Package 'networkDynamicData'

July 22, 2025

Type Package

Title Dynamic (Longitudinal) Network Datasets

Version 0.3.0

Date 2025-04-10

Depends R (>= 3.0),networkDynamic (>= 0.6), network (>= 1.9)

Description A collection of dynamic network data sets from various sources and multiple authors represented as 'networkDynamic'-formatted objects.

License GPL-3 + file LICENSE

URL https://statnet.org

NeedsCompilation no

Author Skye Bender-deMoll [cre, aut],
   Martina Morris [ctb],
   Li Wang [ctb],
   Gerhard van de Bunt [ctb],
   Goele Bossaert [ctb],
   Nadine Meidert [ctb],
   SocioPatterns.org [ctb],
   Tore Opsahi [ctb],
   Radoslaw Michalski, (et al) [ctb],
   Allison Davis, (et. al.) [ctb],
   C.E. Priebe, (et. al.) [ctb]

Maintainer Skye Bender-deMoll <skyebend@uw.edu>

Repository CRAN

Date/Publication 2025-04-11 08:20:10 UTC

Contents

---

networkDynamicData-package

A collection of dynamic network data sets

Description

A collection of dynamic network data sets from various sources and multiple authors stored in `networkDynamic` format. The goal of this package is to facilitate reproducible research by providing a common resource of longitudinal relational data sets which can be used for testing dynamic network algorithms and techniques. We are grateful to the authors of each data set for giving us permission to distribute their work. Each dataset has individual copyright and license restrictions on attribution. View the help page for each dataset for additional information. Please contact the package maintainer if you would like to suggest additional appropriate data sets. The release of this package was supported by grant R01HD68395 from the National Institute of Health.

Details

Package:	networkDynamicData
License:	GPL-3 + individual attribution requirements for each dataset

The package includes the following data sets:

- concurrencyComparisonNets: A synthetic dataset of three simulated networks (base,middle,monog) with varying concurrency characteristics.

- harry_potter_support: Harry Potter support networks of Goele Bossaert and Nadine Meidert.

- hospital_contact (hospital): Hospital ward dynamic RFID contact network from SocioPatterns

- onlineNet: UCI Facebook-like Social Network

- vanDeBunt_students: van de Bunt longitudinal student friendship dataset

- davisDyn: dynamic version of Davis, et al, of Southern Women dataset (bipartite and one-mode projection)

- manufacturingEmails: emails and organizational hierarchy for Polish manufacturing company

- enronEmails: a version of the Enron email network

The networkDynamic package also contains several example data sets:

- McFarland_cls33_10_16_96: Daniel McFarland's Streaming Classroom Interactions Data set
- newcomb: Newcomb's Fraternity Networks
- windsurfers: Lin Freeman's Dynamic Network of Windsurfer Social Interactions

Author(s)

Maintainer: Skye Bender-deMoll skyebend@uw.edu

References

Please view the citation reference links for each dataset.

Examples

```
data(harry_potter_support)
?harry_potter

data(vanDeBunt_students)
?vanDeBunt_students

# the networkDynamic package contains a few as well
data(package='networkDynamic')
```

---

concurrencyComparisonNets
A synthetic dataset of three simulated dynamic networks with epidemic spread.

Description

Three single-mode undirected dynamic networks with an infection started from a single seed, The networks were simulated using the tergm and EpiModel packages. All three networks have the same size, relationship duration distribution and cross-sectional mean degree, but different cross-sectional degree distributions. They are intended as examples for illustrating and comparing the effects of concurrent overlapping partnerships on the connectivity and dynamic transmission potential of networks.

Usage

```
data(concurrencyComparisonNets)
```

Format

Three `networkDynamic` objects

`base`   a dynamic network with a poisson cross-sectional degree distribution

`middle` a dynamic network with half the fraction of persons with degree > 1 (having concurrent partners), compared to the base network

`monog` a dynamic network with a bernoulli (0,1) cross-sectional degree distribution

Details

Each network has the following shared characteristics: 1000 nodes, 100 timesteps, a cross-sectional mean degree that varies stochastically around 0.75, and an exponential relationship duration distribution with a mean of 25 timesteps (due to censoring effects, the naive mean duration calculation using all observed partnerships will be around 20). The only difference in the three networks is the cross-sectional degree distribution, varying from Bernoulli (monog) to Poisson (base), which represent a range from strict serial monogamy in partnerships, to the levels of concurrency that would be present if partnerships are formed independently, without regard for any existing partnerships (an Erdos-Renyi graph). This is accomplished by modifying the the formation model of the STERGM used to simulate edge dynamics (see accompanying code for details).

After simulating the dynamic network, a trivial disease simulation is implemented from a single seed in each network, with transmission probability set to 1.0. For each discordant partnership formed, transmission is therefore guaranteed in one timestep, and the infections trace out the size of a forward-reachable component in each network. Note that because the dynamic network is simulated in its entirety first, this implies the partnership formation/dissolution process is independent of the disease state of the node and the network.

