

Package ‘ppiPre’

July 2, 2014

Type Package

Title Predict protein-protein interactions based on functional and topological similarities.

Version 1.5

Date 2013-07-27

Author Yue Deng, Rongjie Shao, Gang Wang, Yuanjun Sun

Maintainer Yue Deng <anfdeng@163.com>

Description This package implements several functions useful for computing similarities between proteins based on their GO annotation, KEGG annotation and PPI network topology. It integrates these similarities to predict PPIs using an SVM classifier. Twenty species are supported, including Anopheles, Arabidopsis, Bovine, Canine, Chicken, Chimp, E coli strain K12 and strain Sakai, Fly, Human, Malaria, Mouse, Pig, Rhesus, Rat, Worm, Xenopus, Yeast, Streptomyces coelicolor and Zebrafish.

License Artistic-2.0

LazyLoad yes

Depends R (>= 2.15)

Imports AnnotationDbi, igraph, e1071

Suggests

GO.db, org.Sc.sgd.db, org.Hs.eg.db, KEGG.db, org.Ag.eg.db, org.At.tair.db, org.Bt.eg.db, org.Ce.eg.db, org.Cf.eg.db, org.Dr.eg.db, org.Ec.eg.db, org.Fl.eg.db, org.Gg.eg.db, org.Mm.eg.db, org.Ml.eg.db, org.Rn.eg.db, org.Ss.eg.db, org.Xl.eg.db, org.Ye.eg.db, org.Zf.eg.db

NeedsCompilation no

Repository CRAN

Date/Publication 2013-07-28 12:07:52

R topics documented:

ppiPre-package	2
AASim	3
ComputeAllEvidences	4
FNPre	5
GOKEGGSims	6
GOKEGGSimsFromFile	7
IC	8
IntelliGOGeneSim	9
JaccardSim	10
KEGGSim	11
RASim	12
SVMPredict	13
SVMTrain	14
TCSSGeneSim	15
TopologicSims	16
WangGeneSim	17
Index	18

ppiPre-package *Predicting protein-protein interactions*

Description

Predicting protein-protein interactions based on functional and topologic similarities

Details

This package implements several functions useful for computing similarities between proteins based on their GO annotation, KEGG annotation and network topology. It integrates theses similarities to predict PPIs using an SVM classifier. Twenty species are supported.

Package: ppiPre
 Type: Package
 Version: 1.0
 Date: 2011-07-06
 Depends: AnnotationDbi, GO.db, KEGG.db, org.Sc.sgd.db, org.Hs.eg.db, igraph, e1071
 Suggests: org.Ag.eg.db, org.At.tair.db, org.Bt.eg.db, org.Ce.eg.db, org.Cf.eg.db, org.Dm.eg.db,
 org.Dr.eg.db, org.EcK12.eg.db, org.EcSakai.eg.db, org.Gg.eg.db, org.Mm.eg.db,
 org.Mmu.eg.db, org.Pf.plasmo.db, org.Pt.eg.db, org.Rn.eg.db, org.Sco.eg.db,
 org.Ss.eg.db, org.Xl.eg.db
 License: GPL Version 2
 LazyLoad: yes

Author(s)

Yue Deng

Maintainer: Yue Deng <anfdeng@163.com>

See Also[GOKEGGSims](#) [IntelliGOGeneSim](#) [TCSSGeneSim](#) [SVMPredict](#) [TopologicSims](#)

AASim*Compute Adamic-Adar Index Between Two Nodes in PPI Network*

Description

Given a network and two nodes, this function will calculate the Adamic-Adar Index between the two nodes

Usage

```
AASim(node1, node2, graph)
```

Arguments

node1	Name of one node in the PPI network.
node2	Name of another node in the PPI network.
graph	A PPI network

Value

The Adamic-Adar Index between the two nodes

Author(s)

Yue Deng <anfdeng@163.com>

References

[1] L. A. Adamic and E. Adar, "Friends and neighbors on the Web", *Social Networks*, vol. 25, no. 3, pp. 211-230, 2003.

