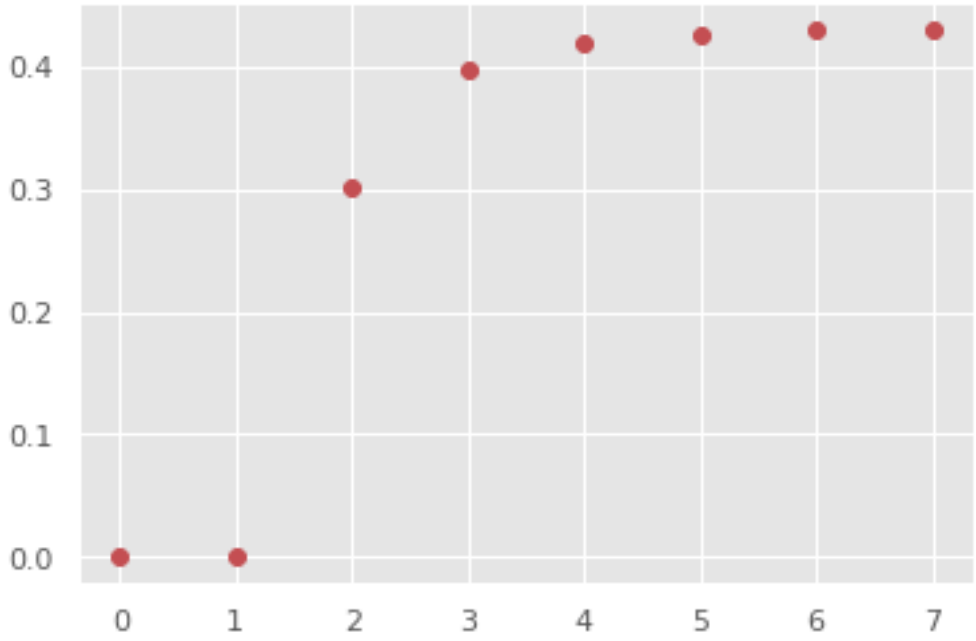
First we divide by the total number of imports per library and then we normalize (L2) on the devs. The goal is to have a vector that would represent it's expertise.


Johan Svensson
Mattias Persson
Tobias Ivarsson
Emil Eifrem
Patrik Larsson
Anders Nawroth
Peter Neubauer
Henrik Larsson
Dev
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




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