Each network has a dynamic 'status' nodal attribute indicating the infection status of each node at each timestep in each network. Comparison of the prevalence and trajectories of the status variable provide insight into the impact of concurrent partnerships on network connectivity and transmission potential. Note that the first infected state does not occur until time 2.

The networks were simulated using the EpiModel package and the code below.

Terms and Conditions: The concurrencyComparisonNets data are provided under the tergms of the Creative Commons Attribution 3.0 License: http://creativecommons.org/licenses/by/3.0/us/

Please cite the dataset authors and the networkDynamicData package (`citation(package='networkDynamicData')`) with any redistribution or published use of this data.

Author(s)

Martina Morris <morrism@uw.edu> and Li Wang <lxwang@gmail.com>

Source

https://statnet.org

References

Morris M., Kurth A., Hamilton D.T., Moody J., and Wakefield S., for The Network Modeling Group (2009) "Concurrent Partnerships and HIV Prevalence Disparities by Race: Linking Science and Public Health Practice" *American Journal of Public Health* 1023-1031, Vol 99, No. 6

Jenness S, Goodreau S, Wang L and Morris M (2014). *EpiModel: Mathematical Modeling of Infectious Disease*. The Statnet Project (https://statnet.org). R package version 0.95, https://CRAN.R-project.org/package=EpiModel.

Examples

```
data(concurrencyComparisonNets)

## Not run:

# compare plots of each network at time 50
plot(network.extract(base,at=50),vertex.cex=0.5,edge.lwd=2)
plot(network.extract(monog,at=50),vertex.cex=0.5,edge.lwd=2)
plot(network.extract(middle,at=50),vertex.cex=0.5,edge.lwd=2)

# compare mean duration. These are at ~20, due to censoring
mean(as.data.frame(base)$duration)
mean(as.data.frame(middle)$duration)
mean(as.data.frame(monog)$duration)

# compare infection time series

plot(sapply(1:100,function(t){
    sum(get.vertex.attribute.active(base,'status',at=t)==1)
  }),col='black',xlab='time step', ylab='# infected'
)
points(sapply(1:100,function(t){
    sum(get.vertex.attribute.active(monog,'status',at=t)==1)
  }),col='blue')
points(sapply(1:100,function(t){
    sum(get.vertex.attribute.active(middle,'status',at=t)==1)
  }),col='red')

## End(Not run)

## The code below can be used generate the three example networks using EpiModel (v1.1.2)
## note that the networks have some attached simulation control variables deleted before
## being saved as the datasets. More recent versions of EpiModel use a different
## representation of the infection status variable.

## Not run:

library(EpiModel)

# === example network parameters setup ===

params.base = list(
```

```
    sim.length = 100,
    size = 1000,
    mean.deg = .75,
    mean.rel.dur = 25,

    net = network.initialize(1000, directed = F),
    formation = ~edges,
    dissolution = ~offset(edges)
)

params.middle = list(
    sim.length = 100,
    size = 1000,
    mean.deg = .75,
    mean.rel.dur = 25,

    net = network.initialize(1000, directed = F),
    formation = ~edges+concurrent,
    dissolution = ~offset(edges),
    targets = 90  # concurrent = 90
)

params.monog = list(
    sim.length = 100,
    size = 1000,
    mean.deg = .75,
    mean.rel.dur = 25,

    net = network.initialize(1000, directed = F),
    formation = ~edges+concurrent,
    dissolution = ~offset(edges),
    targets = 0  # concurrent = 0
)


# === function for estimating stergm, simulating network, and simulating epidemic ===

net.init <- function(params, nsims, adjust=F) {
  for (name in names(params)) assign(name, params[[name]])

  message('network init')

  # generate initial network (defaults if not specified in params)
  if (!exists('net', inherits=F)) {
    net <- network.initialize(size,directed=F)
    net
  }
  if (!exists('formation', inherits=F)) {
    formation = ~edges
  }
  if (!exists('dissolution', inherits=F)) {
    dissolution = ~offset(edges)
  }
```

```
if (!is.null(mean.deg)) {
  target.edges <- size/2 * mean.deg
  density = target.edges / choose(size,2)
} else {
  target.edges <- round(density*choose(size, 2))
}
print(target.edges)

# cludge to fix the monogamy bias in simulate
if (adjust) target.edges = target.edges*adjust

# target stats that does not include edges
if (exists('targets', inherits=F)) {
  target.stats = c(target.edges, targets)
} else {
  target.stats = target.edges
}

coef.diss <- dissolution_coefs(dissolution, mean.rel.dur)

# estimate the stergm

net.est = netest(net, formation, dissolution, target.stats, coef.diss
                    ,set.control.ergm=control.ergm(MCMLE.maxit=200))

stats.form = update(formation, ~.+concurrent)

# simulate the dynamic network
#net.sim = netsim(net.est, nsteps = sim.length, nsims=nsims, stats.form=stats.form,
#                        control = control.simulate.network(MCMC.burnin.add=10))

# simulate the network dynamics and the epidemic
net.sim = netsim(net.est,
                 param.net(inf.prob=1),
                 init.net(i.num=1),
                 control.net(type='SI',
                             nsteps = sim.length,
                             nsims=nsims,
                             keep.network = TRUE)
                )



#trans.sim = epiNet.simTrans(net.sim, "SI", vital=FALSE, i.num=1, trans.rate=1, tea=TRUE)

#print(summary(net.sim$stats[[1]]))
#plot(net.sim$stats[[1]][,'edges'], ylab='edges', xlab='time')

return(get_network(net.sim, sim = 1))
}
```

```
# === simulate example networks ===

base <- net.init(params.base, 1)

middle <- net.init(params.middle, 1)

monog <- net.init(params.monog, 1)


## End(Not run)
```

