

LABORATORY 4

1. Topic Modelling und Analysis

```
> ldaOut.terms <- as.matrix(terms(ldaOut,6))
> ldaOut.terms
      Topic 1      Topic 2
[1,] "use"      "work"
[2,] "softwar"  "licens"
[3,] "copyright" "deriv"
[4,] "provid"   "shall"
[5,] "notic"    "sourc"
[6,] "follow"   "notic"
```

Abbildung 1: The six significant terms in two Topics

```
      Topic 1      Topic 2      Topic 3
[1,] "use"      "work"      "copyright"
[2,] "softwar"  "licens"  "notic"
[3,] "disclaim" "deriv"   "provid"
[4,] "redistribut" "form"   "condit"
[5,] "damag"    "shall"   "includ"
[6,] "liabil_"  "mean"    "contributor"
```

Abbildung 2: The six significant terms in three Topics

```
      Topic 1      Topic 2      Topic 3      Topic 4
[1,] "softwar"  "condit"   "work"      "use"
[2,] "provid"   "licens"   "deriv"     "copyright"
[3,] "disclaim" "notic"    "licens"    "includ"
[4,] "notic"    "file"     "shall"     "follow"
[5,] "damag"    "distribut" "sourc"     "redistribut"
[6,] "must"     "otherwis" "form"      "warranti"
```

Abbildung 3: The six significant terms in four Topics

#	Topic Labels	Topic Description	Example	Topic Proportion
Abbildung 1 and 2	Topic 1	Software use	IN NO EVENT SHALL ISC BE LIABLE FOR ANY SPECIAL, DIRECT, INDIRECT, OR CONSEQUENTIAL DAMAGES OR ANY DAMAGES WHATSOEVER RESULTING FROM LOSS OF USE , DATA OR PROFITS, WHETHER IN AN ACTION OF CONTRACT, NEGLIGENCE OR OTHER TORTIOUS ACTION, ARISING OUT OF OR IN CONNECTION WITH THE USE OR PERFORMANCE OF THIS SOFTWARE	20%
Abbildung 1 ,2 and 3	Topic 1	Software Provider	THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED.	15%
Abbildung 1 and 2	Topic 2	Derivative Works	For the purposes of this License, Derivative Works shall not include works that remain separable from, or merely link (or bind by name) to the interfaces of, the Work and Derivative Works thereof.	25%
Abbildung 2 ,3 and 4	Topics 2,3	License distribution	Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License .	25%
Abbildung 2 and 3	Topics 3,4	Copyright	THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.	20%
Abbildung 2 and 3	Topics 1,4	Redistribution use	Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met: * Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer. * Redistributions in binary form must reproduce the above	25%

