

Package ‘diffusr’

November 28, 2016

Type Package

Title Network Diffusion Algorithms

Version 0.1.0

Maintainer Simon Dirmeier <simon.dirmeier@gmx.de>

Description Implementation of network diffusion algorithms such as insulated heat propagation or Markov random walks. Network diffusion algorithms generally spread information in the form of node weights along the edges of a graph to other nodes. These weights can for example be interpreted as temperature, an initial amount of water, the activation of neurons in the brain, or the location of a random surfer in the internet. The information (node weights) is iteratively propagated to other nodes until a equilibrium state or stop criterion occurs.

URL <https://github.com/dirmeier/diffusr>

BugReports <https://github.com/dirmeier/diffusr/issues>

License GPL (>= 3)

LazyData TRUE

LinkingTo Rcpp, RcppEigen

Imports Rcpp

Suggests knitr, rmarkdown, testthat, lintr, igraph

VignetteBuilder knitr

RoxygenNote 5.0.1

SystemRequirements C++11

NeedsCompilation yes

Author Simon Dirmeier [aut, cre]

Repository CRAN

Date/Publication 2016-11-28 00:33:47

R topics documented:

diffusr-package	2
insulated.heat.diffusion	3
laplacian.heat.diffusion	4
nearest.neighbors	5
normalize.laplacian	6
normalize.stochastic	6
random.walk	7

Index	9
--------------	----------

diffusr-package	<i>diffusr</i>
-----------------	----------------

Description

Network diffusion algorithms in R.

Author(s)

Simon Dirmeier <simon.dirmeier@gmx.de>

References

Tong, H., Faloutsos, C., & Pan, J. Y. (2006), Fast random walk with restart and its applications.

Koehler, S., Bauer, S., Horn, D., & Robinson, P. N. (2008), Walking the interactome for prioritization of candidate disease genes. *The American Journal of Human Genetics*

Bonacich, P. (1987), Power and centrality: A family of measures. *American Journal of Sociology*

Leiserson, M. D., Vandin, F., Wu, H. T., Dobson, J. R., Eldridge, J. V., Thomas, J. L., ... & Lawrence, M. S. (2015), Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes. *Nature genetics*

https://en.wikipedia.org/wiki/Laplacian_matrix

https://en.wikipedia.org/wiki/Heat_equation

`insulated.heat.diffusion`*Graph diffusion using an insulated heat diffusion process*

Description

An amount of starting heat h_0 of nodes on a graph is distributed across its edges. The insulated heat diffusion calculates the stationary distribution of the different temperatures on the nodes.

Usage

```
insulated.heat.diffusion(h0, graph, r = 0.5, ...)
```

Arguments

<code>h0</code>	an n-dimensional numeric non-negative vector of starting temperatures
<code>graph</code>	an (n x n)-dimensional numeric non-negative adjacency matrix representing the graph
<code>r</code>	rate of heat diffusion, where 1 is the maximum diffusion and 0 no diffusion at all
<code>...</code>	additional parameters

Value

returns the heat on every node as numeric vector

Author(s)

Simon Dirmeier, <simon.dirmeier@gmx.de>

References

Power and centrality: A family of measures. *American Journal of Sociology*

Leiserson, M. D., Vandin, F., Wu, H. T., Dobson, J. R., Eldridge, J. V., Thomas, J. L., ... & Lawrence, M. S. (2015), Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes. *Nature genetics*

Examples

```
# count of nodes
n <- 5
# starting distribution (has to sum to one)
h0 <- as.vector(rmultinom(1, 1, prob=rep(.2, n)))
# adjacency matrix (either normalized or not)
```

```
graph <- matrix(abs(rnorm(n*n)), n, n)
# computation of stationary distribution
ht <- insulated.heat.diffusion(h0, graph)
```

laplacian.heat.diffusion

Graph diffusion using a heat diffusion process on a Laplacian matrix.

Description

An amount of starting heat gets distribution using the Laplacian matrix of a graph. Every iteration (or time interval) t heat streams from the starting nodes into surrounding nodes.