---

davisDyn   *dynamic versions of the Southern Women Data Set (Davis)*

Description

This is a data set of 18 women observed over a nine-month period. During that period, various subsets of these women had met in a series of 14 informal social events. The data recored which women met for which events. The data is originally from Davis, Gardner and Gardner (1941) via UCINET and stored as a networkDynamic object.

Usage

```
data("davisDyn")
data("davisActorDyn")
```

Format

a networkDynamic data object

Details

This version includes event timings according to the chart extracted by Berger-Wolf from Davis, et al. stored as instantaneous events in numeric POSIX time. In both networks the vertices are marked as 'always active', although the actuall availibility for event membership is not known. This version includes the two (overlapping) group classifications reported by Davis, et al. (via Freeman). The name "Myra" is corrected from the latentnet version of the dataset.

The davisDyn object is a bi-partite network relating the actors to the events.

The davisActorsDyn is one-mode projection of the bipartite network to create a network of the women mutually connected by the events they attend.

The documentation below is taken from Freeman (2003) in his usual lucid description. See the reference to the paper below:

In the 1930s, five ethnographers, Allison Davis, Elizabeth Stubbs Davis, Burleigh B. Gardner, Mary R. Gardner and J. G. St. Clair Drake, collected data on stratification in Natchez, Mississippi (Warner, 1988, p. 93). They produced the book cited below [DGG] that reported a comparative

study of social class in black and in white society. One element of this work involved examining the correspondence between people's social class levels and their patterns of informal interaction. DGG was concerned with the issue of how much the informal contacts made by individuals were established solely (or primarily) with others at approximately their own class levels. To address this question the authors collected data on social events and examined people's patterns of informal contacts.

In particular, they collected systematic data on the social activities of 18 women whom they observed over a nine-month period. During that period, various subsets of these women had met in a series of 14 informal social events. The participation of women in events was uncovered using "interviews, the records of participant observers, guest lists, and the newspapers"" (DGG, p. 149). Homans (1950, p. 82), who presumably had been in touch with the research team, reported that the data reflect joint activities like, "a day's work behind the counter of a store, a meeting of a women's club, a church supper, a card party, a supper party, a meeting of the Parent-Teacher Association, etc."

This data set has several interesting properties. It is small and manageable. It embodies a relatively simple structural pattern, one in which, according to DGG, the women seemed to organize themselves into two more or less distinct groups. Moreover, they reported that the positions - core and peripheral - of the members of these groups could also be determined in terms of the ways in which different women had been involved in group activities. At the same time, the DGG data set is complicated enough that some of the details of its patterning are less than obvious. As Homans (1950, p. 84) put it, "The pattern is frayed at the edges." And, finally, this data set comes to us in a two-mode "woman by event" form. Thus, it provides an opportunity to explore methods designed for direct application to two-mode data. But at the same time, it can easily be transformed into two one-mode matrices (woman by woman or event by event) that can be examined using tools for one-mode analysis.

Because of these properties, this DGG data set has become something of a touchstone for comparing analytic methods in social network analysis. Davis, Gardner and Gardner presented an intuitive interpretation of the data, based in part on their ethnographic experience in the community. Then the DGG data set was picked up by Homans (1950) who provided an alternative intuitive interpretation. In 1972, Phillips and Conviser used an analytic tool, based on information theory, that provided a systematic way to reexamine the DGG data. Since then, this data set has been analyzed again and again. It reappears whenever any network analyst wants to explore the utility of some new tool for analyzing data.

License

Unknown. Based on original publication date, the data are believed to be public domain and have been previously widely circulated in various accademic sources.

