

Understanding Diffusion with **netdiffuseR**

Survival Analysis

Sunbelt 2016 INSNA

George Vega Yon Thomas Valente

Department of Preventive Medicine
University of Southern California

Newport Beach, CA
April 5, 2016

Setup

We will use the medical innovations data

```
# Loading the required packages
```

```
library(survival)
```

```
library(netdiffuseR)
```

```
# Loading the data
```

```
data("medInnovationsDiffNet")
```

```
medInnovationsDiffNet
```

```
## Dynamic network of class -diffnet-
```

```
## # of nodes : 125 (1001, 1002, 1003, 1004, 1005, 1006, 1007, 1008, ...)
```

```
## # of time periods : 18 (1 - 18)
```

```
## Type : directed
```

```
## Final prevalence : 1.00
```

```
## Static attributes : city, detail, meet, coll, attend, proage, length, ... (58)
```

```
## Dynamic attributes : -
```

Preparing the data

From **netdiffuseR** we will get the following covariates:

- ▶ Cohesive exposure: Proportion of ego's adopters at each time period
- ▶ Structural equivalence exposure: Same as before but using the structural equivalence graph instead of the baseline network.

```
# Structural equivalence exposure  
medInnovationsDiffNet[["seexp"]] <- exposure(  
  medInnovationsDiffNet, groupvar="city", alt.graph = "se")
```

```
## Warning in exposure(medInnovationsDiffNet, groupvar = "city", alt.graph =  
## "se"): To use alt.graph="se" -valued- has been switched to TRUE.
```

```
# Cohesive exposure  
medInnovationsDiffNet[["cohexp"]] <- exposure(medInnovationsDiffNet)
```

Coercing data into a dataframe

```
# Getting the data for running the cox regression
dat <- diffnet.attrs(medInnovationsDiffNet, as.df = TRUE)
dat <- subset(
  dat,
  subset =
    per <= toa &
    per < 18 &
    !is.na(date),
  select = c(id, per, toa, date, city, proage, proage2, seexp, cohexp))

# Creating the event variable
dat$event <- with(dat, toa==per)

# Here, since its survival, we only care from when the doctor is aware.
dat <- subset(dat, per - date >=0)

# Checking out the data
dat <- dat[with(dat, order(id, per)),]
head(dat,10)
```

##	id	per	toa	date	city	proage	proage2	seexp	cohexp	event
## 377	1002	4	12	4	1	6	0	0.4111432	1.0000000	FALSE
## 502	1002	5	12	4	1	6	0	0.5414957	1.0000000	FALSE
## 627	1002	6	12	4	1	6	0	0.6102774	1.0000000	FALSE
## 752	1002	7	12	4	1	6	0	0.6955077	1.0000000	FALSE
## 877	1002	8	12	4	1	6	0	0.7508093	1.0000000	FALSE
## 1002	1002	9	12	4	1	6	0	0.7717741	1.0000000	FALSE
## 1127	1002	10	12	4	1	6	0	0.7809304	1.0000000	FALSE
## 1252	1002	11	12	4	1	6	0	0.8284503	1.0000000	FALSE
## 1377	1002	12	12	4	1	6	0	0.8518452	1.0000000	TRUE
## 879	1004	8	9	8	1	5	0	0.7550771	0.6666667	FALSE

Notice that `diffnet.attrs` generates two extra variables: `per` (time period) and `id`.

The survival package

- ▶ In order to work, the `survival` package works with `Surv` objects.
- ▶ These store the response/events and the time frame during which these occurred.
- ▶ Usually take the following form: `Surv(start, end, event)`.
- ▶ For this tutorial we will use the Cox model, from Andersen and Gill (1982)

$$\mathcal{L}(\beta) = \prod_{i=1}^n \left\{ \frac{\exp \beta' x_i(T_i)}{\sum_j \exp \beta' x_j(T_j)} \right\}^{\delta_i}$$

Which can extended to time-variant covariates.

The survival package

The Survreg object

- ▶ First, we need to create the Survreg object using the function with the same name

```
# Needs a start, stop, event  
surv_mi <- with(dat, Surv(per-1, per, event))
```

Notice the warning as the time frames should be greater than 1.

