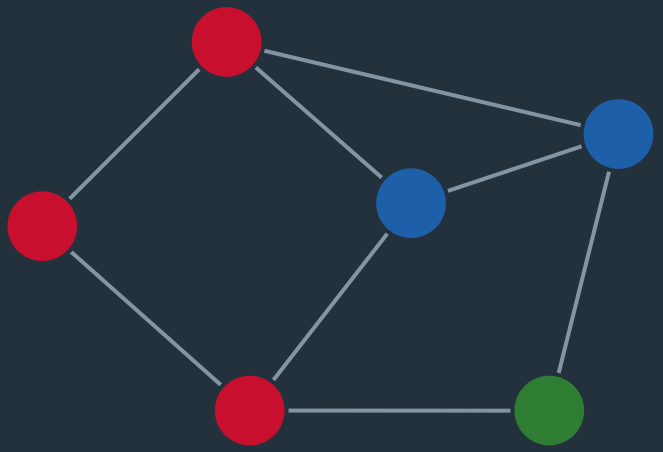
Network analysis in Stata

Session 3 · Relational events with nwrem and nwdynam

Three hours, an optional third session after the six-hour workshop, with the data and commands of Network Analysis with Stata (Stata Press)

Instructors are free to use, adapt and share these slides.



Today: six blocks, one break

0:00	**1 Events and event data**	15 min
0:15	**2 The relational event model: nwrem**	45 min
1:00	**3 The DyNAM: who acts, and whom they choose**	40 min
1:40	Break	15 min
1:55	**4 Comparing the two models**	20 min
2:15	**5 Postestimation: fit, outliers, change points**	25 min
2:40	**6 Limits, pitfalls and where to go**	20 min

Before we start

```
. nwrem version
. nwdynam version
```

Both are installed by nwinstall, everything: nwrem 1.5.0 and nwdynam 1.9.0.

```
. nwwebuse manufacturing, nwclear
. keep if t < 14
```

The e-mails of the first two weeks of the data.

E-mail log

2,738 e-mails among 129 employees, with the time in days.

Toy log

Ten events among three actors, typed in, for the arithmetic.

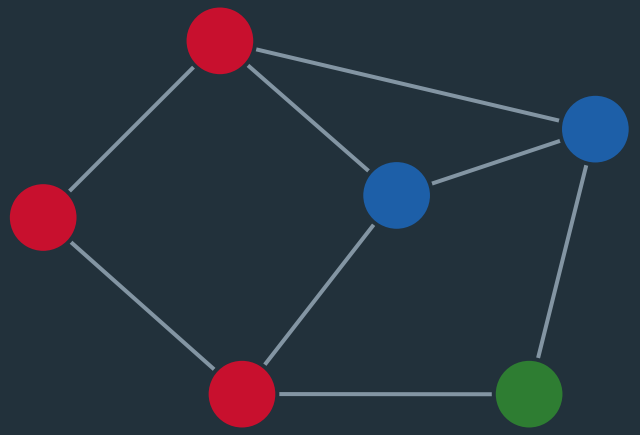
Employees

Supervisor, level and division, to ask who writes to whom.

1

Events and event data

15 min · Chapter 27



A relational event is a dated act

Event data

Each row is one act: a sender, a receiver and a time. The order of events matters, and so does who acted before. There are no waves.

Panel data

Networks observed at a few moments. What happened between the observations is unknown and has to be modeled as a process (Session 2).

The question changes: not which ties exist, but which event happens next, given everything that has happened so far.

The e-mail log as an event network

```
. nwwebuse manufacturing, nwclear
. keep if t < 14
. nwset sender receiver, eventtime(t) name(mail)
. nwsummarize mail
   Network name:  mail
   Directed: true
   Nodes: 129
   Arcs: 1162
   Density:  .0704
   Temporal: true
      Temporal type: event
```

2,738

e-mails among 129 employees in the first 14 days

`eventtime()` declares the log as an event network: one row per event, kept in time order.