Source

This dataset was re-assembled from multiple sources:

Davis, A., Gardner, B. B. and M. R. Gardner (1941) Deep South, Chicago: The University of Chicago Press.

Breiger R. (1974). The duality of persons and groups. Social Forces, 53, 181-190

Linton C. Freeman (2003). Finding Social Groups: A Meta-Analysis of the Southern Women Data, In Ronald Breiger, Kathleen Carley and Philippa Pattison, eds. Dynamic Social Network Modeling

and Analysis. Washington: The National Academies Press. http://intersci.ss.uci.edu/wiki/pub/FreemanSouthernWomen85.pc

Berger-Wolf, T. Y., & Saia, J. (2006). A framework for analysis of dynamic social networks. In Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 523-528). ACM. http://www.cs.unm.edu/~saia/papers/kdd.pdf

Krivitsky P and Handcock M (2015). _latentnet: Latent Position and Cluster Models for Statistical Networks_. The Statnet Project (<URL: http://www.statnet.org>). R package version 2.7.1, <URL: http://CRAN.R-project.org/package=latentnet>.

Examples

```
data(davisDyn)
davisDyn

# convert the dates of the events from numeric seconds
as.POSIXlt(get.change.times(davisDyn),origin="1970-01-01")

data(davisActorsDyn)
davisActorsDyn
```

---

enronEmails   *Enron Emails*

Description

A version of the "Enron Email Network" formatted as a networkDynamic object with edge spells corresponding to individual emails and vertices as email addresses. Data was downloaded form http://www.cis.jhu.edu/~parky/Enron/, with the presumed upstream source of http://www.cs.cmu.edu/~enron/

Usage

```
data("enronEmails")
```

Format

A networkDynamic object.

Details

The edge spells in this network correspond to individual emails sent between 184 addresses in the Enron email corpus. The network is represented as a continuous time event temporal model (onset=terminus). Edge timing is coded as numeric posix time (seconds). The time range is from 315522000 (1979-12-31) to 1024688419 (2002-06-21) but some email timestamps are invalid and most analsyes use the range 1998 (883612800) to 2002. No email content or attachments are included in this version of the dataset.

The vertex ides have been incremented by 1 (compared to the Y. Park version) to follow R's convention of avoid 0-based indices.

Vertex attributes have been attached as follows:

- email_id the non-redundant part of the email (i.e. with @enron.com removed) used as the id in constructing the networks

- role A 'role' associated with the email address (i.e. "Vice President", "Director") (missing and/or redacted for some vertices)

- name The name of the person associated with the email address (missing and/or redacted for some vertices)

- dept A name of the individual's department or subsidiary where known (missing and/or redacted for many vertices)

Original sources

From http://www.cs.cmu.edu/~enron/

"[the Enron email corpus] was collected and prepared by the CALO Project (A Cognitive Assistant that Learns and Organizes). It contains data from about 150 users, mostly senior management of Enron, organized into folders. The corpus contains a total of about 0.5M messages. This data was originally made public, and posted to the web, by the Federal Energy Regulatory Commission during its investigation."

"The email dataset was later purchased by Leslie Kaelbling at MIT, and turned out to have a number of integrity problems. A number of folks at SRI, notably Melinda Gervasio, worked hard to correct these problems, and it is thanks to them (not me [William W. Cohen]) that the dataset is available. The dataset here does not include attachments, and some messages have been deleted "as part of a redaction effort due to requests from affected employees". Invalid email addresses were converted to something of the form user@enron.com whenever possible (i.e., recipient is specified in some parseable format like "Doe, John" or "Mary K. Smith") and to no_address@enron.com when no recipient was specified."

From C.E. Priebe, et al:

"The data are collected from "about 150 users" – mostly Enron executives, but also some energy traders, executive assistants, etc. However, our graphs are based on 184 users, which is the number of unique addresses we obtain from the 'From' line of emails in the 'Sent' boxes after manually removing some addresses which are clearly not associated with the 150 users. [...] In addition, some of the time stamps in the original data are clearly invalid, occurring before Enron existed, so we restrict our attention to a period of 189 weeks, from 1998 through 2002"

License

Creative Commons Attribution Share-Alike license 4.0 https://creativecommons.org/licenses/by-sa/4.0/

Source

downloaded from http://www.cis.jhu.edu/~parky/Enron/ upstream source: http://www.cs.cmu.edu/~enron/

References

C.E. Priebe, J.M. Conroy, D.J. Marchette, and Y. Park, "Scan Statistics on Enron Graphs," Computational and Mathematical Organization Theory, Volume 11, Number 3, p229 - 247, October 2005,

Springer Science+Business Media B.V.. `http://www.cis.jhu.edu/~parky/CEP-Publications/PCMP-CMOT2005.pdf`

Examples

```
data(enronEmails)
enronKnownDates<-network.extract(enronEmails,onset=883612800,terminus=1024688419)
```

---

harry_potter_support   *Harry Potter support networks of Goele Bossaert and Nadine Meidert.*

Description

Goele Bossaert and Nadine Meidert have coded the peer-support ties observed between 64 characters in the the text of the well-known J. K. Rowling fictional novels about Harry Potter.

