Package 'Linkage'

October 12, 2022

Encoding UTF-8

Type Package

Title Clustering Communication Networks Using the Stochastic Topic Block Model Through Linkage.fr

Version 0.9

Depends R (>= 3.5.0)

Imports httr, jsonlite, RColorBrewer, sna, network

Date 2022-04-08

Author Charles Bouveyron, Pierre Latouche, Stéphane Petiot, Carlos Ocanto

Maintainer Charles Bouveyron <charles.bouveyron@gmail.com>

Description It allows to cluster communication networks using the Stochastic Topic Block Model <doi:10.1007/s11222-016-9713-7> by posting jobs through the API of the linkage.fr server, which implements the clustering method. The package also allows to visualize the clustering results returned by the server.

License GPL-3

NeedsCompilation no

Repository CRAN

Date/Publication 2022-04-12 09:02:30 UTC

R topics documented:

Linkage-package *Clustering Communication Networks Using the Stochastic Topic Block Model Through Linkage.fr*

Description

It allows to cluster communication networks using the Stochastic Topic Block Model <doi:10.1007/s11222-016-9713-7> by posting jobs through the API of the linkage.fr server, which implements the clustering method. The package also allows to visualize the clustering results returned by the server.

Details

The DESCRIPTION file:

Encoding:	UTF-8
Package:	Linkage
Type:	Package
Title:	Clustering Communication Networks Using the Stochastic Topic Block Model Through Linkage.fr
Version:	0.9
Depends:	R (>= 3.5.0)
Imports:	httr, jsonlite, RColorBrewer, sna, network
Date:	2022-04-08
Author:	Charles Bouveyron, Pierre Latouche, Stéphane Petiot, Carlos Ocanto
Maintainer:	Charles Bouveyron <charles.bouveyron@gmail.com>
Description:	It allows to cluster communication networks using the Stochastic Topic Block Model <doi:10.1007/s11222-016
License:	GPL-3

Index of help topics:

Enron	The Enron email network
Linkage-package	Clustering Communication Networks Using the Stochastic Topic Block Model Through Linkage.fr
linkage.check	Monitor achievment of the current job
linkage.getresults	Retrieve results for a specific job.
linkage.post	Post a job on Linkage.fr to cluster a network with STBM
plot.linkage	The plot function for 'linkage' objects.

It allows to cluster communication networks using the Stochastic Topic Block Model (Bouveyron et al., 2018, <doi:10.1007/s11222-016-9713-7>) by posting jobs through the API of the linkage.fr server, which implements the clustering method. The package also allows to visualize the clustering results returned by the server.

Author(s)

Charles Bouveyron, Pierre Latouche, Stéphane Petiot, Carlos Ocanto

Maintainer: Charles Bouveyron <charles.bouveyron@gmail.com>

References

C. Bouveyron, P. Latouche and R. Zreik, The Stochastic Topic Block Model for the Clustering of Networks with Textual Edges, Statistics and Computing, vol. 28(1), pp. 11-31, 2017 <doi:10.1007/s11222-016-9713-7>

Examples

```
## Not run:
data(Enron)
write.table(Enron, file="Enron.csv",row.names=FALSE,col.names=FALSE, sep=",")
file = "Enron.csv"

# Provide the user token, which is provided on "developers" page
# of http://linkage.fr (after registration)
token = "xxxxxxxxxxxxxxxxxxx"

# Post the job
job_id = linkage.post(file, token, job_title="My job: Enron",
                      clusters_min = 8, clusters_max = 8,
                      topics_min = 6,topics_max = 6,
                      filter_largest_subgraph = TRUE)

# Monitor achievment of the current job
ans = linkage.check(token)

# Retrieve results (once achievment is 100
res = linkage.getresults(job_id,token)

# Plot the results
plot(res,type='all')

## End(Not run)
```

Enron *The Enron email network*

Description

This data set contains an extract of the email network of the Enron company. This extract focuses on the emails exchanged between Enron employees in October 2001. The reported texts of the emails are only the email subjects. The full email data set is available at https://www.cs.cmu.edu/~enron/.

Usage

```
data(Enron)
```

Format

The data frame is organized as follows:

- the first column contains the id of the sender,

- the second column contains the id of the receiver,

- the third column contains the text of the email

Source

The full email data set is available at https://www.cs.cmu.edu/~enron/.