See Also[RASim](#) [JaccardSim](#)**Examples**

```
edges <- data.frame(node1=c("1132", "1133", "1134", "1134", "1145", "1147", "1147", "1147"),
                    node2=c("1134", "1134", "1145", "1147", "1147", "1148", "1149", "1150"))
graph<-igraph::graph.data.frame(edges,directed=FALSE)
AASim("1134","1147",graph)
```

ComputeAllEvidences *Compute the Biological and Topological Similarities Between Protein Pairs*

Description

Given an input file, this function will calculate 4 biological and 3 topological similarities between each protein pair in the file.

Usage

```
ComputeAllEvidences(input,output="AllEvidences-ppiPre.csv",organism="yeast", drop ="IEA",
header=TRUE, sep=",")
```

Arguments

input	CSV format PPI network, each line of which includes two interacting proteins and another column which is 1 or 0 indicating that the interaction is positive or negative.
output	Result will be saved in the output file.
organism	One of "anopheles", "arabidopsis", "bovine", "canine", "chicken", "chimp", "ecoli12", "ecsakai", "fly", "human", "malaria", "mouse", "pig", "rat", "rhesus", "worm", "xenopus", "yeast", "coelicolor" and "zebrafish".
drop	A set of evidence codes based on which certain annotations are dropped. Use NULL to keep all GO annotations.
header	If the input file contains header.
sep	The separator of the input file.

Author(s)

Yue Deng <anfdeng@163.com>

References

- [1] S. Benabderrahmane, M. Smail-Tabbone, O. Poch, A. Napoli, and M.-D. Devignes, "IntelliGO: a new vector-based semantic similarity measure including annotation origin", BMC Bioinformatics, vol. 11, no. 1, p. 588, 2010.
- [2] J. Z. Wang, Z. Du, R. Payattakool, P. S. Yu, and C.-F. Chen, "A new method to measure the semantic similarity of GO terms", Bioinformatics, vol. 23, no. 10, pp. 1274-1281, May. 2007.
- [3] S. Jain and G. Bader, "An improved method for scoring protein-protein interactions using semantic similarity within the gene ontology", BMC Bioinformatics, vol. 11, no. 1, p. 562, 2010.
- [4] T. Zhou, L. Lv, and Y.-C. Zhang, "Predicting missing links via local information", The European Physical Journal B - Condensed Matter and Complex Systems, vol. 71, no. 4, pp. 623-630, Oct. 2009

- [5] L. A. Adamic and E. Adar, "Friends and neighbors on the Web", *Social Networks*, vol. 25, no. 3, pp. 211-230, 2003.
- [6] P. Jaccard, "Etude comparative de la distribution florale dans une portion des Alpes et des Jura", *Bull. Soc. Vaud. Sci. Nat.*, vol. 37, p. 541, 1901.
- [7] M. Kanehisa and S. Goto, "KEGG: Kyoto Encyclopedia of Genes and Genomes", *Nucleic Acids Research*, vol. 28, no. 1, pp. 27 -30, 2000.

See Also

[TCSSGeneSim](#) [WangGeneSim](#) [IntelliGOGeneSim](#) [GOKEGGSims](#)

Examples

```
#edges <- data.frame(node1=c("1132", "1133", "1134", "1134", "1145", "1147"),
#                    node2=c("1134", "1134", "1145", "1147", "1147", "1149"),
#                    label=c(1, 1, 1, 0, 0, 0))
#graph<-igraph::graph.data.frame(edges,directed=FALSE)
#samplefile <- "ppiPre-ComputeAllEvidences-sample.csv"
#write.csv(edges,file=samplefile,row.names=FALSE)
#ComputeAllEvidences(input=samplefile,output="AllEvidences-ppiPre.csv",organism="human")
#result<-read.csv(file="AllEvidences-ppiPre.csv")
#print(result)
```

FNPre

Predict false negative interactions based on topological similarities

Description

User can predict false negative interactions from an given PPI network, based on one to three topological similarities.

