Package ‘igraphdata’

February 20, 2015

Version 0.2
Date 2014-02-04
Title A collection of network data sets for the igraph package
Author Gabor Csardi <csardi.gabor@gmail.com>
Maintainer Gabor Csardi <csardi.gabor@gmail.com>
Depends R (>= 2.10)
Suggests igraph
Description A small collection of various network data sets, to use with the igraph package. They also work with the igraph0 package.
License CC BY-SA 2.0 + file LICENSE
URL http://igraph.org
BugReports https://github.com/igraph/igraph/issues
NeedsCompilation no
Repository CRAN
Date/Publication 2014-02-04 22:47:20

R topics documented:

igraphdata-package ................................................. 2
foodwebs .......................................................... 2
immuno ............................................................. 5
karate ............................................................... 5
Koenigsberg ......................................................... 6
macaque ............................................................. 7
UKfaculty ........................................................... 8
USairports ........................................................... 8
yeast ................................................................. 9

Index 11
The igraphdata package

Description

The igraphdata package provides various data sets.

How to use the data sets

After loading the igraphdata package, the various data sets can be loaded using the regular `data` command.

Type in

```r
data(package="igraphdata")
```

to get a list of data sets included in this package.

Author(s)

Gabor Csardi &lt;csardi.gabor@gmail.com&gt;

References

- `data`

foodwebs A collection of food webs

Description

A list of graphs. Each one is a food web, i.e. a directed graph of predator-prey relationships.

Usage

foodwebs

Format

A named list of directed igraph graph objects. Here are the list of the graphs included:

‘ChesLower’ Lower Chesapeake Bay in Summer.


‘ChesMiddle’ Middle Chesapeake Bay in Summer.

Reference: same as for ‘ChesLower’.
‘ChesUpper’ Upper Chesapeake Bay in Summer.
   Reference: same as for ‘ChesLower’.

‘Chesapeake’ Chesapeake Bay Mesohaline Network.

‘CrystalC’ Crystal River Creek (Control).

‘CrystalD’ Crystal River Creek (Delta Temp).
   Reference: same as for ‘CrystalC’.

‘Maspalomas’ Charca de Maspalomas.

‘Michigan’ Lake Michigan Control network.

‘Mondego’ Mondego Estuary - Zostrea site.

‘Narragan’ Narragansett Bay Model.

‘Rhode’ Rhode River Watershed - Water Budget.
   Reference: Correll, D. (Unpublished manuscript) Smithsonian Institute, Chesapeake Bay Center for Environmental Research, Edgewater, Maryland 21037-0028 USA.

‘StMarks’ St. Marks River (Florida) Flow network.

‘baydry’ Florida Bay Trophic Exchange Matrix, dry season.

‘baywet’ Florida Bay Trophic Exchange Matrix, wet season.
   Reference: same as for ‘baydry’.

‘cypdry’ Cypress, dry season.
‘cypwet’ Cypress, wet season.
   Reference: same as for ‘cypdry’.

‘gramdry’ Everglades Graminoids - Dry Season.

‘gramwet’ Everglades Graminoids - Wet Season.
   Reference: same as for ‘gramdry’.

‘mangdry’ Mangrove Estuary, Dry Season.

‘mangwet’ Mangrove Estuary, Wet Season.
   Reference: same as for ‘mangdry’.

Each graph has the following vertex attributes: ‘name’ is the name of the species, ‘ECO’ is the type of the node, and integer value between one and five, meaning:

1. Living/producing compartment
2. Other compartment
3. Input
4. Output
5. Respiration.

The ‘Biomass’ vertex attribute contains the biomass of the species.

Edges are weighted, and the weights denote energy flux between the species involved.

The graphs also contain some informative graph attributes: ‘Author’, ‘Citation’, ‘URL’, and ‘name’.

Source

See references for the individual webs above. The data itself was downloaded from http://vlado.fmf.uni-lj.si/pub/networks/data/bio/foodweb/foodweb.htm.

