Network Text Analysis of R Mailing Lists
UseR! Rennes 2009

Angela Bohn, Ingo Feinerer, Kurt Hornik, Patrick Mair, Stefan Theußl

7/10/2009
A mailing list social network

R-help mailing list:
Jan 2008 to May 2009
Number of authors: 5326
Number of mails: 41457
Avg. degree: 4.4
Diameter: 7

Legend:

Author A answered Author B

Vertex' distance to nearest center:
- zero
- one
- two
- three
- four

Vertex size indicates number of e-mails sent (not proportionally).
Number of polygon edges of vertex border corresponds to vertex degree.
Combine SNA and TM

- **Goal:** Combine social network analysis (SNA) and text mining (TM) to find out more
- **Data:** Mailing lists R-help and R-devel
- **Packages:** `sna` and `tm`
- **Results:**
  - “Interest maps” of R users
  - Detection of bottlenecks in communication
Data preparation for social network analysis

Create a social network from e-mail headers (tm):

From: dwinsemius at comcast.net (David Winsemius)
Date: Thu, 30 Apr 2009 18:49:55 -0400
Subject: [R] Extracting Element from S4 objects
In-Reply-To: <23302265.post@talk.nabble.com>
Message-ID: <A6039F4E-ABF4-41C5-B03E-FFF32E07C37A@comcast.net>

Author A: "Hallo, I have a question."
Author B: "This is the answer."
Author C: "This, too."
Author D: "And this."
Author D: "And this."
Author A: "Thank you."

Find aliases:
knoblauch at lyon.inserm.fr (knoblauch)
knoblauch at lyon.inserm.fr (Ken Knoblauch)
ken.knoblauch at inserm.fr (Kenneth Knoblauch)
ken.knoblauch at inserm.fr (Ken Knoblauch)

Levensthein Distance:
agrep(base)
Data preparation for text mining

E-mail subjects:

[R] passing args from the command line
[R] navigating ggplot viewports
[R] how to go to a line in R

... Term Frequencies (termFreq(tm)): 

[... chart with terms like function, question, data, etc.]

R-help

R-devel
Word Networks

“lattice”
Word Networks

“legend”
Word Networks

“boxplot”
# Centrality Measures

<table>
<thead>
<tr>
<th>Notion</th>
<th>lattice</th>
<th>ggplot</th>
<th>legend</th>
<th>boxplot</th>
</tr>
</thead>
<tbody>
<tr>
<td>Most central persons</td>
<td>Deepayan Sarkar, Sundar Dorai Raj, baptiste auguie</td>
<td>hadley wickham, Thierry ONKELINX</td>
<td>Duncan Murdoch, hadley wickham, Greg Snow</td>
<td>Gabor Grothendieck, hadley wickham</td>
</tr>
</tbody>
</table>

![Graph image]

- Duncan Murdoch
- Greg Snow
- legend
- hadley wickham
- Thierry ONKELINX
- Deepayan Sarkar
- boxplot
- Sundar Dorai Raj
- Gabor Grothendieck
- baptiste auguie

![Graph image]
Results: Interest maps
Results: Communication bottlenecks

Good.

Can be improved.
Thank you!

Packages:


References:


Contact:

angela.bohn@gmail.com, www.angela-bohn.de