Usage

```
FNPre(file,indicator=c("RA","AA","Jaccard"),threshold=0.1,
      output="FalseNegativePreResult-ppiPre.csv")
```

Arguments

file	CSV format PPI network given by user, each line of which includes two interacting proteins.
indicator	Can be any combination of "RA", "AA", and "Jaccard", indicate the similarities used.
threshold	The ratio of false negative interactions to positive interactions in the network.
output	Result will be saved in the output file.

Author(s)

Yue Deng <anfdeng@163.com>

References

- [1] T. Zhou, L. Lv, and Y.-C. Zhang, "Predicting missing links via local information", *The European Physical Journal B - Condensed Matter and Complex Systems*, vol. 71, no. 4, pp. 623-630, Oct. 2009
- [2] L. A. Adamic and E. Adar, "Friends and neighbors on the Web", *Social Networks*, vol. 25, no. 3, pp. 211-230, 2003.
- [3] P. Jaccard, "Etude comparative de la distribution florale dans une portion des Alpes et des Jura", *Bull. Soc. Vaud. Sci. Nat.*, vol. 37, p. 541, 1901.

See Also

[AASim](#) [JaccardSim](#) [RASim](#)

Examples

```
edges <- data.frame(node1=c("1132", "1133", "1134"),node2=c("1134", "1134", "1145"))
graph<-igraph::graph.data.frame(edges,directed=FALSE)
samplefile <- "ppiPre-FNPre-sample.csv"
write.csv(edges,file=samplefile,row.names=FALSE)
FNPre(file=samplefile, indicator = c("RA", "AA"), threshold = 0.1)
result<-read.csv(file="FalseNegativePreResult-ppiPre.csv")
print(result)
```

GOKEGGSims

GO- and KEGG- based Similarities Between two Genes

Description

Given two genes, this function will calculate KEGG-based similarity and three GO-based similarities between them

Usage

```
GOKEGGSims(gene1, gene2, organism = "yeast", drop = "IEA")
```

Arguments

gene1	Entrez gene id.
gene2	Another entrez gene id.
organism	One of "anopheles", "arabidopsis", "bovine", "canine", "chicken", "chimp", "ecoli12", "ecsakai", "fly", "human", "malaria", "mouse", "pig", "rat", "rhesus", "worm", "xenopus", "yeast", "coelicolor" and "zebrafish."
drop	A set of evidence codes based on which certain annotations are dropped. Use NULL to keep all GO annotations.

Author(s)

Yue Deng <anfdeng@163.com>

References

- [1] S. Benabderrahmane, M. Smail-Tabbone, O. Poch, A. Napoli, and M.-D. Devignes, "IntelliGO: a new vector-based semantic similarity measure including annotation origin", *BMC Bioinformatics*, vol. 11, no. 1, p. 588, 2010.
- [2] J. Z. Wang, Z. Du, R. Payattakool, P. S. Yu, and C.-F. Chen, "A new method to measure the semantic similarity of GO terms", *Bioinformatics*, vol. 23, no. 10, pp. 1274-1281, May. 2007.
- [3] S. Jain and G. Bader, "An improved method for scoring protein-protein interactions using semantic similarity within the gene ontology", *BMC Bioinformatics*, vol. 11, no. 1, p. 562, 2010.

See Also

[TCSSGeneSim](#) [WangGeneSim](#) [IntelliGOGeneSim](#) [GOKEGGSimsFromFile](#)

Examples

```
#GOKEGGSims("1134", "1145", organism = "human", drop = "IEA")
#GOKEGGSims("YER133W", "YBR045C", organism = "yeast")
```

GOKEGGSimsFromFile *GO- and KEGG- based Similarities Between two Genes*

Description

Given an input file, this function will calculate KEGG-based similarity and three GO-based similarities between each protein pair in the file