- ▶ Now, let's take a look at the object itself

```
head(surv_mi, 10)  
  
## [1] ( 3, 4+] ( 4, 5+] ( 5, 6+] ( 6, 7+] ( 7, 8+] ( 8, 9+] ( 9,10+]  
## [8] (10,11+] (11,12] ( 7, 8+]
```

Fitting the model

All cities

```
# Fitting a model  
set.seed(1988)  
mymodel <- formula(surv_mi ~ factor(city) + proage + I(proage^2) +  
                    seexp + cohexp + cluster(id))  
out <- coxph(mymodel, data=dat)
```

Fitting the model

All cities (cont. 1)

Table 1: Fitting Proportional Hazards Regression Model: mymodel

	coef	exp(coef)	robust se	z	p
factor(city)2	-0.6598	0.517	0.4593	-1.437	0.2
factor(city)3	-0.3608	0.6971	0.5688	-0.6343	0.5
factor(city)4	-1.046	0.3514	0.8926	-1.172	0.2
proage	1.184	3.268	0.4396	2.694	0.007
l(proage^2)	-0.1541	0.8572	0.05499	-2.803	0.005
seexp	-0.3141	0.7304	1.075	-0.2923	0.8
cohexp	0.4132	1.512	0.6524	0.6334	0.5

Likelihood ratio test=8.48 on 7 df, p=0.2922049 n= 310, number of events= 37

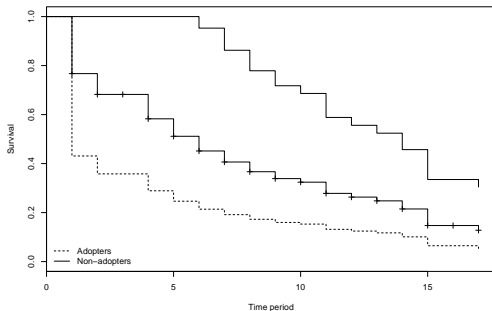
Fitting the model

All cities (cont. 2)

More diagnostics can be done as follows:

```
# Diagnostics
fit <- survfit(out)

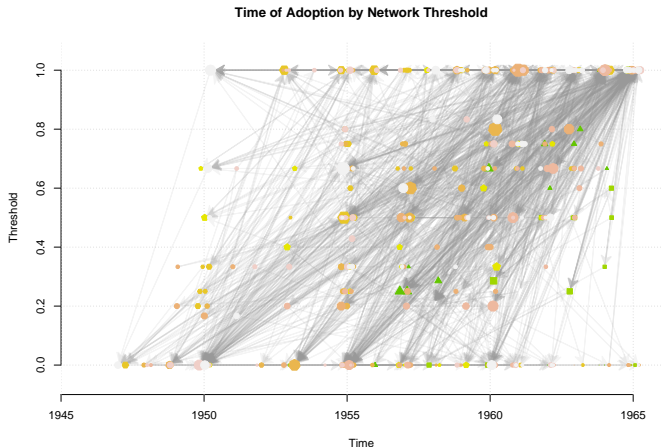
plot(fit, mark.time=TRUE, lty=1:2,
     xlab="Time period", ylab="Survival")
legend("bottomleft", c("Adopters", "Non-adopters"),
     lty=2:1, bty="n")
```



Example with brazilian farmers

```
data("brfarmersDiffNet")

# Creating a classification for village
village <- as.integer(factor(brfarmersDiffNet[["village"]])) + 2
nvillage <- length(unique(village))
plot_threshold(brfarmersDiffNet,
              vertex.sides = village, # Creates polygons
              vertex.col   = terrain.colors(nvillage)[village]) # Colors!
```



Example with brazilian farmers

Preparing the data

```
# Exposure variables
brfarmersDiffNet[["seexp"]] <- exposure(brfarmersDiffNet, alt.graph = "se",
                                       groupvar="village", valued = TRUE)
brfarmersDiffNet[["cohexp"]] <- exposure(brfarmersDiffNet)

# Creating a dynamic version of age
age <- brfarmersDiffNet[["age"]]
pers <- brfarmersDiffNet$meta$pers
brfarmersDiffNet[["age_dyn"]] <- lapply(
  seq_len(nslices(brfarmersDiffNet)), function(x) {
    age + (pers[x] - 1966) # Surveyed in 1966
  })

# Subset
dat <- diffnet.attrs(brfarmersDiffNet, as.df = TRUE)
dat <- subset(dat, per <= toa, select=c(per, toa, age_dyn, village, seexp, cohexp, id))

# Creating the event variable
dat$event <- with(dat, toa==per)

# Checking out the data
dat <- dat[with(dat, order(id, per)),]
head(dat,10)
```