A ten-event toy log

```
. input sender receiver t
. nwset sender receiver, eventtime(t) name(chat)
. nwrem chat, nodsnd nidrec

Relational event model (ordinal partial likelihood, MLE)
Network: chat                  Actors: 3
Events: 10              Log likelihood:  -15.9033

       chat | Coefficient  Std. err.      z    P>|z|
------------+----------------------------------------
     nodsnd |   1.268585   1.354195     0.94   0.349
     nidrec |  -5.567766   3.511139    -1.59   0.113
```

−15.90

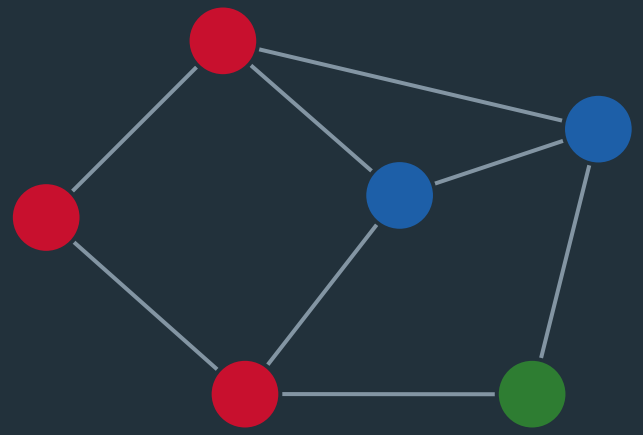
log likelihood of ten events among three actors

Three actors and ten events carry little information: the standard errors are large. The toy log is for arithmetic, not for conclusions.

2

The relational event model: nwrem

45 min · Chapter 27



What the model compares at every event

16,512

ordered pairs at risk at every event: 129 × 128

The pair that fired against all pairs at risk

At each event the model asks how likely the observed pair was among all ordered pairs of actors.

No intercept

Every pair is compared at the same instant, so a common level shifts all of them equally and cannot be estimated.

Written for **`relevent::rem.dyad()`** in R (Butts 2008). nwrem follows relevent and was cross-checked with it.

Activity: who writes, and who is written to

```
. nwrem mail, nodsnd nidrec

Relational event model (ordinal partial likelihood, MLE)
Network: mail                    Actors: 129
Events: 2738              Log likelihood: -2.53e+04

        mail | Coefficient  Std. err.      z    P>|z|
-------------+----------------------------------------
      nodsnd |   36.99531   1.478789    25.02   0.000
      nidrec |   35.10519   .6882604    51.01   0.000
```

nodsnd

the share of all events so far that this sender sent

nidrec

the share of all events so far that this receiver received

Both are positive and large: busy senders keep writing, popular receivers keep receiving. Neither coefficient is stable: both shift when other effects enter.

Inertia: do people write to those they wrote to?

```
. nwrem mail, nodsnd nidrec frpsndsnd

Relational event model (ordinal partial likelihood, MLE)
Network: mail                  Actors: 129
Events: 2738            Log likelihood: -2.44e+04

        mail | Coefficient  Std. err.      z    P>|z|
-------------+----------------------------------------
      nodsnd |   39.51321    1.54486    25.58   0.000
      nidrec |   26.72635   .7611982    35.11   0.000
   frpsndsnd |   4.406264   .0747082    58.98   0.000
```

× 1.55

in the odds of a pair per 10 percentage points more of the sender's past mail to that receiver

frpsndsnd is the fraction of the sender's earlier events that went to this receiver: exp(4.406 × 0.1) = 1.55.

Frequency or recency? Two different claims

```
. nwrem mail, nodsnd nidrec rsndsnd

        mail | Coefficient  Std. err.      z    P>|z|
-------------+----------------------------------------
      nodsnd |    30.2814   1.465531    20.66   0.000
      nidrec |   20.73686   .7949586    26.09   0.000
     rsndsnd |   3.592788   .0498983    72.00   0.000

. estimates stats freq rec
       Model |          N   ll(null)  ll(model)      df        AIC
        freq |      2,738  -26591.03  -24364.72       3   48735.45
         rec |      2,738  -26591.03  -23599.47       3   47204.93
```

1,531

AIC points in favor of recency, with the same number of parameters (3 and 3)

`rsndsnd:` how recently the sender wrote to this receiver, as a rank: the last receiver ranks first. Recent contact predicts the next event better than frequent contact.