Usage

```
GOKEGGSimsFromFile(input, output = "GOKEGGSims-ppiPre.csv", header = TRUE, sep = ",",
                   organism = "yeast", drop = "IEA")
```

Arguments

input	CSV format PPI network, each line of which is two interacting proteins.
output	Result will be saved in output file.
header	If the input file contains header.
sep	The separator of the input file.
organism	One of "anopheles", "arabidopsis", "bovine", "canine", "chicken", "chimp", "ecoli12", "ecsakai", "fly", "human", "malaria", "mouse", "pig", "rat", "rhesus", "worm", "xenopus", "yeast", "coelicolor" and "zebrafish."
drop	A set of evidence codes based on which certain annotations are dropped. Use NULL to keep all GO annotations.

Author(s)

Yue Deng <anfdeng@163.com>

References

- [1] S. Benabderrahmane, M. Smail-Tabbone, O. Poch, A. Napoli, and M.-D. Devignes, "IntelliGO: a new vector-based semantic similarity measure including annotation origin", *BMC Bioinformatics*, vol. 11, no. 1, p. 588, 2010.
- [2] J. Z. Wang, Z. Du, R. Payattakool, P. S. Yu, and C.-F. Chen, "A new method to measure the semantic similarity of GO terms", *Bioinformatics*, vol. 23, no. 10, pp. 1274-1281, May. 2007.
- [3] S. Jain and G. Bader, "An improved method for scoring protein-protein interactions using semantic similarity within the gene ontology", *BMC Bioinformatics*, vol. 11, no. 1, p. 562, 2010.

See Also

[TCSSGeneSim](#) [WangGeneSim](#) [IntelliGOGeneSim](#) [GOKEGGSims](#)

Examples

```
#edges <- data.frame(node1=c("1132", "1133", "1134", "1134", "1145", "1147"),
#                    node2=c("1134", "1134", "1145", "1147", "1147", "1149"))
#graph<-igraph::graph.data.frame(edges,directed=FALSE)
#samplefile <- "ppiPre-GOKEGGSimsFromFile-sample.csv"
#write.csv(edges,file=samplefile,row.names=FALSE)
#GOKEGGSimsFromFile(input=samplefile,output="GOKEGGSims-ppiPre.csv", header=TRUE, sep=",",
#                   organism="human")
#result<-read.csv(file="GOKEGGSims-ppiPre.csv")
#print(result)
```

IC

Information content of GO terms

Description

The information contents of GOterms, corresponding to three GO ontologies and 20 species.

IntelliGOGeneSim *IntelliGO Semantic Similarity Between two Genes*

Description

Given two genes, this function will calculate the IntelliGO Semantic Similarity between them

Usage

```
IntelliGOGeneSim(gene1, gene2, w1 = 1, w2 = 1, ont = "MF",
                 organism = "yeast", drop = "IEA")
```

Arguments

gene1	Entrez gene id.
gene2	Another entrez gene id.
w1	EC weight
w2	EC weight
ont	One of "MF", "BP", and "CC" subontologies.
organism	One of "anopheles", "arabidopsis", "bovine", "canine", "chicken", "chimp", "ecoli12", "ecsakai", "fly", "human", "malaria", "mouse", "pig", "rat", "rhesus", "worm", "xenopus", "yeast", "coelicolor" and "zebrafish."
drop	A set of evidence codes based on which certain annotations are dropped. Use NULL to keep all GO annotations.

Value

geneSim	Semantic similarity.
G01	Corresponding GO terms for gene1.
G02	Corresponding GO terms for gene2.

Author(s)

Yue Deng <anfdeng@163.com>

References

[1] S. Benabderrahmane, M. Smail-Tabbone, O. Poch, A. Napoli, and M.-D. Devignes, "IntelliGO: a new vector-based semantic similarity measure including annotation origin", BMC Bioinformatics, vol. 11, no. 1, p. 588, 2010.