References

See them above.
immuno

Immunoglobulin interaction network

Description

The undirected and connected network of interactions in the immunoglobulin protein. It is made up of 1316 vertices representing amino-acids and an edge is drawn between two amino-acids if the shortest distance between their C_alpha atoms is smaller than the threshold value \( \theta = 8 \) Angstrom.

Usage

immuno

Format

An undirected igraph graph object.

Graph attributes: ‘name’, ‘Citation’, ‘Author’.

Source

See reference below.

References


karate

Zachary’s karate club network

Description

Social network between members of a university karate club, led by president John A. and karate instructor Mr. Hi (pseudonyms).

The edge weights are the number of common activities the club members took part of. These activities were:

1. Association in and between academic classes at the university.
2. Membership in Mr. Hi’s private karate studio on the east side of the city where Mr. Hi taught nights as a part-time instructor.
3. Membership in Mr. Hi’s private karate studio on the east side of the city, where many of his supporters worked out on weekends.
4. Student teaching at the east-side karate studio referred to in (2). This is different from (2) in that student teachers interacted with each other, but were prohibited from interacting with their students.
5. Interaction at the university rathskeller, located in the same basement as the karate club’s workout area.
6. Interaction at a student-oriented bar located across the street from the university campus.
7. Attendance at open karate tournaments held through the area at private karate studios.
8. Attendance at intercollegiate karate tournaments held at local universities. Since both open and intercollegiate tournaments were held on Saturdays, attendance at both was impossible.

Zachary studied conflict and fission in this network, as the karate club was split into two separate clubs, after long disputes between two factions of the club, one led by John A., the other by Mr. Hi.

The ‘Faction’ vertex attribute gives the faction memberships of the actors. After the split of the club, club members chose their new clubs based on their factions, except actor no. 9, who was in John A.’s faction but chose Mr. Hi’s club.

Usage
karate

Format
An undirected igraph graph object. Vertex no. 1 is Mr. Hi, vertex no. 34 corresponds to John A.
Graph attributes: ‘name’, ‘Citation’, ‘Author’.
Vertex attributes: ‘name’, ‘Faction’.
Edge attribute: ‘weight’.

Source
See reference below.

References

---

**Koenigsberg**

*Bridges of Koenigsberg from Euler’s times*

**Description**

The Seven Bridges of Koenigsberg is a notable historical problem in mathematics. Its negative resolution by Leonhard Euler in 1735 laid the foundations of graph theory and presaged the idea of topology.

The city of Koenigsberg in Prussia (now Kaliningrad, Russia) was set on both sides of the Pregel River, and included two large islands which were connected to each other and the mainland by seven bridges.
The problem was to find a walk through the city that would cross each bridge once and only once. The islands could not be reached by any route other than the bridges, and every bridge must have been crossed completely every time (one could not walk half way onto the bridge and then turn around and later cross the other half from the other side).
Euler proved that the problem has no solution.

Usage
Koenigsberg

Format
An undirected igraph graph object with vertex attributes ‘name’ and ‘Euler_letter’, the latter is the notation from Eulers original paper; and edge attributes name (the name of the bridge) and ‘Euler_letter’, again, Euler’s notation from his paper.
This dataset is in the public domain.

Source
Wikipedia, http://de.wikipedia.org/wiki/K%C3%B6nigsberger_Br%C3%BCcken

macaque
Visuotactile brain areas and connections

Description
Graph model of the visuotactile brain areas and connections of the macaque monkey. The model consists of 45 areas and 463 directed connections.

Usage
macaque

Format
A directed igraph graph object with vertex attributes ‘name’ and ‘shape’.
This dataset is licensed under a Creative Commons Attribution-Share Alike 2.0 UK: England & Wales License, see http://creativecommons.org/licenses/by-sa/2.0/uk/ for details. Please cite the reference below if you use this dataset.

Source
See reference below.

References
USairports

**UKfaculty**

*Friendship network of a UK university faculty*

**Description**

The personal friendship network of a faculty of a UK university, consisting of 81 vertices (individuals) and 817 directed and weighted connections. The school affiliation of each individual is stored as a vertex attribute. This dataset can serve as a testbed for community detection algorithms.

