R_code_manuscript.R

mghancean

2021-02-18

```
# Exploring the formation processes of political discussion networks ####
# Hancean, M.-G., Gheorghita, A., Vega Yon, G.G., Mihaila, B.-E.
#
library(easypackages)
libraries("statnet", "igraph", "intergraph", "ergmito", "texreg", "dplyr", "ggpubr")
```

```
## Loading required package: statnet
```

```
## Loading required package: tergm
```

```
## Loading required package: ergm
```

```
## Loading required package: network
```

```
## network: Classes for Relational Data
## Version 1.16.1 created on 2020-10-06.
## copyright (c) 2005, Carter T. Butts, University of California-Irvine
##                     Mark S. Handcock, University of California -- Los Angeles
##                     David R. Hunter, Penn State University
##                     Martina Morris, University of Washington
##                     Skye Bender-deMoll, University of Washington
##  For citation information, type citation("network").
##  Type help("network-package") to get started.
```

```
##
## ergm: version 3.11.0, created on 2020-10-14
## Copyright (c) 2020, Mark S. Handcock, University of California -- Los Angeles
##                     David R. Hunter, Penn State University
##                     Carter T. Butts, University of California -- Irvine
##                     Steven M. Goodreau, University of Washington
##                     Pavel N. Krivitsky, UNSW Sydney
##                     Martina Morris, University of Washington
##                     with contributions from
##                     Li Wang
##                     Kirk Li, University of Washington
##                     Skye Bender-deMoll, University of Washington
##                     Chad Klumb
##                     Michal Bojanowski, Kozminski University
##                     Ben Bolker
## Based on "statnet" project software (statnet.org).
## For license and citation information see statnet.org/attribution
## or type citation("ergm").
```

```
## NOTE: Versions before 3.6.1 had a bug in the implementation of the bd()
## constraint which distorted the sampled distribution somewhat. In
## addition, Sampson's Monks datasets had mislabeled vertices. See the
## NEWS and the documentation for more details.

## NOTE: Some common term arguments pertaining to vertex attribute and
## level selection have changed in 3.10.0. See terms help for more
## details. Use 'options(ergm.term=list(version="3.9.4"))' to use old
## behavior.

## Loading required package: networkDynamic

##
## networkDynamic: version 0.10.1, created on 2020-01-16
## Copyright (c) 2020, Carter T. Butts, University of California -- Irvine
##                     Ayn Leslie-Cook, University of Washington
##                     Pavel N. Krivitsky, University of Wollongong
##                     Skye Bender-deMoll, University of Washington
##                     with contributions from
##                     Zack Almquist, University of California -- Irvine
##                     David R. Hunter, Penn State University
##                     Li Wang
##                     Kirk Li, University of Washington
##                     Steven M. Goodreau, University of Washington
##                     Jeffrey Horner
##                     Martina Morris, University of Washington
## Based on "statnet" project software (statnet.org).
## For license and citation information see statnet.org/attribution
## or type citation("networkDynamic").