See Also

[TCSSGeneSim](#) [WangGeneSim](#)

Examples

```
#IntelliGOGeneSim("1134", "1145", ont="MF", organism="human")
#IntelliGOGeneSim("YOL001W", "YPL031C", ont="MF", organism="yeast")
```

JaccardSim

Compute Jaccard Index Between Two Nodes in PPI Network

Description

Given a network and two nodes, this function will calculate the Jaccard Index between the two nodes

Usage

```
JaccardSim(node1, node2, graph)
```

Arguments

node1	Name of one node in the PPI network.
node2	Name of another node in the PPI network.
graph	A PPI network

Value

The Jaccard Index between the two nodes

Author(s)

Yue Deng <anfdeng@163.com>

References

[1] P. Jaccard, "Etude comparative de la distribution florale dans une portion des Alpes et des Jura", Bull. Soc. Vaud. Sci. Nat, vol. 37, p. 541, 1901.

See Also

[AASim](#) [RASim](#)

Examples

```
edges <- data.frame(node1=c("1132", "1133", "1134", "1134", "1145", "1147", "1147", "1147"),
                    node2=c("1134", "1134", "1145", "1147", "1147", "1148", "1149", "1150"))
graph<-igraph::graph.data.frame(edges,directed=FALSE)
JaccardSim("1134", "1147", graph)
```

KEGGSim

KEGG Semantic Similarity Between two Genes

Description

Given two genes, this function will calculate the KEGG semantic similarity between them, based on the shared pathways between the genes.

Usage

```
KEGGSim(protein1, protein2)
```

Arguments

protein1	Entrez gene id.
protein2	Another entrez gene id.

Value

The KEGG semantic similarity, using the Jaccard index.

Author(s)

Yue Deng <anfdeng@163.com>

References

- [1] M. Kanehisa and S. Goto, "KEGG: Kyoto Encyclopedia of Genes and Genomes", Nucleic Acids Research, vol. 28, no. 1, pp. 27 -30, 2000.
- [2] P. Jaccard, "Etude comparative de la distribution florale dans une portion des Alpes et des Jura", Bull. Soc. Vaud. Sci. Nat, vol. 37, p. 541, 1901.

See Also

[GOKEGGSims](#) [GOKEGGSimsFromFile](#)

Examples

```
#KEGGSim("YJL026W", "YGR180C")  
#KEGGSim("1134", "1145")
```

RASim	<i>Compute Resource Allocation Index Between Two Nodes in PPI Network</i>
-------	---

Description

Given a network and two nodes, this function will calculate the Resource Allocation Index between the two nodes

Usage

```
RASim(node1, node2, graph)
```

Arguments

node1	Name of one node in the PPI network.
node2	Name of another node in the PPI network.
graph	A PPI network

Value

The Resource Allocation Index between the two nodes

Author(s)

Yue Deng <anfdeng@163.com>

References

[1] T. Zhou, L. Lv, and Y.-C. Zhang, "Predicting missing links via local information", The European Physical Journal B - Condensed Matter and Complex Systems, vol. 71, no. 4, pp. 623-630, Oct. 2009.

See Also

[AASim](#) [JaccardSim](#)

Examples

```
edges <- data.frame(node1=c("1132", "1133", "1134", "1134", "1145", "1147", "1147", "1147"),
                    node2=c("1134", "1134", "1145", "1147", "1147", "1148", "1149", "1150"))
graph<-igraph::graph.data.frame(edges,directed=FALSE)
RASim("1134", "1147", graph)
```

SVMPredict

*Predict false interactions using a training set***Description**

Using a training set to train the SVM classifier, and predict false interactions in a user given PPI network.

Usage

```
SVMPredict(training_set,predict_set,output="falsePPIs-ppiPre.csv",organism="yeast",
           drop ="IEA", replaceNA=0)
```

Arguments

training_set	CSV format golden standard training set
predict_set	PPI network to do the prediction
output	CSV format file to ave the result
organism	One of "anopheles", "arabidopsis", "bovine", "canine", "chicken", "chimp", "ecoli12", "ecsakai", "fly", "human", "malaria", "mouse", "pig", "rat", "rhesus", "worm", "xenopus", "yeast", "coelicolor" and "zebrafish."
drop	A set of evidence codes based on which certain annotations are dropped. Use NULL to keep all GO annotations.
replaceNA	The value to replace NA in training and predict set.