##	per	toa	age_dyn	village	seexp	cohexp	id	event
## 1	1946	1961	21	10	0.009157484	0	1001	FALSE
## 693	1947	1961	22	10	0.009157484	0	1001	FALSE
## 1385	1948	1961	23	10	0.009157484	0	1001	FALSE
## 2077	1949	1961	24	10	0.009157484	0	1001	FALSE
## 2769	1950	1961	25	10	0.009157484	0	1001	FALSE
## 3461	1951	1961	26	10	0.009157484	0	1001	FALSE
## 4153	1952	1961	27	10	0.009157484	0	1001	FALSE
## 4845	1953	1961	28	10	0.009157484	0	1001	FALSE
## 5537	1954	1961	29	10	0.009157484	0	1001	FALSE
## 6229	1955	1961	30	10	0.028998057	0	1001	FALSE

Example with brazilian farmers

Fitting the data

```
out <- coxph(Surv(per-1, per, event) ~ factor(village) + seexp + cohexp + age_dyn + cluster(id),  
             data=dat)
```

Table 2: Fitting Proportional Hazards Regression Model: $\text{Surv}(\text{per} - 1, \text{per}, \text{event}) \sim \text{factor}(\text{village}) + \text{seexp} + \text{cohexp} + \text{age_dyn} + \text{cluster}(\text{id})$

	coef	exp(coef)	robust se	z	p
factor(village)22	-0.1878	0.8287	0.1366	-1.375	0.2
factor(village)23	0.7926	2.209	0.1855	4.274	2e-05
factor(village)24	0.4047	1.499	0.1913	2.115	0.03
factor(village)30	0.5677	1.764	0.1823	3.115	0.002
factor(village)31	0.2811	1.325	0.1638	1.716	0.09
factor(village)43	0.2333	1.263	0.1595	1.462	0.1
factor(village)70	0.8174	2.265	0.2188	3.736	2e-04
factor(village)71	0.2508	1.285	0.1669	1.503	0.1
factor(village)80	0.2486	1.282	0.1777	1.399	0.2
factor(village)82	0.4322	1.541	0.1703	2.538	0.01
seexp	0.71	2.034	0.2786	2.548	0.01
cohexp	0.441	1.554	0.1083	4.073	5e-05
age_dyn	0.001521	1.002	0.002515	0.6049	0.5

Likelihood ratio test=140.58 on 13 df, p=0 n= 10244, number of events= 692

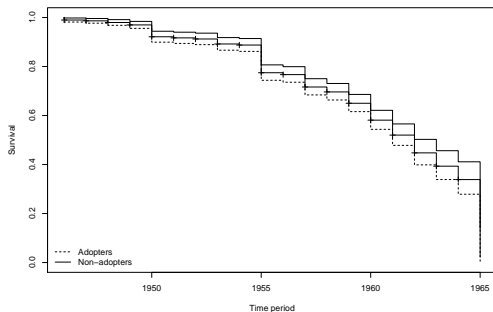
Example with brazilian farmers

Diagnostics

More diagnostics can be done as follows:

```
# Diagnostics
fit <- survfit(out)

plot(fit, mark.time=TRUE, lty=1:2,
     xlab="Time period", ylab="Survival", firstx = min(dat$per))
legend("bottomleft", c("Adopters", "Non-adopters"),
     lty=2:1, bty="n")
```



Example with brazilian farmers

Diagnoststics (cont.)

```
cox.zph(out)
```

```
##              rho      chisq      p
## factor(village)22 -0.00049 8.24e-05 9.93e-01
## factor(village)23 -0.05327 1.38e+00 2.40e-01
## factor(village)24 -0.12147 7.99e+00 4.71e-03
## factor(village)30 -0.08969 4.20e+00 4.03e-02
## factor(village)31 -0.06936 2.26e+00 1.33e-01
## factor(village)43  0.01657 1.05e-01 7.46e-01
## factor(village)70 -0.15421 1.97e+01 8.83e-06
## factor(village)71 -0.05388 1.34e+00 2.47e-01
## factor(village)80 -0.01568 1.04e-01 7.47e-01
## factor(village)82 -0.01233 7.04e-02 7.91e-01
## seexp              0.12935 1.14e+01 7.43e-04
## cohexp             -0.11052 7.71e+00 5.50e-03
## age_dyn            -0.11508 7.16e+00 7.45e-03
## GLOBAL              NA 5.11e+01 1.89e-06
```