Interaction Networks in ML: a socio-meter for Debian

Alexandre Delanoë
Bernard Conein

Mini DebConf, 25 November 2012

Subject : "Making debian-* discussions more viable"

Russ Allbery changed the topic "From exim to postfix"

- DPL (Zack, 3 May 2012) :
  1. "don't feed the troll"
  2. "don't feed the troll" + report abuses to listmasters and act accordingly
  3. public, but contributors-only list
- ⇒ A lot Energy to know what you have to ignore (or to create new rules)

Proposition : from antispam to antitroll

Existing solutions (mutt)

- ThreadIgnore.pl pipe
- Thread killer : fgrep Archive >/dev/null

Socio-meter as an *Agnostic* Thread Index

- Socio-meter to specify interactions in ML
- Anonymous statistics (!= teammetrics)
- Each one could filter threads (with procmail and notmuch)

Summary

1. Mailing list landscape
2. Threads as a group
3. Threads as a network

Debian Users : Membership/posts
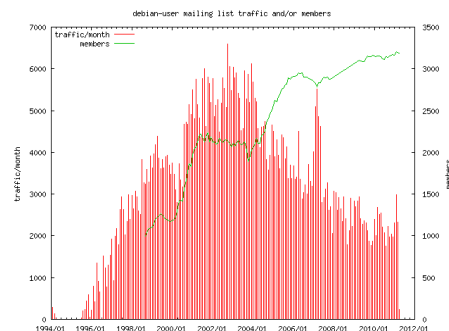
(a) Nb of mails and Members / year

Paradox
- Membership increase
- Post decrease

The first corpus

Wget pages archives

for distribution in Debian Ubuntu ; do

    for ml in users devel vote etc. ; do

        wget \$ml ;

done[2]

- www.debian.org/MailingLists/ html
- https ://lists.ubuntu.com/ mbox format


First Corpus, may 2011

Debian, mainly from 1994
- user (646 005 mails, 17 years, 104 mails/day average)
- user-french (178 844 mails, 10 years, 49 mails/day a.)
- devel (305 106 mails, 17 years, 49 mails/day a.)
- vote (17 266 mails, 12 years, 4 mails/day a.)
- mentors (58 778 mails, 13 years, 12 mails/day a.)

Ubuntu, mainly from 2004
- user (246 233 mails, 7 years, 96 mails/day a.)
- fr (37 073 mails, 7 years, 15 mails/day a.)
- devel (33 134 mails, 7 years, 13 mails/day a.)
- devel-discuss (12 676 mails, 5 years, 7 mails/day a.)

But
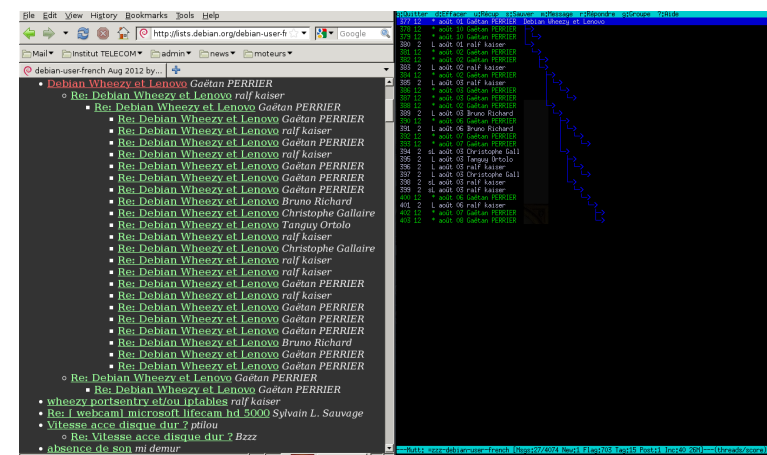
(b) Archives and reality

Thx gmane

Second corpus

- NNTPSERVER='news.gmane.org' && export NNTPSERVER
- nntp-pull gmane.linux.debian.user*

Mutt threads reader

- Mutt groups threads with Subject *AND* References headers
- /.muttrc : macro index,pager <f8> "<tag-thread><tag-prefix-cond><pipe-entry>cat » /tmp/threadXXXX<enter><untag-pattern> T n<pipe-entry> /.mutt/threads.sh<enter>"
- for thread in {1..$nbThread} ; do xdotool F8 ; sleep 3 ; done