Author(s)

Yue Deng <anfdeng@163.com>

See Also

[TopologicSims](#) [GOKEGGSims](#) [ComputeAllEvidences](#)

Examples

```
#edges <- data.frame(node1=c("1132", "1133", "1134", "1134", "1145", "1147"),
#                    node2=c("1134", "1134", "1145", "1147", "1147", "1149"),
#                    label=c(1, 1, 1, 0, 0, 0))
#graph<-graph.data.frame(edges,directed=FALSE)
#trainsample <- "ppiPre-SVMPredict-trainsample.csv"
#write.csv(edges,file=trainsample,row.names=FALSE)
#edges <- data.frame(node1=c("1132", "1133", "1134", "1134", "1146", "1147"),
#                    node2=c("1133", "1134", "1142", "1147", "1147", "1149"),
#                    label=c(1, 0, 1, 0, 1, 0))
#graph<-igraph::graph.data.frame(edges,directed=FALSE)
#predictsample <- "ppiPre-SVMPredict-predictsample.csv"
#write.csv(edges,file=predictsample,row.names=FALSE)
#SVMPredict(trainsample, predictsample, organism="human", replaceNA=0)
```

Description

Golden standard data sets is given by user, the trained SVM classifier will be returned

Usage

```
SVMTrain(input, replaceNA = 0)
```

Arguments

input	Gold-standard data sets, including positive and negative samples
replaceNA	The value to replace NA in training set

Value

The trained SVM classifier

Author(s)

Yue Deng <anf deng@163.com>

See Also

[SVMPredict](#)

Examples

```
#edges <- data.frame(node1=c("1132", "1133", "1134", "1134", "1145", "1147"),
#                    node2=c("1134", "1134", "1145", "1147", "1147", "1149"),
#                    label=c(1, 1, 1, 0, 0, 0))
#graph<-graph.data.frame(edges,directed=FALSE)
#samplefile <- "ppiPre-ComputeAllEvidences-sample.csv"
#write.csv(edges,file=samplefile,row.names=FALSE)
#ComputeAllEvidences(input=samplefile,output="AllEvidences-svmtrainsample.csv",organism="human",
#                    drop ="IEA", header=TRUE, sep=",")
#SVMTrain("AllEvidences-svmtrainsample.csv", replaceNA=0)
```

TCSSGeneSim	<i>Topological Clustering Semantic Similarity(TCSS) Between two Genes</i>
-------------	---

Description

Given two genes, this function will calculate the TCSS between them

Usage

```
TCSSGeneSim(gene1, gene2, ont = "MF", organism = "yeast", drop = "IEA")
```

Arguments

gene1	Entrez gene id.
gene2	Another entrez gene id.
ont	One of "MF", "BP", and "CC" subontologies.
organism	One of "anopheles", "arabidopsis", "bovine", "canine", "chicken", "chimp", "ecoli12", "ecsakai", "fly", "human", "malaria", "mouse", "pig", "rat", "rhesus", "worm", "xenopus", "yeast", "coelicolor" and "zebrafish."
drop	A set of evidence codes based on which certain annotations are dropped. Use NULL to keep all GO annotations.

Value

geneSim	Semantic similarity.
G01	Corresponding GO terms for gene1.
G02	Corresponding GO terms for gene2.

Author(s)

Yue Deng <anfdeng@163.com>

References

[1] S. Jain and G. Bader, "An improved method for scoring protein-protein interactions using semantic similarity within the gene ontology", BMC Bioinformatics, vol. 11, no. 1, p. 562, 2010.