Usage

```
laplacian.heat.diffusion(h0, graph, t = 0.5, ...)
```

Arguments

<code>h0</code>	an n -dimensional numeric non-negative vector of starting temperatures
<code>graph</code>	an $(n \times n)$ -dimensional numeric non-negative adjacency matrix representing the graph
<code>t</code>	time point when heat is measured
<code>...</code>	additional parameters

Value

returns the heat on every node as numeric vector

Author(s)

Simon Dirmeier, <simon.dirmeier@gmx.de>

References

https://en.wikipedia.org/wiki/Laplacian_matrix
https://en.wikipedia.org/wiki/Heat_equation

Examples

```
# count of nodes
n <- 5
# starting distribution (has to sum to one)
h0 <- as.vector(rmultinom(1, 1, prob=rep(.2, n)))
# adjacency matrix (either normalized or not)
graph <- matrix(abs(rnorm(n*n)), n, n)
# computation of stationary distribution
ht <- laplacian.heat.diffusion(h0, graph)
```

nearest.neighbors *Graph diffusion using nearest neighbors*

Description

For every node in a set of nodes the graph gets traversed along the node's shortest paths to its neighbors. Nearest neighbors are added until a maximum depth of k is reached. For settings where there are more than k neighbors having the same distance, all neighbors are returned.

Usage

```
nearest.neighbors(nodes, graph, k = 1L, ...)
```

Arguments

nodes	a n -dimensional integer vector of node indexes (1-based) for which the algorithm is applied iteratively
graph	an $(n \times n)$ -dimensional numeric non-negative adjacency matrix representing the graph
k	the depth of the nearest neighbor search, e.g. the depth of the graph traversal
...	additional parameters

Value

returns the kNN nodes as list of integer vectors of node indexes

Author(s)

Simon Dirmeier, <simon.dirmeier@gmx.de>

Examples

```
# count of nodes
n <- 10
# indexes (integer) of nodes for which neighbors should be searched
node.idx <- c(1L, 5L)
# the adjacency matrix (does not need to be symmetric)
graph <- rbind(cbind(0, diag(n-1)), 0)
# compute the neighbors until depth 3
neigh <- nearest.neighbors(node.idx, graph, 3)
```

normalize.laplacian *Calculate the Laplacian of a matrix*

Description

Calculate the Laplacian of a matrix

Usage

```
normalize.laplacian(obj, ...)
```

Arguments

obj	matrix for which the Laplacian is calculated
...	additional params

Value

returns the Laplacian

Author(s)

Simon Dirmeier, <simon.dirmeier@gmx.de>

Examples

```
W <- matrix(abs(rnorm(10000)), 100, 100)
lapl.W <- normalize.laplacian(W)
```

normalize.stochastic *Create a stochastically normalized matrix/vector*

Description

Create a stochastically normalized matrix/vector

Usage

```
normalize.stochastic(obj, ...)
```

Arguments

obj	matrix/vector that is stochastically normalized
...	additional params

Value

returns the normalized matrix/vector

Author(s)

Simon Dirmeier, <simon.dirmeier@gmx.de>

Examples

```
W <- matrix(abs(rnorm(10000)), 100, 100)
stoch.W <- normalize.stochastic(W)
```

random.walk

Graph diffusion using a Markov random walk

Description

A Markov Random Walk takes an initial distribution p_0 and calculates the stationary distribution of that. The diffusion process is regulated by a restart probability r which controls how often the MRW jumps back to the initial values.

Usage

```
random.walk(p0, graph, r = 0.5, ...)
```

Arguments

p_0	an n -dimensional numeric non-negative vector representing the starting distribution of the Markov chain (does not need to sum to one)
graph	an $(n \times n)$ -dimensional numeric non-negative adjacency matrix representing the graph
r	a scalar between $(0, 1)$. restart probability if a Markov random walk with restart is desired
...	additional parameters

Value

returns the stationary distribution as numeric vector

Author(s)

Simon Dirmeier, <simon.dirmeier@gmx.de>

References

Tong, H., Faloutsos, C., & Pan, J. Y. (2006), Fast random walk with restart and its applications.

Koehler, S., Bauer, S., Horn, D., & Robinson, P. N. (2008), Walking the interactome for prioritization of candidate disease genes. *The American Journal of Human Genetics*

Examples

```
# count of nodes
n <- 5
# starting distribution (has to sum to one)
p0 <- as.vector(rmultinom(1, 1, prob=rep(.2, n)))
# adjacency matrix (either normalized or not)
graph <- matrix(abs(rnorm(n*n)), n, n)
# computation of stationary distribution
pt <- random.walk(p0, graph)
```

Index

*Topic **package**

diffusr-package, [2](#)

diffusr-package, [2](#)

insulated.heat.diffusion, [3](#)

laplacian.heat.diffusion, [4](#)

nearest.neighbors, [5](#)

normalize.laplacian, [6](#)

normalize.stochastic, [6](#)

random.walk, [7](#)