Second Debian Corpus

- gmane.linux.debian.user, 423 275 mails
- gmane.linux.debian.user.french, 174 026 mails
- gmane.linux.debian.devel.general, 166 609 mails
- gmane.linux.debian.devel.legal, 33 617 mails
- gmane.linux.debian.devel.mentors, 54 178 mails
- gmane.linux.debian.devel.policy, 14 868 mails
- gmane.linux.debian.devel.project, 18 769 mails
- gmane.linux.debian.devel.vote, 15 552 mails

Debian and Ubuntu Users
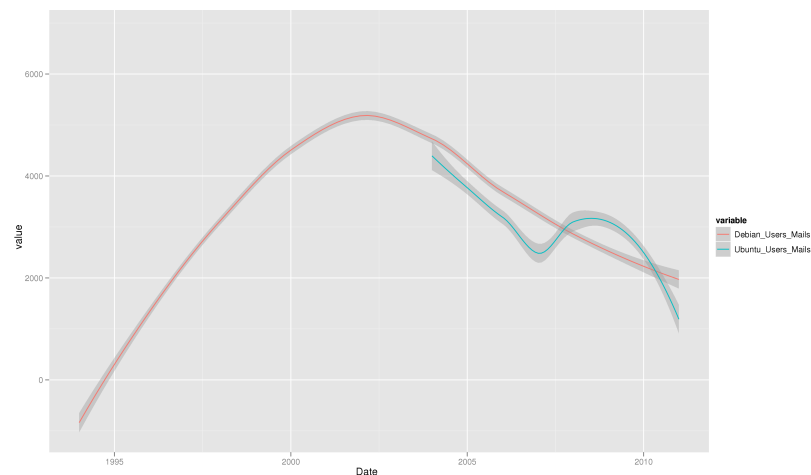
(c) Nb of mails / year
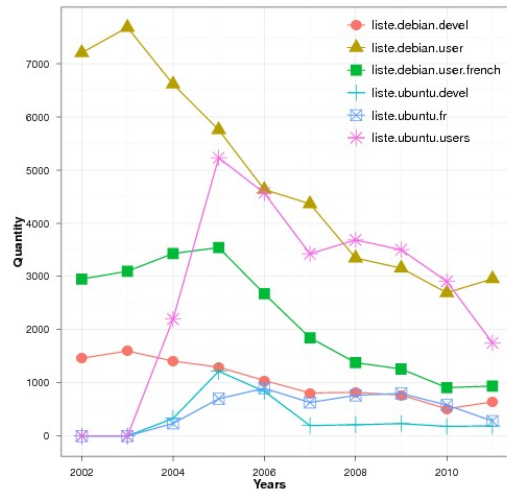
Debian and Ubuntu Users
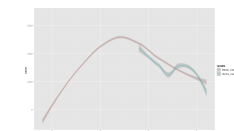
(d) Nb of mails / year

- Increase of mail's numbers
- Decrease of mail's numbers after Ubuntu's birth ?

- But the cause of the phenomenon is not clear

- It cannot be explained only by Ubuntu itself
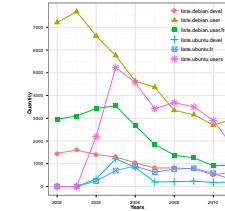- Other factors should be taken into account :
⇒ Less bugs, less requests, other tools, other areas for posting

(e) Nb of threads / year
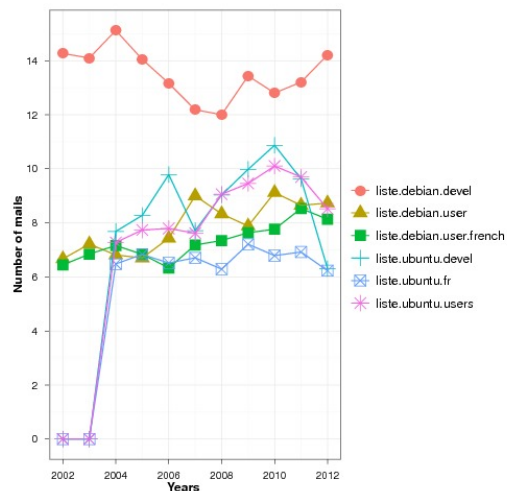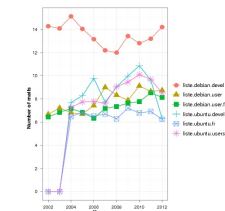
(f) Nb of threads / year

- Decrease of thread's number ?
- No clear explanation
- Less topics to be discussed
- But does it imply that there are less discussions ?
- NO

(g) Nb of mails / thread

(h) Nb of mails / thread

- Increase of mails number :
  - Continuous for Debian user
  - Discontinous for Debian Devel
- Size of thread counts ?
  - Maybe less topics but complex topics
  - Other explanation : the size is not enough as the thread structure might count