Same supervisor, once history is controlled

```
. nwload mail, xvars
. rename _nwnode label
. merge 1:1 label using "$nwwebpath/manufacturing_nodes.dta", ///
>     keep(master match) nogen
. nwexpand supervisor, name(samesup)
. nwrem mail, nodsnd nidrec frpsndsnd covevent(samesup)

        nodsnd |   37.68373   1.525432    24.70   0.000
        nidrec |   27.30805   .7524725    36.29   0.000
     frpsndsnd |    4.10912   .0760244    54.05   0.000
covevent_samesup |   1.135317   .0545936    20.80   0.000
```

× 3.1

in the odds of a pair that reports to the same supervisor: exp(1.135)

samesup is a network over the same employees: 1 for two people with the same supervisor, built from the employee file with nwexpand. covevent() puts it into the model.

Names in full, intervals on request

```
. nwrem, ci

         mail | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
--------------+----------------------------------------------------------------
       nodsnd |   37.68373   1.525432    24.70   0.000     34.69394    40.67353
       nidrec |   27.30805   .7524725    36.29   0.000     25.83323    28.78287
    frpsndsnd |    4.10912   .0760244    54.05   0.000     3.960115    4.258125
covevent_sa~p |   1.135317   .0545936    20.80   0.000     1.028316    1.242319
```

The same holds for nwdynam, nwergm and nwsaom. noci leaves the intervals out.

By default

term names are printed in full and the interval columns are left out when linesize is below 100.

ci

brings the intervals back; the long names are then cut.

Hands-on (15 min)

Ch. 27, Ex. 2

The e-mail log has 129 actors and 2,738 events. (a) How many ordered pairs are in the risk set at each event? (b) With all coefficients zero, what is the log likelihood of the 2,738 events? (c) Compare it with the baseline model with nodsnd and nidrec (-2.53e4).

Hint Each of the 129 employees can write to any of the 128 others. With equal weights every event has probability one over the size of the risk set, so the log likelihood is 2,738 times its logarithm; display computes it.

Ch. 27, Ex. 4

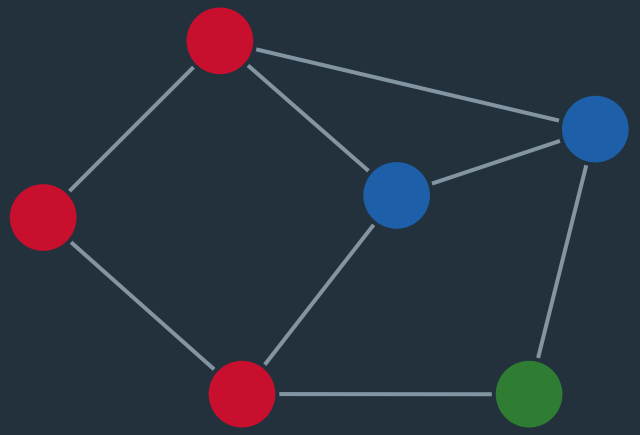
After keep if t < 14 and nwset sender receiver, eventtime(t) name(mail), fit nwrem mail, nodsnd nidrec rsndsnd rrecsnd. Report rsndsnd, rrecsnd and the log likelihood next to the fit with rsndsnd alone (3.59; -2.36e4). Does writing again to one's latest partner still matter once answering the latest writer is in the model, and by how much does its coefficient shrink?

Hint display e(ll) after each fit prints the log likelihood with all its digits. Divide the rsndsnd coefficient of the joint fit by its value in the fit with rsndsnd alone.

3

The DyNAM: who acts, and whom they choose

40 min · Chapter 28



Two questions, two sub-models

Rate: who acts next?

