

# Package ‘surveygraph’

April 2, 2024

**Title** Network Representations of Attitudes

**Version** 0.1.1

**Author** Samuel Unicomb [aut, cre],  
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. By treating a survey as a bipartite network, we measure the similarity between respondents and survey items to produce network edges. We do this in both a respondent network, as well as a survey item network. Used in combination with graph visualisation libraries, this technique helps practitioners in the social sciences identify network structure that may be present within a survey.

**Depends** R (>= 2.15.1)

**URL** <https://surveygraph.ie/>

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

**License** MIT + file LICENSE

**Encoding** UTF-8

**RoxygenNote** 7.2.3

**Suggests** covr, ggplot2, igraph, knitr, rmarkdown, testthat (>= 3.0.0)

**NeedsCompilation** yes

**Repository** CRAN

**Date/Publication** 2024-04-02 12:10:02 UTC

## R topics documented:

make_projection	2
make_synthetic_data	3
make_threshold_profile	4
surveygraph	5

<b>Index</b>	<b>6</b>
--------------	----------

---

make_projection	<i>Outputs the survey projection onto the agent or symbolic layer</i>
-----------------	---

---

### 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,
  threshold_method = NULL,
  method_value = NULL,
  centre = NULL,
  similarity_metric = NULL
)
```

### Arguments

data	A data frame corresponding to a survey
layer	A string flag specifying which layer to project <ul style="list-style-type: none"> <li>• "agent" produces the network corresponding to the agents, which we assume to be rows in data</li> <li>• "symbolic" produces the network corresponding to the symbols, or items, which we assume to be columns in data</li> </ul>
threshold_method	A string flag specifying how edges are selected in the network representation. <ul style="list-style-type: none"> <li>• "raw_similarity" means we remove all edges whose weight, meaning node similarity, is below a specified threshold.</li> <li>• "target_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.</li> <li>• "target_ad" finds the value of the threshold that results in the network whose average degree is as close as possible to a specified value.</li> </ul>
method_value	A utility variable that we interpret according to the threshold_method chosen.

- If `threshold_method = "raw_similarity"`, then `method_value` is interpreted as the similarity threshold, and thus is in the range  $[-1, 1]$ . A value of -1 means no edges are removed, and a value of 1 means all edges are removed.
- If `threshold_method = "target_lcc"`, then `method_value` 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 `threshold_method = "target_ad"`, then `method_value` is interpreted as the desired average degree. We assume that `method_value` 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.

`centre` If FALSE, we shift edge weights by 1 from  $[-1, 1]$  to  $[0, 2]$ . Defaults to TRUE.

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

### 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 = 0.5,
  polarisation = 0,
```

```

    correlation = 0.85,
    scale = 10
  )

```

### 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
polarisation	The degree of polarisation among the system's agents
correlation	Probability that group item corresponds to polarisation
scale	Range of the Likert scale

### 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)
```

### 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, <ul style="list-style-type: none"> <li>• "agent" produces the network corresponding to the agents, which we assume to be rows in data</li> <li>• "symbolic" produces the network corresponding to the symbols, or items, which we assume to be columns in data</li> </ul>

**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**

The surveygraph package provides the following functions....

**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++, and blah blah blah.

# Index

make\_projection, 2  
make\_synthetic\_data, 3  
make\_threshold\_profile, 4  
surveygraph, 5