Joey Hess' Stuff, 30 october 2005 : it's all about pattern recognition

This is the "think before you post" pattern.

```
bob    +-> Foo
bob    +->
```

```
bob       +-> Foo?
```

```
bob    +-> Foo?
fred      +->
alice     +->
george    +->
nancy     +->
steven    +->
nick      +->
```

(i) Think before!   (j) Uninteresting   (k) Obvious Answer

(l) Take it Private   (m) Worthwhile   (n) Too late

Joey Hess explains how to subscribe to mailing lists with a combined total posts of 2000 or more per day, and live. It's all

Examples

(o) Start shaped thread : vote by agreement   (p) Line shaped thread : vote by deed

Two types of thread graph

Star Shape
Depth =1

Line Shape
Depth= 3

- Star shaped = responses as answers
- Line shaped = responses as oriented to the prior response

Example

(q) Average depth by year

Hauteur moyenne des fils de discussion par liste et par année

Listes des développeurs: niveau de profondeur des fils plus importante que celles des utilisateurs

Apprentissage aux fils filiformes de + en + longs

Election du Leader Debian qui fait l'unanimité (peu de discussions)

(r) Average depth by year

lidi	debian-devel	debian-mentors	debian-user	debian-user-french	debian-vote
dd	100	25	11	1	36
dm	25	100	10	2	19
du	11	10	100	2	5
duf	1	2	2	100	1
dv	36	19	5	1	100

(s) Top 100 contributors contribution each list

- Contributors of each list are "specialized"
- Entry barriers
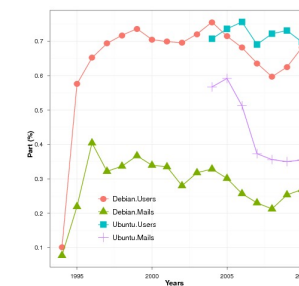
(t) % of mails and Passive Members / year

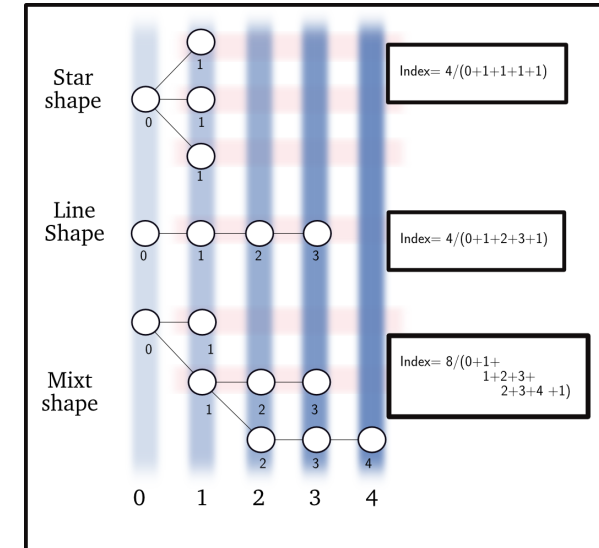- Members who stop posting after posting one at least
- Only active members re-intervene

Structure Thread index

Calcul

$$ThreadIndex = \frac{\displaystyle\sum_{1}^{N}\text{Mails}_i}{\displaystyle\sum_{1}^{N}\text{References}_i + 1}$$

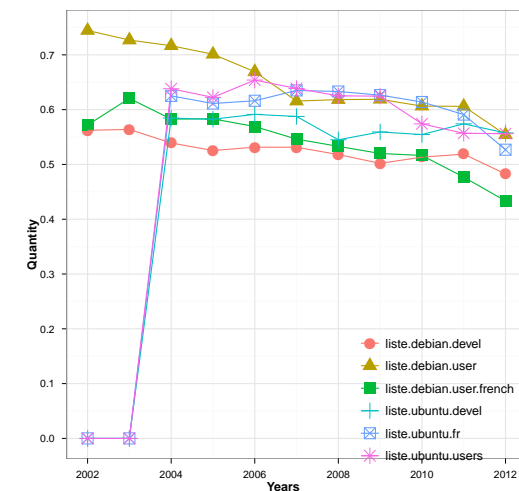## Consider thread as a directed graph

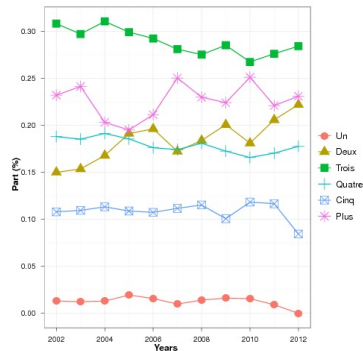(u) 3 examples of index

Structure Thread index

- Index comprehension :
  - Strictly Star Shape thread Index = 1
  - Strictly Line Shape thread Index tends to 0
- Comprehensive analysis
  - Star Shape = vote (secunded)
  - Strict line shape = collaboration
  - Big Mix analysis of flaming

Diachronic Thread index View

(v) Index evolution throw time

How many authors per thread ?