Which actor sends the next event, and, with an intercept, how long the waiting time is. Statistics describe the sending actor: indeg and outdeg of the sender.

Choice: to whom?

Given the sender, which receiver is chosen among the candidates. Statistics describe the candidate: inertia, recip, indeg of the candidate.

The same effect name means different things in the two sub-models: indeg of the sender in the rate model, indeg of the candidate in the choice model. nwdynam follows goldfish (Stadtfeld and Block 2017).

The choice sub-model

```
. nwset sender receiver, eventtime(t) directed name(mail)
. nwdynam mail

Dynamic Network Actor Model - choice sub-model (MLE)
Network: mail                  Actors: 129
Events: 2738            Log likelihood: -9.96e+03

        mail | Coefficient  Std. err.      z    P>|z|
-------------+----------------------------------------
     inertia |   2.036706   .0636816    31.98   0.000
       recip |   1.883335   .0609398    30.90   0.000
       indeg |    .019011   .0011356    16.74   0.000
```

× 7.7

in the odds of choosing a receiver the sender wrote to before: exp(2.037)

recip works the same way for receivers who wrote to the sender before: exp(1.883) = 6.6.

The rate sub-model: who sends next

```
. nwdynam mail, submodel(rate)

Dynamic Network Actor Model - rate sub-model, no intercept (MLE)
Events: 2738            Log likelihood: -1.27e+04

        mail | Coefficient  Std. err.      z    P>|z|
-------------+----------------------------------------
       indeg |  -.0020857    .002177    -0.96   0.338
      outdeg |   .1307426   .0043335    30.17   0.000
```

It says

Actors who have sent many e-mails are more likely to send the next one; the number of e-mails received does not matter (p = 0.34).

It does not say

How long the actor waits. Without an intercept the rate model uses only the order of events.

Same branch, net of structure

```
. encode division, generate(branch)
. nwdynam mail, inertia recip indeg same(branch)

Dynamic Network Actor Model - choice sub-model (MLE)
Events: 2738              Log likelihood: -9.91e+03

        mail | Coefficient  Std. err.      z    P>|z|
-------------+----------------------------------------
     inertia |   1.987294    .063794    31.15   0.000
       recip |   1.842186   .0609646    30.22   0.000
       indeg |   .0194229   .0011295    17.20   0.000
        same |   .4944068   .0467183    10.58   0.000
```

× 1.6

in the odds of a receiver from the sender's own branch: exp(0.494)

same(branch) is 1 when sender and candidate belong to the same branch of the company (12 branches, 72 of the 129 employees in the largest). The effect stays after inertia, reciprocity and popularity.

When real time matters: the rate with an intercept

```
. nwdynam mail, submodel(rate) intercept indeg outdeg

Dynamic Network Actor Model - rate sub-model, with intercept (MLE)
Events: 2738            Log likelihood: -435.5168

        mail | Coefficient  Std. err.      z    P>|z|
-------------+----------------------------------------
   Intercept |   .1240626   .0317477     3.91   0.000
       indeg |  -.0018054   .0022274    -0.81   0.418
      outdeg |   .0889101    .003791    23.45   0.000
```

−435.5

log likelihood of the continuous-time model: it is not comparable with the ordinal models

With the intercept every actor has a Poisson clock with rate exp(intercept + effects), in events per day. The model uses the waiting times between events.

Hands-on (15 min)

Ch. 28, Ex. 3

Using the rate sub-model with intercept, indeg and outdeg (intercept 0.1241, outdeg 0.0889), compute the expected waiting time until the next e-mail for an employee who has written to 20 different colleagues and not yet been written to (in-degree 0), and compare it with the 21 hours of an employee who has written to none. Give hours and days.

Hint The hazard is exp(0.1241 + 0.0889 x out-degree) e-mails per day; the expected waiting time is its reciprocal, exp(-(0.1241 + 0.0889 x out-degree)) days; multiply by 24 for hours.