Usage

```
data(harry_potter_support)
```

Format

The format is is a networkDynamic object with node and edge activity.attributes.

Details

The data in this network was originally collected Goele Bossaert and Nadine Meidert in 2013. They made the data available for general use at `http://www.stats.ox.ac.uk/~snijders/siena/HarryPotterData.html` and it was downloaded and converted to a networkDynamic object.

The data collection is described by the authors as follows:

Contact between the 64 Hogwarts students was coded as peer support when one of the four types of peer support, described in Tardy's model, were found: 1) Student A supports student B emotionally, e.g., in Book 1: Harry, Ron and Hermione assure Neville that he is definitely a Gryffindor when he doubts he is not brave enough to be part of the house; 2) Student A gives students B instrumental help; e.g., in Book 1: Fred and George Weasley help Harry Potter to get his trunk into the compartment of the Hogwarts Express; 3) Student A gives student B certain information to help student B, e.g., in Book 1: Hermione Granger helps Harry Potter with his homework and; 4) Student A praises student B, e.g., in book 5: Terry Boot praises Hermione Granger, for doing a Protean Charm, which is advanced magic. Furthermore, two extra conditions regarding the context in which peer support appeared needed to be fulfilled as well.

First, contact between students was only coded if the peer support was offered voluntarily. Second, only interactions occurring between two living characters, attending Hogwarts at the same moment, were coded as peer support. Consequently, when dead characters reappeared in the books, interactions between these dead characters and living students were not coded. One example for such reappearance is Cedric Diggory's return at the end of book 4, when Cedric asks Harry to return his dead body to his parents. Furthermore, interactions with former or future Hogwarts students at a certain point in time were not included. For example, although Harry and Ginny met before

Ginny attended Hogwarts, peer support relations between both characters were only coded when both students attended Hogwarts together.

The network contains the following vertex attributes:

id: the integer id used by Bossaert and Meidert in their paper

schoolyear: year at which students first attended Hogwarts

gender: 1 is male, 2 is female

house: number indicating which house student was a member of. 1=Gryffindor, 2=Hufflepuff, 3=Ravenclaw, 4=Slytherin

vertex.names: The full name of each student hpnames.txt the names!

Source

Siena datasets: <http://www.stats.ox.ac.uk/~snijders/siena/HarryPotterData.html>

References

Goele Bossaert and Nadine Meidert (2013). 'We are only as strong as we are united, as weak as we are divided'. A dynamic analysis of the peer support networks in the Harry Potter books. *Open Journal of Applied Sciences*, Vol. 3 No. 2, pp. 174-185. doi:10.4236/ojapps.2013.32024

Examples

```
data(harry_potter_support)

# which vertex is Harry Potter?
which(network.vertex.names(harry_potter_support)=="Harry James Potter")
```

| hospital | *Hospital ward dynamic contact network* |

Description

This dataset contains the SocioPatterns temporal network of contacts between patients, patients and health-care workers (HCWs) and among HCWs in a hospital ward in Lyon, France, from Monday, December 6, 2010 at 1:00 pm to Friday, December 10, 2010 at 2:00 pm. The study included 46 HCWs and 29 patients.

Usage

```
data(hospital)
```

Format

The format is is a [networkDynamic](#) object. The network contains the vertex attribute `role` with the values:

- `'NUR'`=paramedical staff, i.e. nurses and nurses' aides;
- `'PAT'`=Patient;
- `'MED'`=Medical doctor;
- `'ADM'`=administrative staff

The `net.obs.period` network attribute describes an observation range from 120 to 347640 seconds. Observations are discrete 20-second intervals.

Details

The details below are excerpted from the paper describing the dataset:

Study Design and Data Collection: Data were collected in a short stay geriatric unit (19 beds) of a university hospital of almost 1000 beds [3] in Lyon, France, from Monday, December 6, 2010 at 1:00 pm to Friday, December 10, 2010 at 2:00 pm. During that time, 50 professional staff worked in the unit and 31 patients were admitted. We collected data on the contacts between 46 staff members (92% participation rate) and 29 patients (94% participation rate). The participating staff members were 27 nurses or nurses' aides, 11 medical doctors and 8 administrative staff.