(w) Nb. of threads / nb. of persons

- Number of dyads increases whereas others are stable (monad, triad, tetrad, pintad and more)
- Relation between number of authors and structure of the thread[Conein and Latapy, 2008]
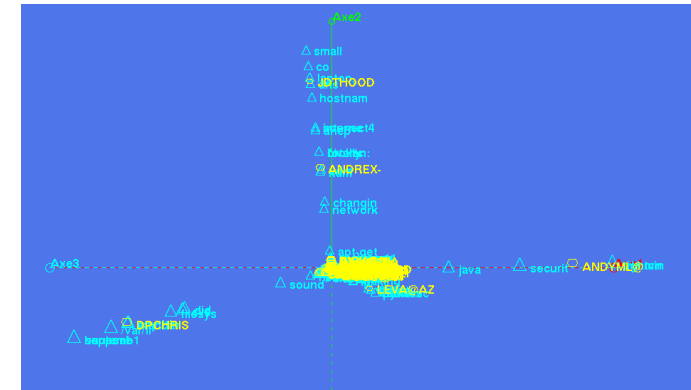
Content analysis

FIGURE: Factor analysis, Subject/From

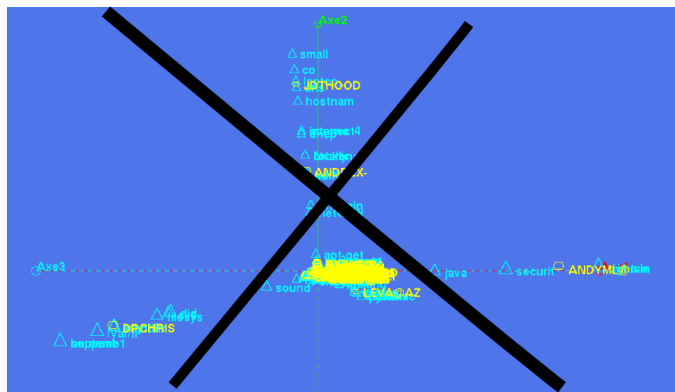## Content analysis

FIGURE: Factor analysis, Subject/From

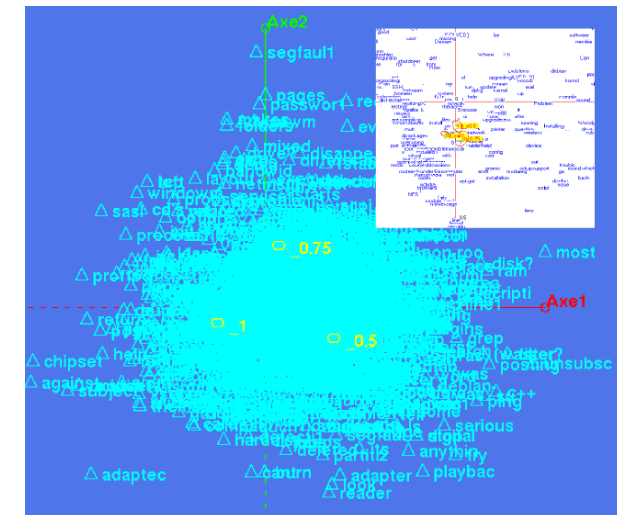## Stability of the Thread Index

FIGURE: Factor analysis, Subject/Index
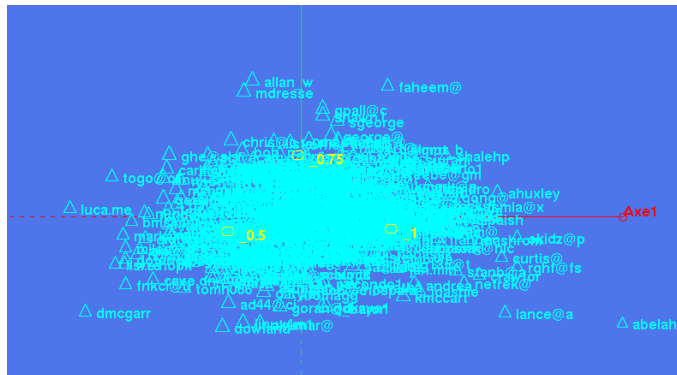
Stability of the Thread Index