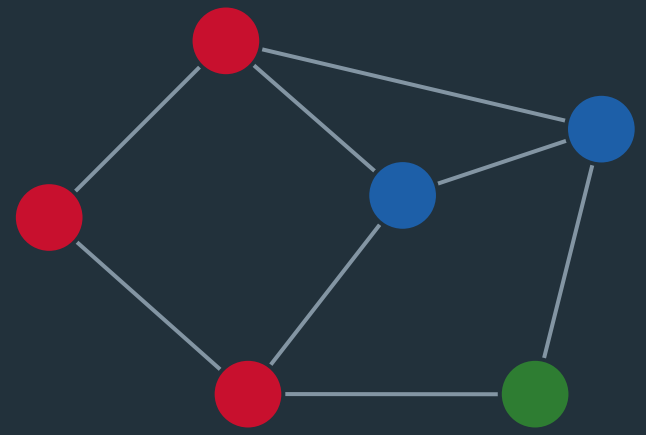
Ch. 28, Ex. 7

Declare the log undirected (nwset sender receiver, undirected eventtime(t) name(mailu)), merge branch, and fit nwdynam mailu, submodel(choice_coordination) inertia same(branch): 1.86 and 0.27, against 3.27 and 0.58 in the directed model. (a) Why are the coordination coefficients not the directed ones? (b) Add recip: what does Stata say? (c) Why can the log likelihoods (-2.12e4 and -1.04e4) not be compared?

Hint (a) The coordination likelihood multiplies a preference in both directions and normalizes over unordered pairs. (b) The error lists the effects that exist. (c) With all coefficients zero each event contributes -ln of the number of alternatives: 128 receivers, or 8,256 unordered pairs.

Break

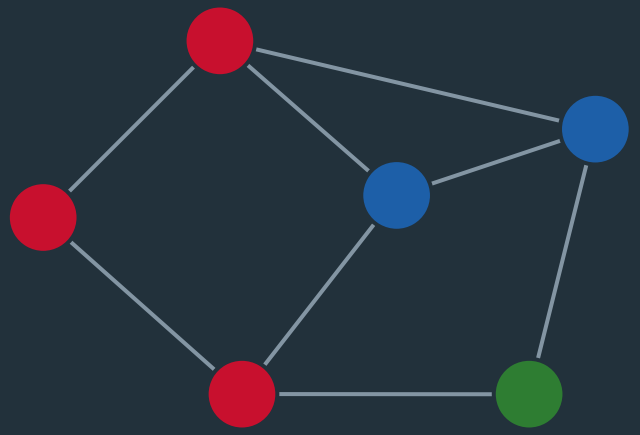
15 minutes



4

Comparing the two models

20 min · Chapters 27 and 28



REM and DyNAM side by side

	nwrem	nwdynam
R reference	`relevent::rem.dyad()`	`goldfish::estimate()`
Question	which pair fires next	which actor acts, and whom they choose
Steps	one: the pair among all pairs at risk	two: rate, then choice
Time	order only; optionally waiting times	order, or exact times with an intercept
Typical effects	nodsnd, nidrec, frpsndsnd, rsndsnd, covevent()	inertia, recip, indeg, outdeg, same()
Intercept	none in the ordinal model	only in the rate sub-model

Written from Butts (2008) and Stadtfeld and Block (2017); both are checked against the R packages.

Same log, two models

```
. nwrem mail, nodsnd nidrec frpsndsnd
   nodsnd |   39.51321
   nidrec |   26.72635
frpsndsnd |   4.406264
```

REM: the pair among all pairs

```
. nwdynam mail, inertia recip indeg
  inertia |   2.036706
    recip |   1.883335
    indeg |    .019011
```

DyNAM choice: the receiver among candidates

Both find the same story: people write to those they have written to before (positive, large, significant), and, in the DyNAM, to those who wrote to them. The sizes are not comparable: the statistics are defined differently, and the choice model conditions on the sender who acts.

Which model for which question?

Who is likely to write at all?

The rate sub-model of nwdynam: sender statistics only.