In the ward, all rooms but 2 were single-bed rooms. Each day 2 teams of 2 nurses and 3 nurses' aides worked in the ward: one of the teams was present from 7:00 am to 1:30 pm and the other from 1:30 pm to 8:00 pm. An additional nurse and an additional nurse' aid were moreover present from 9:00 am to 5:00 pm. Two nurses were present during the nights from 8:00 pm to 7:00 am. In addition, a physiotherapist and a nutritionist were present each day at various points in time, with no fixed schedule, and a social counselor and a physical therapist visited on demand (in our analysis they are considered as nurses). Two physicians and 2 interns were present from 8:00 am to 17:00 pm each day. Visits were allowed from 12:00 am to 8:00 pm but visitors were not included in the study.

The measurement system, developed by the SocioPatterns collaboration, is based on small active RFID devices ("tag") that are embedded in unobtrusive wearable badges and exchange ultra-low-power radio packets. The power level is tuned so that devices can exchange packets only when located within 1-1.5 meters of one another, i.e., package exchange is used as a proxy for distance (the tags do not directly measure distances). Individuals were asked to wear the devices on their chests using lanyards, ensuring that the RFID devices of two individuals can only exchange radio packets when the persons are facing each other, as the human body acts as a RF shield at the frequency used for communication. In summary the system is tuned so that it detects and records close-range encounters during which a communicable disease infection could be transmitted, for example, by cough, sneeze or hand contact. The information on face-to-face proximity events detected by the wearable sensors is relayed to radio receivers installed throughout the hospital ward (bedrooms, offices and hall).

The system was tuned so that whenever two individuals wearing the RFID tags were in face-to-face proximity the probability to detect such a proximity event over an time interval of 20 seconds was larger than 99%. We therefore define two individuals to be in "contact" during a 20-second interval if and only if their sensors exchanged at least one packet during that interval. A contact is therefore symmetric by definition, and in case of contacts involving three or more individuals in

the same 20-second interval, all the contact pairs were considered. After the contact is established, it is considered ongoing as long as the devices continue to exchange at least one packet for every subsequent 20 s interval. Conversely, a contact is considered broken if a 20-second interval elapses with no exchange of packets. We emphasize that this is an operational definition of the human proximity behavior that we choose to quantify, and that all the results we present correspond to this precise and specific definition of "contact". We make the raw data we collected available to the public as Datasets S1-S5 in File S1 and on the website of the SocioPatterns collaboration (www. sociopatterns.org).

Terms and conditions: The data are distributed to the public under a Creative Commons Attribution-NonCommercial-ShareAlike license http://creativecommons.org/licenses/by-nc-sa/3.0/. When this data is used in published research or for visualization purposes, please cite the following paper: P. Vanhems et al., Estimating Potential Infection Transmission Routes in Hospital Wards Using Wearable Proximity Sensors, PLoS ONE 8(9): e73970 (2013). Please also acknowledge the SocioPatterns collaboration and provide a link to http://www.sociopatterns.org.

Author(s)

Philippe Vanhems, Alain Barrat, Ciro Cattuto, Jean-Francois Pinton, Nagham Khanafer, Corinne Regis, Byeul-a Kim, Brigitte Comte, Nicolas Voirin. info@sociopatterns.org

Source

http://www.sociopatterns.org/datasets/hospital-ward-dynamic-contact-network/ https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0073970

References

P. Vanhems et al., Estimating Potential Infection Transmission Routes in Hospital Wards Using Wearable Proximity Sensors, PLoS ONE 8(9): e73970 (2013). https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0073970

Examples

```
data(hospital_contact)
## Not run:
# get an overview of temporal density
plot(table(get.edge.activity(hospital,as.spellList=TRUE)[,1]),
xlab='time',ylab='activity count',col="#00000055")

# define a mapping of roles to colors
rolecolors<-function(roles){
  roles[roles=='NUR']<-'blue'
  roles[roles=='PAT']<-'gray'
  roles[roles=='MED']<-'green'
  roles[roles=='ADM']<-'orange'
  return(roles)
}

# network plot aggregating across all days
```

```
slice<-network.collapse(hospital,rm.time.info=FALSE)
plot(slice,vertex.col=rolecolors(slice%v%'role'),
  edge.lwd=slice%e%'activity.duration'/300,edge.col='#00000044')

# plot one hour of interaction
plot(network.extract(hospital,onset=160000,terminus=163600),
vertex.cex=0.5,vertex.col=rolecolors(hospital%v%'role'))

## End(Not run)
```

---

manufacturingEmails   *Internal Emails from a Polish Manufacturing Company*

Description

The source is a longitudinal network describing the history of internal e-mail communication (sender, recipient, datetime) between 167 employees of a mid-sized manufacturing company located in Poland. Multiple recipients of the same e-mail (To, CC, BCC) are represented as separate rows without distinguishing the recipient type. The period covered are nine full months of 2010 starting from 2010-01-01 to 2010-09-30 (event dates in local time). Apart from the communication, information about who in the company reports to whom is included . Node #86 is the CEO (the only loop in the graph).

