

# 時系列データやパネルデータからのネットワーク推定

## 10 章の練習問題

ジュリアン・バーガー<sup>1,2,3</sup>, リア・H・A・フークストラ<sup>2</sup>,  
アレッサンドラ・C・マンズト<sup>2,3,4</sup>, & サシャ・エプスカンプ<sup>2,3</sup>

1. フローニンゲン大学医療センター
2. アムステルダム大学心理学科
3. アムステルダム大学都市精神健康センター
4. アムステルダム大学コミュニケーション科学科

2020 年 3 月、新種のコロナウィルスが急速に拡散したことを受けて、多くの国々がロックダウンの状況に陥った。ロックダウンの措置はウィルスの拡散速度を抑えるのに役立ち、そのおかげでヘルスケアのキャパシティが保たれた。一方で、社会的な孤立が個人の人々のメンタルヘルスにネガティブな効果を及ぼすのではないかと、ということに危惧する人も多かった。本章の実践問題では、Fried, Papanikolaou, and Epskamp (2020) が収集したデータを分析してもらおう。このデータセットには、80 人のオランダ人学生から経験サンプリング法 (experiential sampling method: ESM) で収集した時系列データが含まれている。その ESM では、メンタルヘルスの問題に関する項目群が毎日 4 回、2 週間にわたって測定された (つまり、合計で 56 回の測定が行われた)。測定は 3 月末まで実施されたが、その時期はちょうどオランダでロックダウン期が始まった頃であった。データや分析、結果についての詳細情報については、著者らが記した論文やブログ記事を参照してほしい。ここからは、こちらのデータを使用し、個人ごとの時系列ネットワークの演習 (graphicalVAR パッケージと psychonetrics パッケージを活用) と複数名の参加者についての時系列ネットワークの演習 (mlVAR パッケージを活用) の両方を実施していこう\*<sup>1</sup>。この研究で測定されていた項目群は、以下の通りである。

項目番号	略称	項目内容
1	id	参加者 ID
2	Relax	リラックスするのが困難であった
3	Irritable	(とても) イライラした
4	Worry	様々なものごとについて心配していた
5	Nervous	緊張, 不安, 苛立ちを感じた
6	Future	先々に楽しみなことが何もないように感じた
7	Anhedonia	いかなるポジティブ感情も体験できそうになかった
8	Tired	疲れを感じた
9	Alone	仲間がいなかったかのように感じたり, 他人と親密ではないように感じたりした
10	Social_offline	オフラインの有意義な社交に費やした時間は, ... であった。
11	Social_online	ソーシャルメディアを使って時間をつぶしていた/やり過ごしていた時間は, ... であった。
12	Outdoors	外出した時間は, ... であった。
13	C19_occupied	コロナウィルスのことで占められていた (例: ニュースを見る, コロナについて考える, 友達とコロナについて話す) 時間は, ... であった。
14	C19_worry	コロナウィルスの関連で, 自身自身の健康や, 親しい友人や家族の健康について考えた時間は, ... であった。
15	Home	家 (親やパートナーの家を含む) にいた時間は, ... であった。
16	day	その回答が行われたのは, 研究の何日目であったかを示す変数
17	beep	その回答が行われたのは, その日の何回目のビーブ (測定機会) においてであったかを示す変数
18	conc	その回答に紐づけられた, ビーブの識別番号

注: 項目 1 から 15 は, すべて 1-5 の 5 段階で測定された。項目 1 から 9 についての回答の選択肢は, 「まったくなかった」「少しあった」「まあまああった」「かなりあった」「極端にあった」の 5 段階であった。項目 10 から 15 についての回答の選択肢は, 「0 分」「1-15 分」「15-60 分」「1-2 時間」「2 時間以上」であった。

訳注: 項目 10 から 15 については, 「項目内容」の欄に記された文面の...の部分に所要時間を代入する形で回答がなされた (例: 「オフラインの有意義な社交に費やした時間は, 『15-60 分』であった」)。

\*<sup>1</sup> なお, もとの研究においては, 本書の学習範囲を超えた発展的なデータ処理がいくつか施されているということを留意しておいてほしい。この実践問題で得られる結果は, 簡略化したデータ処理に基づくものであるため, Freid et al. (2020) の知見をそのまま再現したものとはならない。

まず、以下のように `library` 関数を用いて、必要なパッケージを読み込もう (まだインストールしていないものがあれば、インストールしておくこと)。

```
library("mlVAR")
library("graphicalVAR")
library("psychometrics")
library("dplyr")
library("qgraph")
```

以下のように、“clean\_network.RData” というファイルを [OSF](#) からダウンロードし、R に読み込もう。

```
load("clean_network.RData")
```

次に、以下のコードを実行してノードのラベルを改良するとともに、論文では活用されなかった項目を除外しよう。

```
# 検討対象となる変数を指定：
vars <- paste0("Q", 1:18)

# ラベルの指定：
varLabs <- c("Relax", "Irritable", "Worry", "Nervous", "Future", "Anhedonia",
             "Tired", "Hungry", "Alone", "Angry", "Social_offline", "Social_online",
             "Music", "Procrastinate", "Outdoors", "C19_occupied",
             "C19_worry", "Home")

# データ内の列の名称を変更：
names(Data2)[names(Data2) %in% vars] <- varLabs

# 不要な項目の除外：
Data2 <- Data2 %>% select(-Hungry, -Angry, -Music, -Procrastinate)
varLabs <- varLabs[!varLabs %in% c("Hungry", "Angry", "Music", "Procrastinate")]
```

これで、あなたの R 環境には 2 つのオブジェクトができたことになる。データを格納した `Data2` というオブジェクトと、分析対象となる変数名を格納した `varLabs` というオブジェクトである。