References

C. Bouveyron, P. Latouche and R. Zreik, The Stochastic Topic Block Model for the Clustering of Networks with Textual Edges, Statistics and Computing, vol. 28(1), pp. 11-31, 2017 <doi:10.1007/s11222-016-9713-7>

Examples

```
## Not run:
data(Enron)
write.table(Enron, file="Enron.csv",row.names=FALSE,col.names=FALSE, sep=",")
file = "Enron.csv"

# Provide the user token, which is provided on "developers" page
# of http://linkage.fr (after registration)
token = "xxxxxxxxxxxxxxxxxxx"

# Post the job
job_id = linkage.post(file, token, job_title="My job: Enron",
                      clusters_min = 8, clusters_max = 8,
                      topics_min = 6,topics_max = 6,
                      filter_largest_subgraph = TRUE)

# Monitor achievment of the current job
ans = linkage.check(token)

# Retrieve results (once achievment is 100%)
res = linkage.getresults(job_id,token)

# Plot the results
plot(res,type='all')

## End(Not run)
```

linkage.check *Monitor achievment of the current job*

Description

Monitor the achievment of the current job by checking on the web server linkage.fr.

Usage

```
linkage.check(token)
```

Arguments

token	The token of the user. This personal token can be found on https://linkage.fr/developers/ after registration. Registration is free of charge for individual and academic users.

Value

It returns a list containing in particular:

id	the job id
progress	the achievment of the current job (in percentage)

Author(s)

Charles Bouveyron <charles.bouveyron@gmail.com>

References

C. Bouveyron, P. Latouche and R. Zreik, The Stochastic Topic Block Model for the Clustering of Networks with Textual Edges, Statistics and Computing, vol. 28(1), pp. 11-31, 2017 <doi:10.1007/s11222-016-9713-7>

Examples

```
## Not run:
data(Enron)
write.table(Enron, file="Enron.csv",row.names=FALSE,col.names=FALSE, sep=",")
file = "Enron.csv"

# Provide the user token, which is provided on "developers" page
# of http://linkage.fr (after registration)
token = "xxxxxxxxxxxxxxxxxxxx"

# Post the job
job_id = linkage.post(file, token, job_title="My job: Enron",
                      clusters_min = 8, clusters_max = 8,
                      topics_min = 6,topics_max = 6,
```

```
                        filter_largest_subgraph = TRUE)

# Monitor achievment of the current job
ans = linkage.check(token)

# Retrieve results (once achievment is 100
res = linkage.getresults(job_id,token)

# Plot the results
plot(res,type='all')

## End(Not run)
```

`linkage.getresults` *Retrieve results for a specific job.*

Description

Retrieve results for a specific job posted on the Linkage.fr server.

Usage

```
linkage.getresults(job_id, token)
```

Arguments

`job_id`	The id of the job to retrieve (as returned by the linkage.post or the linkage.check functions).
`token`	The token of the user. This personal token can be found on https://linkage.fr/developers/ after registration. Registration is free of charge for individual and academic users.

Value

It returns a list containing in particular:

`job_id`	the job id
`nb_nodes`	the number of nodes
`nb_edges`	the number of edges
`clusters_optim`	the optimal number of clusters
`topics_optim`	the optimal number of topics
`dictionary`	the list of words used in the texts
`result`	a list containing the clustering results for the optimal numbers of clusters and topics. This list contains in particular: - clusters_mat: clustering of the nodes - rho_mat: node cluster proportions - pi_mat: estimated connexion probabilities between clusters - theta_qr_mat: estimated proportions of topics in interactions between groups - top_words: most representative words for each topic

Author(s)

Charles Bouveyron <charles.bouveyron@gmail.com>

References

C. Bouveyron, P. Latouche and R. Zreik, The Stochastic Topic Block Model for the Clustering of Networks with Textual Edges, Statistics and Computing, vol. 28(1), pp. 11-31, 2017 <doi:10.1007/s11222-016-9713-7>

Examples

```
## Not run:
data(Enron)
write.table(Enron, file="Enron.csv",row.names=FALSE,col.names=FALSE, sep=",")
file = "Enron.csv"

# Provide the user token, which is provided on "developers" page
# of http://linkage.fr (after registration)
token = "xxxxxxxxxxxxxxxxxxx"

# Post the job
job_id = linkage.post(file, token, job_title="My job: Enron",
                      clusters_min = 8, clusters_max = 8,
                      topics_min = 6,topics_max = 6,
                      filter_largest_subgraph = TRUE)

# Monitor achievment of the current job
ans = linkage.check(token)

# Retrieve results (once achievment is 100
res = linkage.getresults(job_id,token)

# Plot the results
plot(res,type='all')

## End(Not run)
```