Usage

```
data("manufacturingEmails")
```

Format

a networkDynamic object

Details

This dataset consists of two network objects:

The manufacturingEmails network is a networkDynamic object with 82614 edge spells (emails communications) between 176 employees. The network is represented as a continuous time event temporal model (onset=terminus). Edge timing is coded as numeric POSIX time (seconds) with event dates in local time ranging from 1262482810 (2010-01-01) until 1285909692 (2010-09-30). The network contains self-loops. Duplicate rows in the input data (email to the same recipient at the same second using TO, CC, BCC etc) have been collapsed but this information is preserved in the numEmailTypes dynamic edge attribute. The networks included here have a much larger vertex set and so do not correspond exactly to the description in the paper (below).

The manufacturingReportsTo network a static network object which includes the organizational hierarchy. Note that vertices 4, 10, 21, 23, 24, 26 and 46 are technical email accounts not used by employees, and vertices 51, 75, 87, 93, 111 and 139 are email accounts corresponding to former employees and so appear as isolates in the manufacturingReportsTo network.

Description from paper:

... company is a manufacturing company located in Poland. The company employs 300 persons, whereas 1/3 are clerical workers, the rest - laborers. The period analyzed was half a year. The type of organizational structure is functional [3]. However, due to organization operating model and its consequences to organizational structure clarity as well as logs interpretation possibility, only a subset of organization have been chosen for current analysis: 49 clerical employees not directly related to manufacturing process. Three-level management structure exists in the selected company part: management board (2 persons), managers (11 persons) and regular employees (36 persons) and they work in twelve different departments. There were no organizational changes during the analyzed period. Email logs were source data used to build social network . Because of email logs structure, there was no distinction between To, CC and BCC recipients. The resulting set of data contained 11,816 emails in total.

License

The data are distributed to the public under a Creative Commons Attribution-NonCommercial-ShareAlike license http://creativecommons.org/licenses/by-nc-sa/3.0/.

Source

radoslaw.michalski <at> pwr.edu.pl https://www.ii.pwr.wroc.pl/~michalski/index.php?content=datasets and

References

When using this dataset, please cite:

Michalski, R., Kajdanowicz, T., Brodka, P., Kazienko, P.: Seed Selection for Spread of Influence in Social Networks: Temporal vs. Static Approach. New Generation Computing (JCR-listed journal), Vol. 32, Issue 3-4, pp. 213-235. Ohmsha-Japan and Springer (2014))

```
@article{michalski2014seed,
  title={Seed Selection for Spread of Influence in Social Networks: Temporal vs. Static Approach},
  author={Michalski, Rados{\l}aw and Kajdanowicz, Tomasz and Br{\'o}dka, Piotr and Kazienko, Przemys{\l
  journal={New Generation Computing},
  volume={32},
  number={3-4},
  pages={213--235},
  year={2014},
  publisher={Springer}
}
```

Michalski, R., Palus, S., Kazienko, P.: Matching Organizational Structure and Social Network Extracted from Email Communication. Lecture Notes in Business Information Processing LNBIP, vol. 87, pp. 197-206, Springer, Berlin Heidelberg (2011)

Examples

```
data(manufacturingEmails)
## Not run:
# plot the organizational hierarchy
```

```
plot(manufacturingReportsTo,displaylabels=TRUE,
     vertex.cex=0.6,label.cex=0.6,edge.col='gray')

# plot the first two days of emails
plot(network.extract(manufacturingEmails,
     onset=1262482810,length=60*60*24*2))

# plot email density over time
plot(density(as.data.frame(manufacturingEmails)$onset))

# convert date string to POSIX seconds
as.numeric(as.POSIXct('2010-09-30',format='%Y-%m-%d'))

# convert POSIX seconds to date string
as.POSIXct(1285830000,origin='1970-01-01',tz = 'PL')

## End(Not run)
```

---

onlineNet   *UCI Facebook-like Social Network*

Description

The Facebook-like Social Network originated from an online community for students at University
of California, Irvine. The dataset includes the users that sent or received at least one message
(1,899). A total number of 59,835 online messages were sent over 20,296 directed ties among these
users over a period of six months.

Usage

```
data(onlineNet)
```

Format

The format is is a [networkDynamic](#) object.The `net.obs.period` network attribute describes an
observation range from timestamp 1080101515 to 1098777142. The original dates and times were
converted to POSIXct timestamps during import. The original time range is "2004-03-23 20:11:55"
to "2004-10-26 00:52:22"

Details

This network is described in Patterns and Dynamics of Users' Behaviour and Interaction: Network
Analysis of an Online Community and used in a number of articles including Prominence and
control: The weighted rich-club effect and Clustering in weighted networks. Although this dataset
contains many nodal attributes (e.g., gender, age, and course attended), these are not made available
as it would be possible to reverse engineer the anonymisation procedure of users. Self-loops in
the original longitudinal edgelist signal the time that users registered on the site, these have been
converted into vertex onset times and removed.