See Also

[IntelliGOGeneSim](#) [WangGeneSim](#)

Examples

```
#TCSSGeneSim("1134", "1145", ont="MF", organism="human")
#TCSSGeneSim("YOR065W", "YEL024W", ont="CC", organism="yeast", drop="IEA")
```

TopologicSims *Compute topological similarities from user input file*

Description

Given an PPI network, this function will calculate three topological similarities of each protein pair in the network.

Usage

```
TopologicSims(inputfile, outputfile = "TopologicSims-ppiPre.csv", header = TRUE,
              sep = ",")
```

Arguments

inputfile	CSV format PPI network, each line of which is an two interacting proteins.
outputfile	Result will be saved in output file.
header	If the input file contains header.
sep	The separator of the input file.

Author(s)

Yue Deng <anfdeng@163.com>

References

- [1] T. Zhou, L. Ly, and Y.-C. Zhang, "Predicting missing links via local information", The European Physical Journal B - Condensed Matter and Complex Systems, vol. 71, no. 4, pp. 623-630, Oct. 2009
- [2] L. A. Adamic and E. Adar, "Friends and neighbors on the Web", Social Networks, vol. 25, no. 3, pp. 211-230, 2003.
- [3] P. Jaccard, "Étude comparative de la distribution florale dans une portion des Alpes et des Jura", Bull. Soc. Vaud. Sci. Nat, vol. 37, p. 541, 1901.

See Also

[AASim](#) [JaccardSim](#) [RASim](#)

Examples

```
edges <- data.frame(node1=c("1132", "1133", "1134", "1134", "1145", "1147", "1147", "1147"),
                  node2=c("1134", "1134", "1145", "1147", "1147", "1148", "1149", "1150"))
graph<-igraph::graph.data.frame(edges,directed=FALSE)
samplefile <- "ppiPre-TopologicSims-sample.csv"
write.csv(edges,file=samplefile,row.names=FALSE)
TopologicSims(samplefile,outputfile="TopologicSimstest-ppiPre.csv", header=TRUE, sep=",")
result<-read.csv(file="TopologicSimstest-ppiPre.csv")
print(result)
```


WangGeneSim

*Semantic Similarity Between Two Genes by Wang's Method***Description**

Given two genes, this function will calculate the Wang's Semantic Similarity between them

Usage

```
WangGeneSim(gene1, gene2, ont = "MF", organism = "yeast", drop = "IEA")
```

Arguments

gene1	Entrez gene id.
gene2	Another entrez gene id.
ont	One of "MF", "BP", and "CC" subontologies.
organism	One of "anopheles", "arabidopsis", "bovine", "canine", "chicken", "chimp", "ecoli12", "ecsakai", "fly", "human", "malaria", "mouse", "pig", "rat", "rhesus", "worm", "xenopus", "yeast", "coelicolor" and "zebrafish."
drop	A set of evidence codes based on which certain annotations are dropped. Use NULL to keep all GO annotations.

Value

geneSim	Semantic similarity.
G01	Corresponding GO terms for gene1.
G02	Corresponding GO terms for gene2.

Author(s)

Yue Deng <anfdeng@163.com>

References

[1] J. Z. Wang, Z. Du, R. Payattakool, P. S. Yu, and C.-F. Chen, "A new method to measure the semantic similarity of GO terms", *Bioinformatics*, vol. 23, no. 10, pp. 1274-1281, May. 2007.

See Also

[TCSSGeneSim](#) [IntelliGOGeneSim](#)

Examples

```
#WangGeneSim("1134", "1145", ont="MF", organism="human")
#WangGeneSim("YOR065W", "YEL024W", ont="MF", organism="yeast")
```