##
## tergm: version 3.7.0, created on 2020-10-15
## Copyright (c) 2020, Pavel N. Krivitsky, UNSW Sydney
##                     Mark S. Handcock, University of California -- Los Angeles
##                     with contributions from
##                     David R. Hunter, Penn State University
##                     Steven M. Goodreau, University of Washington
##                     Martina Morris, University of Washington
##                     Nicole Bohme Carnegie, New York University
##                     Carter T. Butts, University of California -- Irvine
##                     Ayn Leslie-Cook, University of Washington
##                     Skye Bender-deMoll
##                     Li Wang
##                     Kirk Li, University of Washington
##                     Chad Klumb
## Based on "statnet" project software (statnet.org).
## For license and citation information see statnet.org/attribution
## or type citation("tergm").

## Loading required package: ergm.count

##
```

```
## ergm.count: version 3.4.0, created on 2019-05-15
## Copyright (c) 2019, Pavel N. Krivitsky, University of Wollongong
##                     with contributions from
##                     Mark S. Handcock, University of California -- Los Angeles
##                     David R. Hunter, Penn State University
## Based on "statnet" project software (statnet.org).
## For license and citation information see statnet.org/attribution
## or type citation("ergm.count").

## NOTE: The form of the term 'CMP' has been changed in version 3.2 of
## 'ergm.count'. See the news or help('CMP') for more information.

## Loading required package: sna

## Loading required package: statnet.common

##
## Attaching package: 'statnet.common'

## The following object is masked from 'package:base':
##
##     order

## sna: Tools for Social Network Analysis
## Version 2.6 created on 2020-10-5.
## copyright (c) 2005, Carter T. Butts, University of California-Irvine
##  For citation information, type citation("sna").
##  Type help(package="sna") to get started.

## Loading required package: tsna

##
## statnet: version 2019.6, created on 2019-06-13
## Copyright (c) 2019, Mark S. Handcock, University of California -- Los Angeles
##                     David R. Hunter, Penn State University
##                     Carter T. Butts, University of California -- Irvine
##                     Steven M. Goodreau, University of Washington
##                     Pavel N. Krivitsky, University of Wollongong
##                     Skye Bender-deMoll
##                     Martina Morris, University of Washington
## Based on "statnet" project software (statnet.org).
## For license and citation information see statnet.org/attribution
## or type citation("statnet").

## unable to reach CRAN

## Loading required package: igraph

##
## Attaching package: 'igraph'
```

```
## The following objects are masked from 'package:sna':
##
##     betweenness, bonpow, closeness, components, degree, dyad.census,
##     evcent, hierarchy, is.connected, neighborhood, triad.census

## The following objects are masked from 'package:network':
##
##     %c%, %s%, add.edges, add.vertices, delete.edges, delete.vertices,
##     get.edge.attribute, get.edges, get.vertex.attribute, is.bipartite,
##     is.directed, list.edge.attributes, list.vertex.attributes,
##     set.edge.attribute, set.vertex.attribute

## The following objects are masked from 'package:stats':
##
##     decompose, spectrum

## The following object is masked from 'package:base':
##
##     union

## Loading required package: intergraph

## Loading required package: ergmito

##
## Attaching package: 'ergmito'

## The following object is masked from 'package:igraph':
##
##     is_directed

## Loading required package: texreg

## Version:  1.37.5
## Date:     2020-06-17
## Author:   Philip Leifeld (University of Essex)
##
## Consider submitting praise using the praise or praise_interactive functions.
## Please cite the JSS article in your publications -- see citation("texreg").

## Loading required package: dplyr

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:igraph':
##
##     as_data_frame, groups, union

## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

## Loading required package: ggpubr

## Loading required package: ggplot2

## All packages loaded successfully
```

```
# 1. read data ####
read_nets <- function(efiles, nfiles) {

  # Checking the dimensions
  if (length (efiles) != length(nfiles))
    stop("The length of the lists do not match.")

  n <- length (efiles)
  out <- vector("list", n)

  for (i in 1:n) {

    # Checking whether the files exists
    if (! file.exists(efiles[i]))
      stop("The file ", efiles[i], "does not exist.")
    if (!file.exists(nfiles[i]))
      stop("The file ", nfiles[i], "does not exist.")