As this dataset excludes isolated vertices, for analyses involving degree distributions it may be desirable to add them back in. There were a total of 2595 users that logged in at least once – and 2995 users that filled in the registration form (might not have validated their email etc).

This dataset is also included in the CRAN package `tnet`.

Author(s)

Tore Opsahl; https://toreopsahl.com

Source

https://toreopsahl.com/datasets/#online_social_network

References

Opsahl, T., Panzarasa, P., 2009. Clustering in weighted networks. Social Networks 31 (2), 155-163, doi: 10.1016/j.socnet.2009.02.002

Panzarasa, P., Opsahl, T., Carley, K.M., 2009. Patterns and dynamics of users' behavior and interaction: Network analysis of an online community. Journal of the American Society for Information Science and Technology 60 (5), 911-932

https://toreopsahl.com/2009/03/06/article-patterns-and-dynamics-of-users-behaviour-and-interaction-

Examples

```
data(onlineNetwork)

# convert timestamp to human-readable
as.POSIXct(1080101515,origin = "1970-01-01")

# plot number in network over time
plot(sapply(seq(from=1080101515, to=1098777142,length.out=100),
     function(t){
       network.size.active(onlineNet,at=t)
     }
     ),ylab= '# members'
)
```

---

vanDeBunt_students	*van de Bunt longitudinal student friendship dataset*

Description

Multiple levels of friendship ties among students reported at 7 time points. This data set was collected by Gerhard van de Bunt, and is discussed extensively in van de Bunt (1999) and van de Bunt, van Duijn, and Snijders (1999). It is used as example in the manual and in various methodological articles about SIENA.

Usage

    data(vanDeBunt_students)

Format

The format is is a networkDynamic object with node and edge activity.attributes.

Details

The dataset was acquired from http://www.stats.ox.ac.uk/~snijders/siena/vdBunt_data. htm. The information below is adapted from the description page:

The data were collected among a group of university freshmen who, except for a few existing relationships (acquaintances from a former school), did not know each other at the first measurement (time=t0). The data were collected at 7 time points. The first four time points are three weeks apart, whereas the last three time points are six weeks apart. The original group consisted of 49 students, but due to 'university drop-outs' and after deleting those who did not fill in the questionnaire four or more times, a group was obtained of 32 students for whom almost complete data are available.

The students were asked to rate their relationships on a six point scale, with response categories described as follows.

1. Best friendship Persons whom you would call your 'real' friends

2. Friendship Persons with whom you have a good relationship, but whom you do not (yet) consider a 'real' friend

3. Friendly relationship Persons with whom you regularly have pleasant contact during classes. The contact could grow into a friendship

4. Neutral relationship Persons with whom you have not much in common. In case of an accidental meeting the contact is good. The chance of it growing into a friendship is not large

0. Unknown person Persons whom you do not know

5. Troubled relationship Persons with whom you can't get on very well, and with whom you definitely do not want to start a relationship. There is a certain risk of getting into a conflict

NOTE: in the import process, ties are not created for value 0. Also, in the original matrix values "6 = item non-response, 9 = actor non-response" These missing data codes were translated to NA values, but not well represented by the conversion to a networkDynamic as we have not set the spec for dynamic missingness yet.

Next to the sociometric data, available individual characteristics are sex, education program, and smoking behavior. Smoking was only allowed in special areas. As a consequence, the 'smokers' had to separate themselves from the 'non-smokers' if they wished to smoke (which they often did during coffee and lunch breaks). Thus, contact opportunities differed between actors because of their smoking behavior. The education program was important because, although all started to study at the same moment, there were three groups, following different courses. During the first months all programs overlapped largely, but after a few months, the programs diverged. Especially the 2-year program was quite different from the other two programs. Therefore, this attribute also gives information on the individuals' contact opportunities. See the references mentioned belo for further information about this network and the actor attributes.

The network contains the following vertex attributes:

gender: integer, 1 is female, 2 is male

program: integer, 2=2-year, 3=3-year, 4=4-year

smoking: integer,1 = yes, 2 = no

vertex.names: integer matrix id used in the input files

As this network was observed at unequal intervals, the `net.obs.period` attribute provides information on the duration between observation windows.

Source

Siena datasets: http://www.stats.ox.ac.uk/~snijders/siena/vdBunt_data.htm

References

Van de Bunt, G.G. 1999. *Friends by choice. An actor-oriented statistical network model for friendship networks through time.* Amsterdam: Thesis Publishers.

Van de Bunt, G.G., M.A.J. van Duijn, and T.A.B. Snijders. 1999. Friendship networks through time: An actor-oriented statistical network model. *Computational and Mathematical Organization Theory*, 5, 167-192.

Examples

```
data(vanDeBunt_students)
```

Index