# Package 'surveygraph'

November 9, 2025

**Title** Network Representations of Attitudes

**Version** 1.0.0

**Author** Samuel Unicomb [aut, cre],
Ana Jovancevic [aut],
Caoimhe O'Reilly [aut],
Alejandro Dinkelberg [aut],
Pádraig MacCarron [aut],
David O'Sullivan [aut],
Paul Maher [aut],
Mike Quayle [aut]

**Maintainer** Samuel Unicomb <samuelunicomb@gmail.com>

**Description** A tool for computing network representations of attitudes,
extracted from tabular data such as sociological surveys. Development of
surveygraph software and training materials was initially funded by the
European Union under the ERC Proof-of-concept programme (ERC,
Attitude-Maps-4-All, project number: 101069264). Views and opinions expressed
are however those of the author(s) only and do not necessarily reflect those
of the European Union or the European Research Council Executive Agency.
Neither the European Union nor the granting authority can be held responsible
for them.

**Depends** R (>= 2.15.1)

**URL** https://github.com/surveygraph/surveygraphr,
https://surveygraph.ie/

**BugReports** https://github.com/surveygraph/surveygraphr/issues

**License** MIT + file LICENSE

**Encoding** UTF-8

**RoxygenNote** 7.3.2

**Suggests** covr, knitr, rmarkdown, testthat (>= 3.0.0)

**VignetteBuilder** knitr

**NeedsCompilation** yes

**Repository** CRAN

**Date/Publication** 2025-11-09 22:20:02 UTC

# Contents

---

data_preprocess              *Outputs a synthetic survey using a simple model*

---

## Description

data_preprocess() outputs a synthetic survey, generated using a simple, stochastic model of polarisation.

## Usage

```
data_preprocess(data, limits = NULL, dummycode = NULL)
```

## Arguments

| data | The number of rows in the survey |
|------|----------------------------------|
| limits | The number of columns in the survey |
| dummycode | The fraction of nodes in the smaller of the two polarised groups |

## Value

A data frame corresponding to a survey.

## Examples

```
S <- make_synthetic_data(200, 8)
```

---

make_projection *Outputs the survey projection onto the agent or symbolic layer*

---

### Description

make_projection() outputs the agent or symbolic network corresponding to a survey, i.e. the row or column projection.

### Usage

```
make_projection(
  data,
  layer = NULL,
  method = NULL,
  methodval = NULL,
  comparisons = NULL,
  metric = NULL,
  limits = NULL,
  dummycode = NULL,
  bootreps = NULL,
  bootval = NULL,
  bootseed = NULL,
  centre = NULL,
  ...
)
```

### Arguments

| | |
|---|---|
| data | A data frame corresponding to a survey |
| layer | A string flag specifying which layer to project |

- "agent" produces the network corresponding to the agents, which we assume to be rows in data
- "symbolic" produces the network corresponding to the symbols, or items, which we assume to be columns in data

| | |
|---|---|
| method | A string flag specifying how edges are thresholded in the network representation. |

- "similarity" means we remove all edges whose weight, meaning node similarity, is below a threshold specified by methodval.
- "lcc" finds the value of the threshold that results in the network whose largest connected component is as close as possible to a specified value. In general a range of thresholds will satisfy this condition, and we choose the upper limit of this range. As such, "lcc" provided is a target.
- "avgdegree" finds the value of the threshold that results in the network whose average degree is as close as possible to a specified value. Like "lcc", this is a target.

methodval        A utility variable that we interpret according to the method chosen.

- If method = "similarity", then methodval is interpreted as the similarity threshold, and thus is in the range [0, 1]. A value of 0 means no edges are removed, and a value of 1 means all edges are removed.
- If method = "lcc", then methodval is interpreted as the desired fractional size of the largest connected component, in the range [0, 1]. E.g., when set to 0, no nodes are connected, and if set to 1, the network is as sparse as possible while remaining fully connected.
- If method = "avgdegree", then methodval is interpreted as the desired average degree. We assume that methodval is normalised to the range [0, 1] When method_value = 0, then no nodes are connected, and if method_value = 1, the network is complete, meaning it contains every possible edge.

comparisons      The minimum number of valid comparisons that must be made when computing the similarity between rows or columns in the data. If at least one of the entries in the fields being compared is NA, then the comparison is invalid.

metric           This currently has just one allowed value, namely the Manhattan distance, which is the default.

limits           Specifies the limits of the Likert scale contained in data.

dummycode        flag that indicates whether we dummycode data.

bootreps         The number of bootstrap realisations to perform. If not specified, bootstrapping is not carried out.

bootval          A sampling probability used when bootstraping. In particular, it provides the probability of sampling a given survey entry in a given bootstrapping step. With probability 1 - bootval, that entry is set to NA.

bootseed         A random number generator seed used when bootstrapping. Mainly used for testing, but maybe useful for reproducibility in general.

centre           If TRUE, we shift edge weights from [0, 1] to [-1, 1]. Defaults to FALSE, as most network analysis applications require positive edge weights.