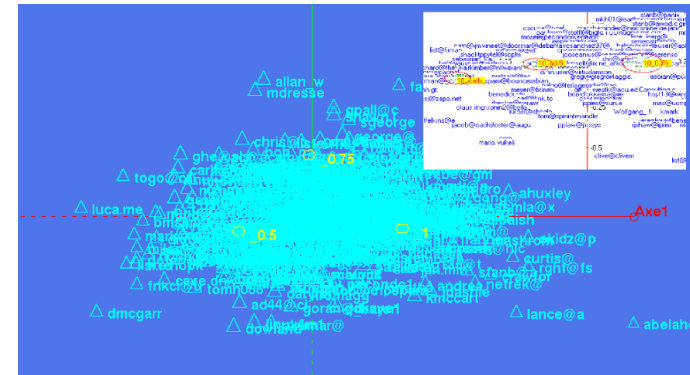FIGURE: Factor analysis, From/Index

Stability of the Thread Index

FIGURE: Factor analysis, From/Index

Interview 2010

Raphaël Hertzog

You contribute regularly to Debian mailing lists, yet I don't remember any aggressive/frustrated mail of you. How do you manage that ? Are you avoiding heated discussions ?

Joey Hess

Rats, sounds like all my wonderful flames of years past have been forgotten !
Seriously though, after I noticed thread patterns, I started trying to avoid participating in the bad patterns myself. Now I limit myself to one expression of an opinion, and I don't care who gets the last word. If people can't be convinced, it's time to find another approach to the problem. Also, code talks.
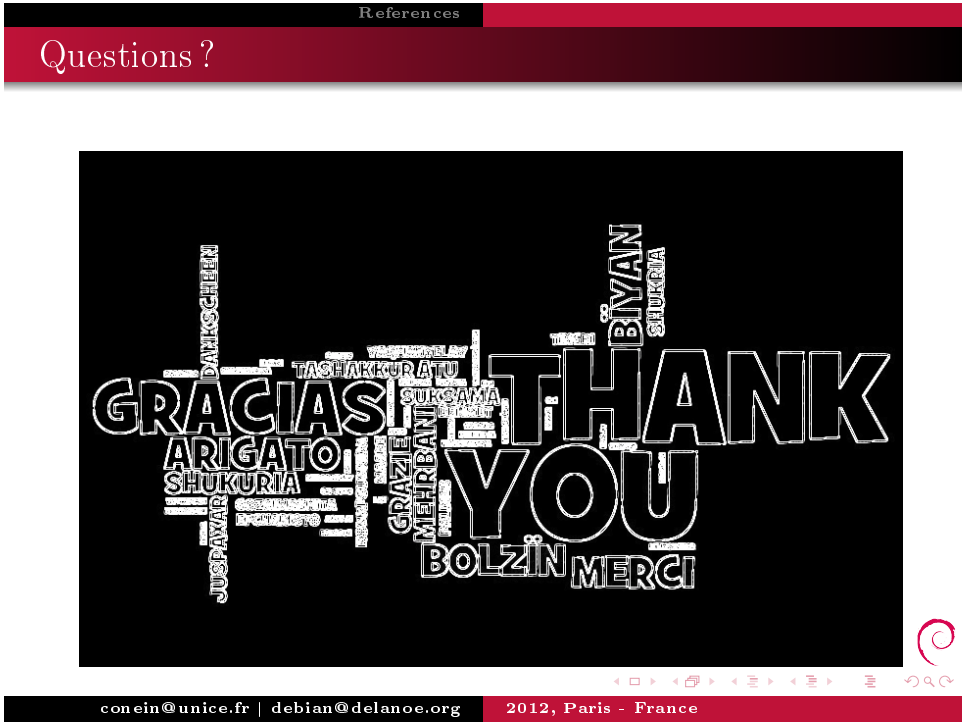
Source

http ://raphaelhertzog.com/2010/11/11/people-behind-debian-joey-hess-of-debhelper-fame/

Conclusion

Work in progress

- To analyze mailing-list archive can be usefull
- To produce not only a new meter
- But an agnostic thread index
- That everyone could manage with (ham/troll)
- But it supposes that everyone have access
- to headers of the mailing-list archive
- → How to preserve privacy and ignore trolls ?

Conein, B. and Latapy, M. (2008).
Les usages épistémiques des réseaux de communication
électronique : Le cas de l'open-source.
Sociologie du travail.