**Usage**

**UKfaculty**

**Format**

A directed igraph graph object with vertex attribute 'Group', the numeric id of the school affiliation, and edge attribute 'weight', i.e. the graph is weighted.

This dataset is licensed under a Creative Commons Attribution-Share Alike 2.0 UK: England & Wales License, see [http://creativecommons.org/licenses/by-sa/2.0/uk/](http://creativecommons.org/licenses/by-sa/2.0/uk/) for details. Please cite the reference below if you use this dataset.

**Source**

See reference below.

**References**


USairports

*US airport network, 2010 December*

**Description**

The network of passenger flights between airports in the United States. The data set was compiled based on flights in 2010 December. This network is directed and edge directions correspond to flight directions. Each edge is specific to a single carrier aircraft type. Multiple carriers between the same two airports are denoted by multiple edges.

See information about the included meta-data below.

**Usage**

**USairports**
**Format**

A directed igraph graph object, with multiple edges. It has a ‘name’ graph attribute, and several vertex and edge attributes. The vertex attributes:

**name** Symbolic vertex name, this is the three letter IATA airport code.

**City** City and state, where the airport is located.

**Position** Position of the airport, in WGS coordinates.

Edge attributes:

**Carrier** Name of the airline. The network includes both domestic and international carriers that performed at least one flight in December of 2010.

**Departures** The number of departures (for a given airline and aircraft type).

**Seats** The total number of seats available on the flights carried out by a given airline, using a given aircraft type.

**Passengers** The total number of passengers on the flights carried out by a given airline, using a given aircraft type.

**Aircraft** Type of the aircraft.

**Distance** The distance between the two airports, in miles.

**Source**

Most of this information was downloaded from The Research and Innovative Technology Administration (RITA). See http://www.rita.dot.gov/about_rita/ for details. The airport position information was collected from Wikipedia and other public online sources.

---

**yeast**  
*Yeast protein interaction network*

---

**Description**

Comprehensive protein-protein interaction maps promise to reveal many aspects of the complex regulatory network underlying cellular function.

This data set was compiled by von Mering et al. (see reference below), combining various sources. Only the interactions that have ‘high’ and ‘medium’ confidence are included here.

**Usage**

yeast
Format

An undirected igraph graph object. Its graph attributes: ‘name’, ‘Citation’, ‘Author’, ‘URL’. ‘Classes’. The ‘Classes’ attribute contain the key for the classification labels of the proteins, in a data frame, the original MIPS categories are given after the semicolon:

E energy production; energy
G aminoacid metabolism; aminoacid metabolism
M other metabolism; all remaining metabolism categories
P translation; protein synthesis
T transcription; transcription, but without subcategory ‘transcriptional control’
B transcriptional control; subcategory ‘transcriptional control’
F protein fate; protein fate (folding, modification, destination)
O cellular organization; cellular transport and transport mechanisms
A transport and sensing; categories ‘transport facilitation’ and ‘regulation of / interaction with cellular environment’
R stress and defense; cell rescue, defense and virulence
D genome maintenance; DNA processing and cell cycle
C cellular fate / organization; categories ‘cell fate’ and ‘cellular communication / signal transmission’ and ‘control of cellular organization’
U uncharacterized; categories ‘not yet clear-cut’ and ‘uncharacterized’

Vertex attributes: ‘name’, ‘Description’, ‘Class’, the last one contains the class of the protein, according to the classification above.

Note that some proteins in the network did not appear in the annotation files, the ‘Class’ and ‘Description’ attributes are NA for these.

Source

The data was downloaded from http://www.nature.com/nature/journal/v417/n6887/suppinfo/nature750.html.

References

Index

*Topic **datasets**
  foodwebs, 2
  immuno, 5
  karate, 5
  Koenigsberg, 6
  macaque, 7
  UKfaculty, 8
  USairports, 8
  yeast, 9

data, 2

foodwebs, 2

igraphdata (igraphdata-package), 2
igraphdata-package, 2
immuno, 5

karate, 5
Koenigsberg, 6

macaque, 7

UKfaculty, 8
USairports, 8

yeast, 9