Tabelle 1: Analysis for Topics

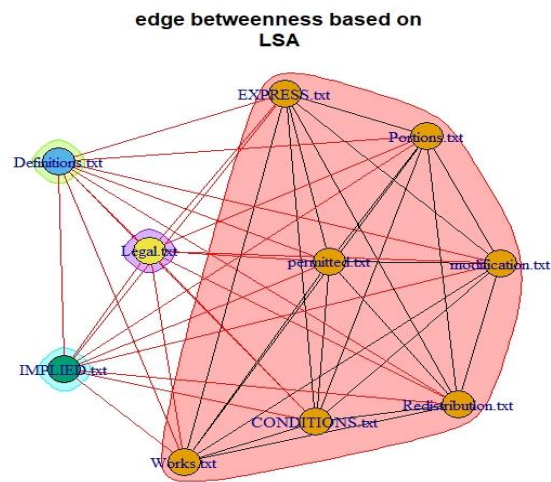


Abbildung 7: Edge betweenness based on LSA

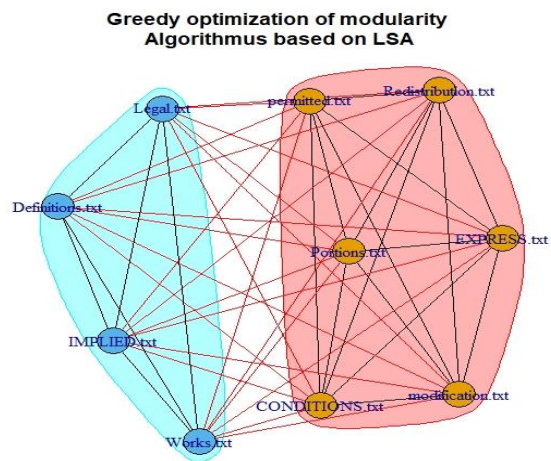
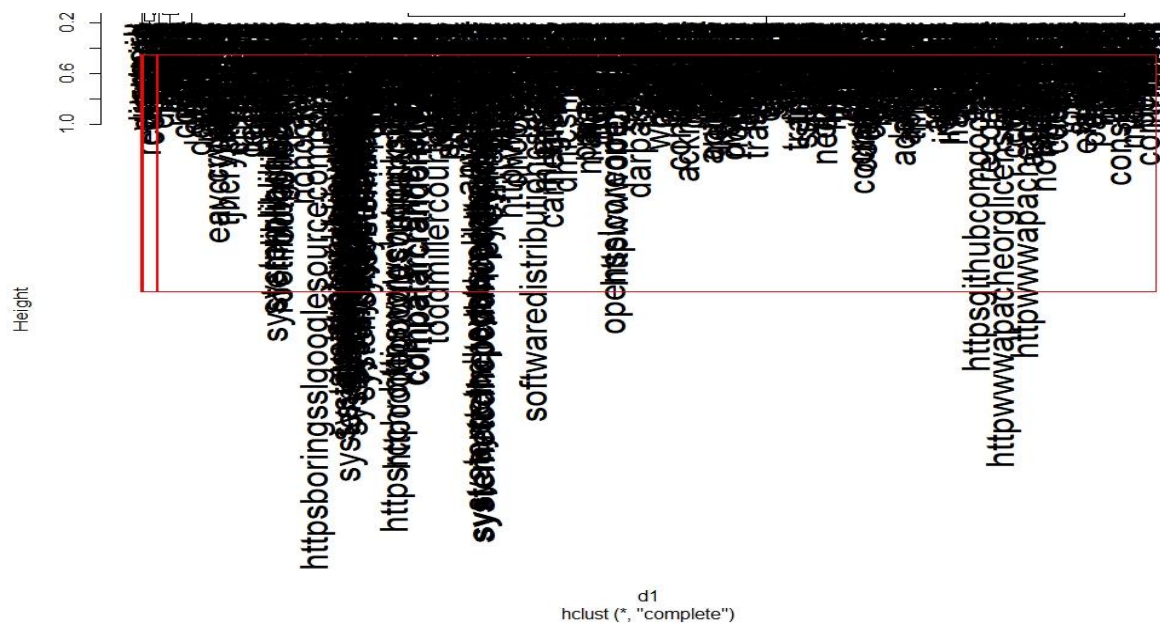


Abbildung 6: Greedy Optimization based on LSA

Abbildung 8: Propagating Labels based on LSA



Method	Topic/Clusters	Source	Topic1/Cluster 1: Software use	Topic 2/Cluster2: Software Provider	Topic 3/Cluster3: Copyright	Opinion/ Comments	Rating
Topic Modelling	4	Dtm	Abbildung 1,2	Abbildung 1 ,2 and 3.	Abbildung 2 and 3	Good result, difficult to interpret topics	2
k-means	3	LSA			Abbildung 5	Good result, difficult to identified	3
Plot Latent Semantic Space		Dtn_Norm	Abbildung 4			Very Good result, little bit difficult to interpret for too many daten	1.5
Hierachical Clustering	2	Dtn_Norm		Abbildung 9		Bad result, difficult to interpret	4
Graph-based algorithms		LSAspace	Abbildung 6,7 and 8			Very Good result, easy to interprete	1

Tabelle 2: Interpretation und Rating for Labels