### N = 1 の時系列分析

N = 1 の時系列分析を練習するために、ある単一の参加者のデータだけを読み込もう。以下のコードでは ID68 の参加者を選択しているが、別の参加者を選んでもよい (分析中にたくさんのエラーが出た場合には、選択した参加者のなかで分散が小さすぎることに起因している可能性がある。)

```
my_data <- Data2[which(Data2$id == 68),]
```

続いて、あなたの興味に応じて 3 個か 4 個の変数を選択しよう (参加者一人ひとりについて観測値の数

が非常に少ないので、5 個以上のノードは分析に含めない方が好ましいだろう)。たとえば、以下のコードを実行してみよう。

```
includevars <- c("C19_worry", "Future")
```

上記のコードにおいて、C19\_worry と Future 以外の変数を選んでも構わない。次に、以下のように *graphicalVAR* パッケージを用いて、(正則化を施した場合の) グラフィカル VAR モデルを推定しよう。

```
result_graphicalVAR <- graphicalVAR(
  my_data,
  vars = includevars,
  dayvar = "day",
  beepvar = "beep",
  gamma = 0.25)

# 経時的ネットワーク：
temp_graphicalVAR <- result_graphicalVAR$PDC

# 同時性ネットワーク：
cont_graphicalVAR <- result_graphicalVAR$PCC
```

あるいは以下のように、*psychonetrics* パッケージを用いた推定も可能である (この場合は、枝刈りを用いている)。

```
mod_psychonetrics <- gvar(
  my_data,
  vars = includevars,
  dayvar = "day",
  beepvar = "beep",
  estimator = "FIML" # <- handles missing data well
) %>%
  runmodel %>%
  prune(
    alpha = 0.05
  )

# 経時的ネットワーク：
temp_psychonetrics <- getmatrix(mod_psychonetrics, "PDC")

# 同時性ネットワーク：
cont_psychonetrics <- getmatrix(mod_psychonetrics, "omega_zeta")
```

### Exercise 1

*graphicalVAR* パッケージと *psychonetrics* パッケージの両方を用い、あなたが選択した参加者 1 名について、任意の変数 3 個か 4 個を含んだ GVAR モデルを推定しよう。2 つの推定方法それぞれについて、経時的ネットワークと同時性ネットワークをプロットし (なお、円形レイアウトもしくは平均レイアウトを指定し、すべてのネットワークを同一のレイアウトで表示すること)、結果を比較しよう。推定された経時的ネットワークと同時性ネットワークについて、短い文章 (英文換算で 100 語以内) で解釈を述べてほしい。

### Exercise 2

上記の関数のなかで、引数 `day` と `beep` はどのような働きをしているだろうか。また、「1 年間にわたる研究を実施し、平日 (週末は除く) に 1 日 1 個の観測値を含んだデータセットが得られた」という場面を想像してほしい。この場面において、金曜日の測定値から月曜日の測定値への回帰が生じないようにするためには、どのような引数を用いるとよいだろうか。

その他にも、時点を表す変数を用いれば (今回の場合は、測定時間の代用として変数 `conc` を用いる)、データ内のトレンドをチェックすることができる。たとえば、変数 “C19\_worry” がトレンドを示しているかどうかを検定するには、以下のように時点を独立変数にした線形回帰を実行することができる。

```
lm_c19_worry <- lm(C19_worry ~ conc, data = my_data)
summary(lm_c19_worry)

##
## Call:
## lm(formula = C19_worry ~ conc, data = my_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.03552 -0.30191 -0.01976  0.16423  1.32564
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.103235   0.152314   13.81  < 2e-16 ***
## conc        -0.022572   0.004542   -4.97 8.58e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5011 on 49 degrees of freedom
## (5 observations deleted due to missingness)
## Multiple R-squared:  0.3351, Adjusted R-squared:  0.3216
## F-statistic: 24.7 on 1 and 49 DF, p-value: 8.576e-06
```

上記のように、トレンドの効果は有意であった。この場合は、以下のようにすれば変数“C19\_worry”を「デトレンド (detrrend)」できる。

```
my_data$C19_worry[!is.na(my_data$C19_worry)] <- residuals(lm_c19_worry)
```

### Exercise 3

あなたがここまで選択した変数についてトレンドの有意性を検定し ( $\alpha = 0.05$ )、トレンドが有意であった場合にはデトレンドを実施しよう。そのうえで、*graphicalVAR* パッケージか *psychometrics* パッケージを用いてネットワークを再推定しよう。ネットワークの推定結果に、何か変化は生じただろうか。

## $N > 1$ の時系列分析

次に、データセット全体 (Data2) を *mlVAR* パッケージとともに用いて、マルチレベル・グラフィカル VAR を推定しよう。

### Exercise 4

*mlVAR* 関数のヘルプファイルを参照したうえで、4 個から 6 個の変数を選択し、データセット全体についてのマルチレベル GVAR モデルを推定しよう。

### Exercise 5

ヘルプファイル (?plot.mlVAR) を参照して *mlVAR* パッケージにおけるプロットの方法を確認し、推定された (固定効果の) 経時的 (temporal) ネットワーク、同時性 (contemporaneous) ネットワーク、個人間 (between-subjects) ネットワークをプロットしよう。なお、すべてのネットワークを同一のレイアウト (円形レイアウト、もしくは平均レイアウト) でプロットし、有意水準  $\alpha = 0.05$  に満たないエッジを隠す (閾値処理を施す) ようにしよう。また、同時性ネットワークと個人間ネットワークについては AND ルールを適用し、エッジを表示する際のタイプ 1 エラー率を最小化すること。推定されたネットワークについて、あなたの解釈を短い文章 (英文換算で 100 語以内) で要約しよう。