    e1.edgelist <- read.csv(efiles[i], header = FALSE, as.is = TRUE) # read edgelist
    e1.nodes <- read.csv(nfiles[i], header = TRUE, as.is = TRUE) # read node set

    e1.igraph <- graph_from_data_frame(
      d = e1.edgelist,
      vertices = e1.nodes,
      directed = F
    ) # create igraph object

    out[[i]] <- intergraph::asNetwork(e1.igraph)

  }

  return(out)
}

# Listing files
efiles <- list.files(path = "F:/manuscripts/adrian/SUBMISSION/data", pattern = "e[0-9]+dir\\.csv", full
nfiles <- gsub("(e[0-9]+)dir\\.csv", "\\1n.csv", efiles)
all_nets <- read_nets(efiles, nfiles)

# Adding extra information
nafiles <- gsub("(e[0-9]+)dir\\.csv", "\\1na.csv", efiles)
for (i in seq_along(all_nets)) {
  f <- read.csv(nafiles[i])
```

```
  all_nets[[i]] %v% "class_name" <- f[["class.name"]]
  all_nets[[i]] %n% "class_name" <- f[["class.name"]][1]
}
# Network of size 2 not very useful
nets_4_5 <- all_nets[nvertex(all_nets) != 2]
# 3. ergmito models ####
# Model 1
model_1 <- ergmito(
  nets_4_5 ~ edges +
    triangle
  # kstar(2)
  ,model_update = ~
    I(edges * (class_name %in% "adult")) +
    I(edges * (class_name %in% "senior"))
  # I(edges * (n == 5))
)
```

```
## Warning: ergmito does not fully support undirected graphs (yet). We will
## continue with the estimation process, but simulation has limited supported for
## now.
```

```
## Warning: The observed statistics (target.statistics) are near or at the boundary
## of its support, i.e. the Maximum Likelihood Estimates maynot exist or be hard
## to be estimated. In particular, the statistic(s) "edges", "triangle", "I(edges *
## (class_name %in% "adult"))", "I(edges * (class_name %in% "senior"))".
```

```
texreg::screenreg(list(model_1))
```

```
##
## ================================================
##                                      Model 1
## ------------------------------------------------
## edges                                  -0.86 ***
##                                        (0.20)
## triangle                                1.17 ***
##                                        (0.14)
## edges * (class_name %in% "adult")       0.01
##                                        (0.16)
## edges * (class_name %in% "senior")      0.04
##                                        (0.16)
## ------------------------------------------------
## AIC                                   328.33
## BIC                                   343.01
## Log Likelihood                       -160.16
## Num. networks                          29
## Convergence                             0
## Time (seconds)                          0.97
## ================================================
## *** p < 0.001; ** p < 0.01; * p < 0.05
```

```r
# Model 2
model_2 <- ergmito(
  nets_4_5 ~ edges +
    # triangle
    kstar(2)
  ,model_update = ~
    I(edges * (class_name %in% "adult")) +
    I(edges * (class_name %in% "senior"))
  # I(edges * (n == 5))
)
```

```
## Warning: ergmito does not fully support undirected graphs (yet). We will
## continue with the estimation process, but simulation has limited supported for
## now.
```

```
## Warning: The observed statistics (target.statistics) are near or at the boundary
## of its support, i.e. the Maximum Likelihood Estimates maynot exist or be hard
## to be estimated. In particular, the statistic(s) "edges", "kstar2", "I(edges *
## (class_name %in% "adult"))", "I(edges * (class_name %in% "senior"))".
```

```r
texreg::screenreg(list(model_2))
```

```
##
## ================================================
##                                       Model 1
## ------------------------------------------------
## edges                                   -1.54 ***
##                                         (0.30)
## kstar2                                   0.58 ***
##                                         (0.08)
## edges * (class_name %in% "adult")        0.01
##                                         (0.18)
## edges * (class_name %in% "senior")       0.05
##                                         (0.17)
## ------------------------------------------------
## AIC                                     348.98
## BIC                                     363.66
## Log Likelihood                         -170.49
## Num. networks                            29
## Convergence                               0
## Time (seconds)                            0.84
## ================================================
## *** p < 0.001; ** p < 0.01; * p < 0.05
```

```r
# Model 3
model_3 <- ergmito(
  nets_4_5 ~ edges +
    triangle +
    kstar(2)
  ,model_update = ~
    I(edges * (class_name %in% "adult")) +
    I(edges * (class_name %in% "senior"))
)
```