...              Mostly used to handle deprecated arguments, and arguments with alternative spellings.

## Value

A data frame corresponding to the edge list of the specified network. It contains three columns named

- u, the first node adjacent to the edge
- v, the second node adjacent to the edge, and
- weight, the similarity between nodes u and v

## Examples

```
S <- make_synthetic_data(20, 5)
```

---

make_synthetic_data *Outputs a synthetic survey using a simple model*

---

### Description

make_synthetic_data() outputs a synthetic survey, generated using a simple, stochastic model of polarisation.

### Usage

```
make_synthetic_data(
  nrow,
  ncol,
  minority = NULL,
  correlation = NULL,
  polarisation = NULL,
  likert = NULL,
  seed = NULL,
  ...
)
```

### Arguments

| | |
|---|---|
| nrow | The number of rows in the survey |
| ncol | The number of columns in the survey |
| minority | The fraction of nodes in the smaller of the two polarised groups |
| correlation | Probability that group item corresponds to polarisation |
| polarisation | The degree of polarisation among the system's agents |
| likert | Range of the Likert scale |
| seed | Seed value for random number generation. |
| ... | Mostly used to handle arguments with alternative spellings. |

### Value

A data frame corresponding to a survey.

### Examples

```
S <- make_synthetic_data(200, 8)
```

---

make_threshold_profile

*Illustrates how network properties vary with the similarity threshold*

---

### Description

make_threshold_profile() outputs properties of the agent or symbolic network as a function of similarity threshold.

### Usage

```
make_threshold_profile(
  data,
  layer = NULL,
  comparisons = NULL,
  metric = NULL,
  count = NULL,
  limits = NULL,
  dummycode = NULL,
  ...
)
```

### Arguments

| | |
|---|---|
| data | A data frame corresponding to the attitudes held by agents with respect to a number of items |
| layer | A string flag specifying the type of network to be extracted, |
| | • "agent" produces the network corresponding to the agents, which we assume to be rows in data |
| | • "symbolic" produces the network corresponding to the symbols, or items, which we assume to be columns in data |
| comparisons | An integer, minimum number of comparisons for valid distance. |
| metric | A string option describing the similarity metric to be used. |
| count | The number of threshold values to include in the description. |
| limits | Specify the limits of the Likert range in during a data preprocessing step. |
| dummycode | Specify whether to apply dummycoding during a data preprocessing step. |
| ... | Used to handle alternative argument spellings. |

### Details

Note that this routine is expensive on large graphs. We study networks over the full range of similarity thresholds [-1, 1], and as a result, produce networks that are complete at the lower limit of that range. Note that by default we will subsample the provided survey with the C++ implementation in order to avoid memory issues. We could then allow a flag that turns off the subsampling step, at the user's peril.

**Value**

A data frame containing properties of the `agent` or `symbolic` network as a function of the similarity threshold. In particular, it contains three columns named

- `threshold`, the value of the similarity threshold
- `ad`, the average degree resulting from `threshold`, and
- `lcc`, the size of the largest connected component resulting from `threshold`

**Examples**

```
S <- make_synthetic_data(20, 5)
```

---

surveygraph                           *surveygraph: network representations of attitudes*

---

**Description**

This page is a work in progress.

**Details**

The surveygraph package functions can be summarised as follows.

**reading functions**

The reading functions import survey datasets to R so they can be passed to C++ routines. A bunch of file formats need to be accounted for.

**network generating functions**

The network functions are implemented in C++.

**Author(s)**

**Maintainer**: Samuel Unicomb <samuelunicomb@gmail.com>

Authors:

- Ana Jovancevic
- Caoimhe O'Reilly
- Alejandro Dinkelberg
- Pádraig MacCarron
- David O'Sullivan
- Paul Maher
- Mike Quayle

**See Also**

Useful links:

- <https://github.com/surveygraph/surveygraphr>
- <https://surveygraph.ie/>
- Report bugs at <https://github.com/surveygraph/surveygraphr/issues>

# Index