`linkage.post` *Post a job on Linkage.fr to cluster a network with STBM*

Description

Post a clustering job on the server Linkage.fr though the API. The Linkage.fr server implements the Stochastic Topic Block Model (STBM, Bouveyron et al., 2018, doi:10.1007/s11222-016-9713-7).

The users should have registered on the web server https://linkage.fr. Registration is free of charge for individual and academic users.

Usage

```
linkage.post(file, token, job_title = "", clusters_min = 2, clusters_max = 10,
             topics_min = 2, topics_max = 10, filter_largest_subgraph = TRUE)
```

Arguments

file	the location on the disk of the CSV file containing the communication network. Each line of tje CSV file should be of the form: sender_id, receiver_id, text of the message.
token	The token of the user. This personal token can be found on https://linkage.fr/developers/ after registration. Registration is free of charge for individual and academic users.
job_title	Title of the job
clusters_min	Minimum number of node clusters to test
clusters_max	Maximum number of node clusters to test
topics_min	Minimum number of topics to test
topics_max	Maximum number of topics to test
filter_largest_subgraph	a boolean indicating if the clustering should be done only on the largest subgraph or not

Value

The id of the job is returned.

Author(s)

Charles Bouveyron <charles.bouveyron@gmail.com>

References

C. Bouveyron, P. Latouche and R. Zreik, The Stochastic Topic Block Model for the Clustering of Networks with Textual Edges, Statistics and Computing, vol. 28(1), pp. 11-31, 2017 <doi:10.1007/s11222-016-9713-7>

Examples

```
## Not run:
data(Enron)
write.table(Enron, file="Enron.csv",row.names=FALSE,col.names=FALSE, sep=",")
file = "Enron.csv"

# Provide the user token, which is provided on "developers" page
# of http://linkage.fr (after registration)
token = "xxxxxxxxxxxxxxxxxxxx"

# Post the job
job_id = linkage.post(file, token, job_title="My job: Enron",
```

```
                      clusters_min = 8, clusters_max = 8,
                      topics_min = 6,topics_max = 6,
                      filter_largest_subgraph = TRUE)

# Monitor achievment of the current job
ans = linkage.check(token)

# Retrieve results (once achievment is 100
res = linkage.getresults(job_id,token)

# Plot the results
plot(res,type='all')

## End(Not run)
```

`plot.linkage` *The plot function for 'linkage' objects.*

Description

This function plots different information about 'linkage' objects.

Usage

```
## S3 method for class 'linkage'
plot(x, type="all", ...)
```

Arguments

`x`	an object of type 'linkage' to plot
`type`	the type of information to plot: - "all": all information, - "network": the clustered network, - "metanetwork": the metanetwork which summarizes all model parameters, - "topics": the most representative words of each topic, - "prop": the node cluster proportions.
`...`	Additional options to pass to the plot function.

Value

No value is returned by this function.

Author(s)

Charles Bouveyron <charles.bouveyron@gmail.com>

References

C. Bouveyron, P. Latouche and R. Zreik, The Stochastic Topic Block Model for the Clustering of Networks with Textual Edges, Statistics and Computing, vol. 28(1), pp. 11-31, 2017 <doi:10.1007/s11222-016-9713-7>

Examples

```
## Not run:
data(Enron)
write.table(Enron, file="Enron.csv",row.names=FALSE,col.names=FALSE, sep=",")
file = "Enron.csv"

# Provide the user token, which is provided on "developers" page
# of http://linkage.fr (after registration)
token = "xxxxxxxxxxxxxxxxxxxx"

# Post the job
job_id = linkage.post(file, token, job_title="My job: Enron",
                      clusters_min = 8, clusters_max = 8,
                      topics_min = 6,topics_max = 6,
                      filter_largest_subgraph = TRUE)

# Monitor achievment of the current job
ans = linkage.check(token)

# Retrieve results (once achievment is 100
res = linkage.getresults(job_id,token)

# Plot the results
plot(res,type='all')

## End(Not run)
```

Index