```
## Warning: ergmito does not fully support undirected graphs (yet). We will
## continue with the estimation process, but simulation has limited supported for
## now.
```

```
## Warning: The observed statistics (target.statistics) are near or at the boundary
## of its support, i.e. the Maximum Likelihood Estimates maynot exist or be hard
## to be estimated. In particular, the statistic(s) "edges", "triangle", "kstar2",
## "I(edges * (class_name %in% "adult"))", "I(edges * (class_name %in% "senior"))".
```

```
texreg::screenreg(list(model_3))
```

```
##
## ================================================
##                                       Model 1
## ------------------------------------------------
## edges                                  0.67
##                                       (0.58)
## triangle                               2.53 ***
##                                       (0.43)
## kstar2                                -0.86 **
##                                       (0.26)
## edges * (class_name %in% "adult")      0.01
##                                       (0.18)
## edges * (class_name %in% "senior")     0.05
##                                       (0.18)
## ------------------------------------------------
## AIC                                  319.98
## BIC                                  338.33
## Log Likelihood                      -154.99
## Num. networks                         29
## Convergence                            0
## Time (seconds)                         1.04
## ================================================
## *** p < 0.001; ** p < 0.01; * p < 0.05
```

```
# Model 4
model_4 <- ergmito(
  nets_4_5 ~ edges +
    triangle +
    kstar(2) +
    nodematch("female")
  ,model_update = ~
    I(edges * (class_name %in% "adult")) +
    I(edges * (class_name %in% "senior"))
)
```

```
## Warning: ergmito does not fully support undirected graphs (yet). We will
## continue with the estimation process, but simulation has limited supported for
## now.
```

```
## Warning: The observed statistics (target.statistics) are near or at the
## boundary of its support, i.e. the Maximum Likelihood Estimates maynot exist or
```

```
## be hard to be estimated. In particular, the statistic(s) "edges", "triangle",
## "kstar2", "nodematch.female", "I(edges * (class_name %in% "adult"))", "I(edges *
## (class_name %in% "senior"))".
```

```
texreg::screenreg(list(model_4))
```

```
##
## ===============================================
##                                      Model 1
## -----------------------------------------------
## edges                                 0.70
##                                      (0.59)
## triangle                              2.53 ***
##                                      (0.43)
## kstar2                               -0.86 **
##                                      (0.26)
## nodematch.female                     -0.06
##                                      (0.25)
## edges * (class_name %in% "adult")     0.01
##                                      (0.18)
## edges * (class_name %in% "senior")    0.04
##                                      (0.18)
## -----------------------------------------------
## AIC                                  321.92
## BIC                                  343.94
## Log Likelihood                      -154.96
## Num. networks                         29
## Convergence                            0
## Time (seconds)                         1.90
## ===============================================
## *** p < 0.001; ** p < 0.01; * p < 0.05
```

```
# Model 5
model_5 <- ergmito(
  nets_4_5 ~ edges +
    triangle +
    kstar(2) +
    absdiff("age")
  ,model_update = ~
    I(edges * (class_name %in% "adult")) +
    I(edges * (class_name %in% "senior"))
)
```

```
## Warning: ergmito does not fully support undirected graphs (yet). We will
## continue with the estimation process, but simulation has limited supported for
## now.
```