Whom do they write to?

The choice sub-model of nwdynam, or nwrem with pair statistics.

Does recency or frequency matter?

nwrem: rsndsnd and frpsndsnd, compared by AIC.

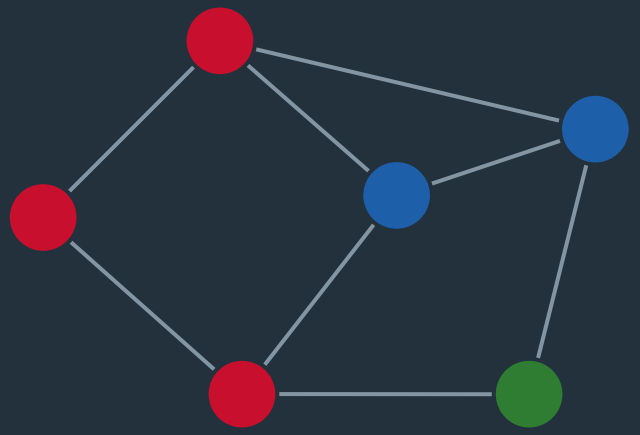
How long until the next event?

The rate sub-model with an intercept: it uses the waiting times.

5

Postestimation: fit, outliers, change points

25 min · Chapter 28



Information criteria and event probabilities

```
. estat ic
      Model |        N           ll       ll/N    df        AIC        BIC
       mail |     2738    -9906.868   -3.61829     4   19821.74    19845.4

. estat eventprob, frame(evp)
  One row per event and candidate receiver: p = probability that the actor is
  the receiver; observed = the realised receiver.
. frame evp: summarize p
    Variable |        Obs        Mean    Std. dev.       Min        Max
           p |    350,464    .0078125    .0229392   .0004076   .5064527
```

1/128

the mean probability: 128 candidates at every event, 350,464 rows

`estat eventprob` gives the probability the model assigned to each candidate at each event. The highest, 0.51, is a candidate the model was very sure of.

Outliers: events the model finds surprising

```
. estat outliers, method(top) parameter(5)
  | event        time   sender   receiver          ll |
  |  1737   10.443461      E45       E101   -7.0065354 |
  |  2031   11.477882     E152       E151   -7.1885935 |
  |  2032   11.478009     E152        E55   -6.9655923 |
  |  2327   12.487662     E136       E153   -7.0069778 |
  |  2631   13.484606     E152        E78   -7.4530721 |

. estat outliers, window(50)
  24 events flagged with Hampel's rule (parameter 3)
```

ll is the log likelihood of one event under the model. The lowest values, about -7, are the events that the model found least likely. Look the senders and receivers up in the employee file.

method(top) flags the five lowest; the default Hampel rule flags the events that fall far below the local median.

Change points: when does the process shift?

```
. estat changepoints, window(20)
  Change points of the log likelihood of the events, pelt method,
  change in the mean, minimum segment length (window) 20
  | event        time   sender   receiver          ll |
  |   180    3.425231     E166       E128   -4.3738517 |
  |   694   5.3855438      E97        E17   -2.0240505 |
  |   797   5.4742818      E30        E73   -5.3070767 |
  |  1480   7.5972338     E156       E121   -1.9201027 |
  |  1517   7.6088538      E98        E45   -4.5015155 |
  |  2214   12.411181      E86       E118   -2.5669688 |
  ... 8 change points in all
```

A change point ends a regime: the events after it fit the model differently. Eight regimes end between day 3.4 and day 12.5. Compare the dates with what you know about the company.

Methods: pelt (default), amoc, binseg. With binseg on the variance, none was found.

Hands-on (10 min)

Try 1

Refit the choice model with inertia, recip and indeg, and list the five events with the lowest log likelihood: estat outliers, method(top) parameter(5). Look the senders up in the employee file. What do they have in common?

Hint Sender E152 appears twice. Merge manufacturing_nodes and look at level, supervisor and division.