Index

*Topic **datasets**

IC, 8

*Topic **manip**

AASim, 3

ComputeAllEvidences, 4

FNPre, 5

GOKEGGSims, 6

GOKEGGSimsFromFile, 7

IntelliGOGeneSim, 9

JaccardSim, 10

KEGGSim, 11

RASim, 12

SVMPredict, 13

SVMTrain, 14

TCSSGeneSim, 15

TopologicSims, 16

WangGeneSim, 17

*Topic **package**

ppiPre-package, 2

AASim, 3, 6, 10, 12, 16

ComputeAllEvidences, 4, 13

FNPre, 5

GOKEGGSims, 3, 5, 6, 8, 11, 13

GOKEGGSimsFromFile, 7, 7, 11

IC, 8

Info_Contents_BP_anopheles (IC), 8

Info_Contents_BP_arabidopsis (IC), 8

Info_Contents_BP_bovine (IC), 8

Info_Contents_BP_canine (IC), 8

Info_Contents_BP_chicken (IC), 8

Info_Contents_BP_chimp (IC), 8

Info_Contents_BP_coelicolor (IC), 8

Info_Contents_BP_ecolik12 (IC), 8

Info_Contents_BP_ecsakai (IC), 8

Info_Contents_BP_fly (IC), 8

Info_Contents_BP_human (IC), 8

Info_Contents_BP_malaria (IC), 8

Info_Contents_BP_mouse (IC), 8

Info_Contents_BP_pig (IC), 8

Info_Contents_BP_rat (IC), 8

Info_Contents_BP_rhesus (IC), 8

Info_Contents_BP_worm (IC), 8

Info_Contents_BP_xenopus (IC), 8

Info_Contents_BP_yeast (IC), 8

Info_Contents_BP_zebrafish (IC), 8

Info_Contents_CC_anopheles (IC), 8

Info_Contents_CC_arabidopsis (IC), 8

Info_Contents_CC_bovine (IC), 8

Info_Contents_CC_canine (IC), 8

Info_Contents_CC_chicken (IC), 8

Info_Contents_CC_chimp (IC), 8

Info_Contents_CC_coelicolor (IC), 8

Info_Contents_CC_ecolik12 (IC), 8

Info_Contents_CC_ecsakai (IC), 8

Info_Contents_CC_fly (IC), 8

Info_Contents_CC_human (IC), 8

Info_Contents_CC_malaria (IC), 8

Info_Contents_CC_mouse (IC), 8

Info_Contents_CC_pig (IC), 8

Info_Contents_CC_rat (IC), 8

Info_Contents_CC_rhesus (IC), 8

Info_Contents_CC_worm (IC), 8

Info_Contents_CC_xenopus (IC), 8

Info_Contents_CC_yeast (IC), 8

Info_Contents_CC_zebrafish (IC), 8

Info_Contents_MF_anopheles (IC), 8

Info_Contents_MF_arabidopsis (IC), 8

Info_Contents_MF_bovine (IC), 8

Info_Contents_MF_canine (IC), 8

Info_Contents_MF_chicken (IC), 8

Info_Contents_MF_chimp (IC), 8

Info_Contents_MF_coelicolor (IC), 8

Info_Contents_MF_ecolik12 (IC), 8

Info_Contents_MF_ecsakai (IC), 8

Info_Contents_MF_fly (IC), 8

Info_Contents_MF_human (IC), 8
Info_Contents_MF_malaria (IC), 8
Info_Contents_MF_mouse (IC), 8
Info_Contents_MF_pig (IC), 8
Info_Contents_MF_rat (IC), 8
Info_Contents_MF_rhesus (IC), 8
Info_Contents_MF_worm (IC), 8
Info_Contents_MF_xenopus (IC), 8
Info_Contents_MF_yeast (IC), 8
Info_Contents_MF_zebrafish (IC), 8
IntelliGOGeneSim, 3, 5, 7, 8, 9, 15, 17

JaccardSim, 3, 6, 10, 12, 16

KEGGSim, 11

ppiPre (ppiPre-package), 2
ppiPre-package, 2

RASim, 3, 6, 10, 12, 16

SVMpredict, 3, 13, 14
SVMTrain, 14

TCSSGeneSim, 3, 5, 7–9, 15, 17
TopologicSims, 3, 13, 16

WangGeneSim, 5, 7–9, 15, 17