```
## Warning: The observed statistics (target.statistics) are near or at the boundary
## of its support, i.e. the Maximum Likelihood Estimates maynot exist or be hard
## to be estimated. In particular, the statistic(s) "edges", "triangle", "kstar2",
## "absdiff.age", "I(edges * (class_name %in% "adult"))", "I(edges * (class_name
## %in% "senior"))".
```

```
texreg::screenreg(list(model_5))
```

```
##
## ============================================
##                                   Model 1
## --------------------------------------------
## edges                              0.85
##                                   (0.60)
## triangle                           2.52 ***
##                                   (0.43)
## kstar2                            -0.86 **
##                                   (0.26)
## absdiff.age                       -0.01
##                                   (0.01)
## edges * (class_name %in% "adult")  -0.04
##                                   (0.19)
## edges * (class_name %in% "senior")  0.04
##                                   (0.18)
## --------------------------------------------
## AIC                              320.29
## BIC                              342.31
## Log Likelihood                  -154.14
## Num. networks                     29
## Convergence                        0
## Time (seconds)                     6.33
## ============================================
## *** p < 0.001; ** p < 0.01; * p < 0.05
```

```
# Model 6
model_6 <- ergmito(
  nets_4_5 ~ edges +
    triangle +
    kstar(2) +
    nodematch("university")
  ,model_update = ~
    I(edges * (class_name %in% "adult")) +
    I(edges * (class_name %in% "senior"))
)
```

```
## Warning: ergmito does not fully support undirected graphs (yet). We will
## continue with the estimation process, but simulation has limited supported for
## now.
```

```
## Warning: The observed statistics (target.statistics) are near or at the boundary
## of its support, i.e. the Maximum Likelihood Estimates maynot exist or be hard
## to be estimated. In particular, the statistic(s) "edges", "triangle", "kstar2",
## "nodematch.university", "I(edges * (class_name %in% "adult"))", "I(edges *
## (class_name %in% "senior"))".
```

```
texreg::screenreg(list(model_6))
```

```
##
```

```
## ================================================
##                                       Model 1
## ------------------------------------------------
## edges                                  0.55
##                                       (0.59)
## triangle                               2.53 ***
##                                       (0.43)
## kstar2                                -0.86 **
##                                       (0.26)
## nodematch.university                   0.21
##                                       (0.22)
## edges * (class_name %in% "adult")     -0.03
##                                       (0.19)
## edges * (class_name %in% "senior")     0.02
##                                       (0.18)
## ------------------------------------------------
## AIC                                   321.13
## BIC                                   343.15
## Log Likelihood                       -154.56
## Num. networks                          29
## Convergence                            0
## Time (seconds)                         2.50
## ================================================
## *** p < 0.001; ** p < 0.01; * p < 0.05
```

```
# Model 7
model_7 <- ergmito(
  nets_4_5 ~ edges +
    triangle +
    kstar(2) +
    absdiff("duration")
  ,model_update = ~
    I(edges * (class_name %in% "adult")) +
    I(edges * (class_name %in% "senior"))
)
```

```
## Warning: ergmito does not fully support undirected graphs (yet). We will
## continue with the estimation process, but simulation has limited supported for
## now.
```

```
## Warning: The observed statistics (target.statistics) are near or at the
## boundary of its support, i.e. the Maximum Likelihood Estimates maynot exist or
## be hard to be estimated. In particular, the statistic(s) "edges", "triangle",
## "kstar2", "absdiff.duration", "I(edges * (class_name %in% "adult"))", "I(edges *
## (class_name %in% "senior"))".
```

```
texreg::screenreg(list(model_7))
```

```
##
## ================================================
##                                       Model 1
## ------------------------------------------------
```

```
## edges                                    1.16 *
##                                         (0.59)
## triangle                                 2.46 ***
##                                         (0.43)
## kstar2                                  -0.84 **
##                                         (0.26)
## absdiff.duration                        -0.04 ***
##                                         (0.01)
## edges * (class_name %in% "adult")       -0.07
##                                         (0.20)
## edges * (class_name %in% "senior")       0.41
##                                         (0.22)
## ------------------------------------------------
## AIC                                    305.55
## BIC                                    327.57
## Log Likelihood                        -146.77
## Num. networks                           29
## Convergence                              0
## Time (seconds)                           7.79
## ================================================
## *** p < 0.001; ** p < 0.01; * p < 0.05
```

```
# Model 8
model_8 <- ergmito(
  nets_4_5 ~ edges +
    triangle +
    kstar(2) +
    nodematch("female") +
    absdiff ("age") +
    nodematch("university") +
    absdiff("duration")
  ,model_update = ~
    I(edges * (class_name %in% "adult")) +
    I(edges * (class_name %in% "senior"))
)
```