Try 2

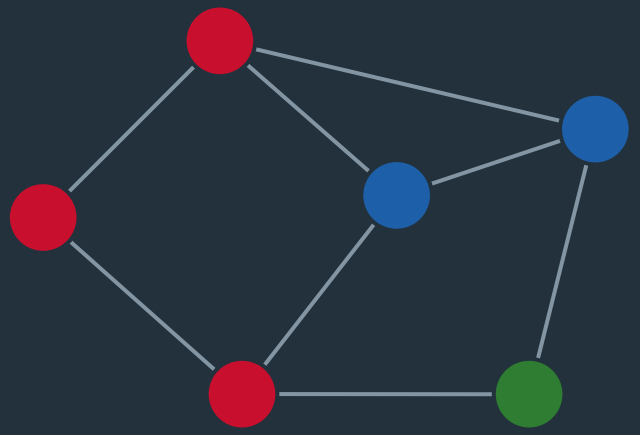
Run estat changepoints, window(20) method(binseg) moment(variance), then again with window(50) and method(pelt). How many regimes does each find?

Hint A larger window means fewer and longer regimes. With binseg on the variance the shorter window found none.

6

Limits, pitfalls and where to go

20 min · Chapters 27 to 29



Limits and pitfalls

Looking for an intercept

The REM compares pairs at the same instant, so none can be estimated.

Quoting nodsnd or nidrec alone

Both shift with the other effects in the model.

indeg and outdeg

In the rate model they describe the acting actor, in the choice model the candidate.

Order or waiting times?

Without intercept the rate model uses the order only; the choice model never has an intercept.

Declaring the event network

nwdynam needs a directed event network; only choice_coordination needs an undirected one.

Simultaneous events

Events with the same time are treated as simultaneous, as in goldfish (the note under the header).

What is refused

```
. capture noisily nwrem mail
nwrem requires at least one effect: nodsnd nidrec nidsnd nodrec ntdegsnd
ntdegrec frpsndsnd frrecsnd covsnd() covrec() covint() covevent() rsndsnd
rrecsnd otpsnd itpsnd ospsnd ispsnd fesnd ferec feint psabba psabby psabxa
psabxb psabxy psabay (see nwrem for what each computes).
invalid syntax
. display _rc
198

. nwrandom 8, prob(.3) name(ordinary)
. capture noisily nwdynam ordinary
nwdynam requires a network declared via nwset's eventtime() option (an
event-type temporal network) - see nwset.
invalid syntax
. display _rc
198
```

For R users

Task	nwcommands	R
Declare an event log	`nwset ..., eventtime()`	`defineNetwork() + linkEvents()`
Relational event model	`nwrem`	`relevent::rem.dyad()`
Dynamic network actor model	`nwdynam`	`goldfish::estimate()`
Information criteria	`estat ic`	`AIC(), BIC()`
Event probabilities	`estat eventprob`	`returnEventProbabilities()`
Outliers and change points	`estat outliers, estat changepoints`	`examineOutliers(), examineChangepoints()`

Equivalents as in Appendix C of the book; the manuals of nwrem and nwdynam give the full translation tables.

Further tools

- **`interval`**
  nwrem: the likelihood that uses the waiting times
- **`riskset(), entry(), exit()`**
  nwrem: restricted risk sets and actors that come and go
- **`netterms()`**
  nwdynam: effects through other networks, windows and weights
- **`submodel(choice_coordination)`**
  nwdynam: events between undirected pairs
- **`Triadic and participation effects`**
  nwrem: otpsnd, itpsnd, psabba and others
- **The manuals**
  nwcommands.org/nwrem/manual and /nwdynam/manual
- **Chapter 29**
  A complete network analysis of the law firm
- **Chapters 27 and 28**
  The relational event models and the DyNAM

To read after the session

Events: Chapters 27 and 28 · Synthesis: Chapter 29

Manuals: nwcommands.org/nwrem/manual · nwcommands.org/nwdynam/manual

Do-files and teaching material: nwcommands.org/book

Instructors are free to use, adapt and share these slides.