```
## Warning: ergmito does not fully support undirected graphs (yet). We will
## continue with the estimation process, but simulation has limited supported for
## now.
```

```
## Warning: The observed statistics (target.statistics) are near or at the boundary
## of its support, i.e. the Maximum Likelihood Estimates maynot exist or be hard
## to be estimated. In particular, the statistic(s) "edges", "triangle", "kstar2",
## "nodematch.female", "absdiff.age", "nodematch.university", "absdiff.duration",
## "I(edges * (class_name %in% "adult"))", "I(edges * (class_name %in% "senior"))".
```

```
texreg::screenreg(list(model_8))
```

```
##
## ================================================
##                                       Model 1
## ------------------------------------------------
```

```
## edges                                  1.14
##                                       (0.62)
## triangle                               2.46 ***
##                                       (0.43)
## kstar2                                -0.83 **
##                                       (0.26)
## nodematch.female                      -0.04
##                                       (0.27)
## absdiff.age                            0.00
##                                       (0.01)
## nodematch.university                  -0.04
##                                       (0.26)
## absdiff.duration                      -0.04 ***
##                                       (0.01)
## edges * (class_name %in% "adult")     -0.04
##                                       (0.21)
## edges * (class_name %in% "senior")     0.44
##                                       (0.24)
## ---------------------------------------------
## AIC                                  311.27
## BIC                                  344.30
## Log Likelihood                      -146.63
## Num. networks                         29
## Convergence                            0
## Time (seconds)                        10.28
## =============================================
## *** p < 0.001; ** p < 0.01; * p < 0.05
```

```
# Models
texreg::screenreg(list(model_1, model_2, model_3, model_4,
                       model_5, model_6, model_7, model_8))
```

```
##
## ============================================================================================================
##                                    Model 1      Model 2      Model 3      Model 4      Model 5
## ------------------------------------------------------------------------------------------------------------
## edges                               -0.86 ***    -1.54 ***     0.67         0.70         0.85
##                                     (0.20)       (0.30)       (0.58)       (0.59)       (0.60)
## triangle                             1.17 ***                  2.53 ***     2.53 ***     2.52 ***
##                                     (0.14)                    (0.43)       (0.43)       (0.43)
## edges * (class_name %in% "adult")    0.01         0.01         0.01         0.01        -0.04
##                                     (0.16)       (0.18)       (0.18)       (0.18)       (0.19)
## edges * (class_name %in% "senior")   0.04         0.05         0.05         0.04         0.04
##                                     (0.16)       (0.17)       (0.18)       (0.18)       (0.18)
## kstar2                                            0.58 ***    -0.86 **     -0.86 **     -0.86 **
##                                                   (0.08)       (0.26)       (0.26)       (0.26)
## nodematch.female                                                            -0.06
##                                                                             (0.25)
## absdiff.age                                                                              -0.01
##                                                                                          (0.01)
## nodematch.university
##
## absdiff.duration
##
```

```
## ------------------------------------------------------------------------------------------------
## AIC                                328.33       348.98       319.98       321.92       320.29
## BIC                                343.01       363.66       338.33       343.94       342.31
## Log Likelihood                   -160.16      -170.49      -154.99      -154.96      -154.14
## Num. networks                      29           29           29           29           29
## Convergence                         0            0            0            0            0
## Time (seconds)                      0.97         0.84         1.04         1.90         6.33
## ================================================================================================
## *** p < 0.001; ** p < 0.01; * p < 0.05
```

```
# GOF
# Model 3
# gof_model_3 <- gof_ergmito(model_1)
# plot(gof_model_3)
# plot(model_3)
# Model 7
# gof_model_7 <- gof_ergmito(model_7)
# plot(gof_model_7)
# plot(model_7)
```