Effect Sizes for Stochastic Actor-oriented Models

Tom A.B. Snijders

University of Oxford
University of Groningen

July, 2022

1. Overview

Contents
Generalities about effect sizes
Change statistics
Question 1
Question 2
Question 3
Implementation in RSiena

2. Effect sizes ....?

Effect sizes aim for **comparability** of parameters.

Comparability may be across variables in the same data set, or between models for the same data set, or across data sets.

In linear regression comparability (all types) can be achieved by using standardized predictors, or equivalently by considering

$$\beta_k \, \text{s.d.}(X_k) \qquad \text{or} \qquad \frac{\beta_k \, \text{s.d.}(X_k)}{\text{s.d.}(Y)} \, .$$

For linear regression, there still is an essential difference between such standardized coefficients ($\sim$ one model) and contributions to R^2 ($\sim$ comparison between models).

The straightforward approach to effect sizes in linear regression models breaks down in most other generalized linear models.

Several approaches possible for defining effect sizes:

- ⇒ Marginal effects:
  expected change in some outcome variable,
  for a ‘unit’ change in explanatory ‘variable’;
- ⇒ Model-based effects:
  definitions within the model,
  making the effect sizes comparable in some way.

Marginal effects are more directly interpretable, because they refer directly to observed variables; they may be complicated because of their dependence on the values of other variables in the model.

Many sociologists are fond of marginal effects for the interpretation of logistic regression results.

Here I present some available techniques
for model-based effects for SAOMs.

3. Change statistics

Consider a SAOM with evaluation function

$$f_i(\beta, x) = \sum_k \beta_k s_{ki}(x) .$$

Define by $\delta_{ij,k}(x)$ the change statistic

$$\delta_{k,ij}(x) \; = \; s_{ki}\big(x^{(+ij)}\big) - s_{ki}\big(x^{(-ij)}\big) ,$$

where $x^{(+ij)}$ and $x^{(-ij)}$ are the networks x
with and without tie $i \rightarrow j$, respectively.

For the model-based effect sizes,
we consider the probabilities in a ministep.

In a ministep, the probability of toggling tie variable x_{ij}, if actor i has the opportunity to make a change, is

$$\pi_{ij}(\beta, x) = \frac{\exp\big(f_i(\beta, x^{(\pm ij)}) - f_i(\beta, x)\big)}{1 + \sum_{h \neq i} \exp\big(f_i(\beta, x^{(\pm ih)}) - f_i(\beta, x)\big)} ,$$

where $x^{(\pm ij)}$ the network in which tie variable x_{ij} is toggled into $1 - x_{ij}$ (and similarly for $x^{(\pm ih)}$).

Note that

$$f_i(\beta, x^{(\pm ij)}) - f_i(\beta, x) = \sum_k \beta_k (1 - 2\, x_{ij})\, \delta_{ij,\, k}(x) .$$

The change statistics (with a + or a –) have the role of the explanatory ~~('independent')~~ variables.

For example:

$$\log\left(\frac{\mathrm{P}\{\text{add tie } i \rightarrow j \text{ to } x\}}{\mathrm{P}\{\text{leave network unchanged}\}}\right) = \sum_k \beta_k\, \delta_{ij,\, k}(x) \qquad \text{if } x_{ij} = 0 \;;$$

$$\log\left(\frac{\mathrm{P}\{\text{drop tie } i \rightarrow j \text{ from } x\}}{\mathrm{P}\{\text{leave network unchanged}\}}\right) = -\sum_k \beta_k\, \delta_{ij,\, k}(x) \quad \text{if } x_{ij} = 1 \;;$$

and

$$\log\left(\frac{\mathrm{P}\{\text{add tie } i \rightarrow j \text{ to } x\}}{\mathrm{P}\{\text{add tie } i \rightarrow h \text{ to } x\}}\right) = \sum_k \beta_k \big(\delta_{ij,\, k}(x) - \delta_{ih,\, k}(x)\big) ,$$

if x does not have the ties $i \rightarrow j$ or $i \rightarrow h$.

4. Question 1: differences in parameter sizes?

Two-mode network for Glasgow data for 14 activities:

		daily	weekly	monthly	less
1	I listen to tapes or CDs	**388**	23	5	16
2	I look around in the shops	**65**	290	48	30
3	I read comics, mags or books	**186**	121	65	60
4	I go to sport matches	**30**	**113**	90	200
5	I take part in sports	**218**	117	30	68
6	I hang round in the streets	**216**	64	26	125
7	I play computer games	**157**	109	45	122
8	I spend time on my hobby (e.g. art, an instrument)	**114**	113	36	170
9	I go to something like B.B., Guides or Scouts	**36**	**81**	1	314
10	I go to cinema	**11**	**81**	269	71
11	I go to pop concerts, gigs	**7**	**6**	**92**	326
12	I go to church, mosque or temple	**2**	**52**	**10**	368
13	I look after a pet animal	**197**	25	6	203
14	I go to dance clubs or raves	**15**	**44**	**104**	266
15	I do nothing much (am bored)	37	39	24	331

Number of students participating in each of a list of activities, summed over three waves, for Glasgow data. Bold-faced are frequencies counted as a tie.

Effect	par.	(s.e.)
rate period 1	4.230	(0.268)
rate period 2	4.046	(0.278)
outdegree	–5.891***	(0.660)
outdegree - activity	0.637***	(0.088)
indegree - popularity (√)	0.790***	(0.100)
out-in degree assortativity	–0.0184***	(0.0025)
4-cycles	0.0389***	(0.0057)

Estimation results for activity participation of Glasgow students.

Question 1: what about the vastly different parameter sizes?

Some kind of standardization

To compare β_k for a given network x,
consider the formulae

$$\pi_{ij}(\beta, x) = \frac{\exp\big(f_i(\beta, x^{(\pm ij)}) - f_i(\beta, x)\big)}{1 + \sum_{h \neq i} \exp\big(f_i(\beta, x^{(\pm ih)}) - f_i(\beta, x)\big)} ,$$

$$f_i(\beta, x^{(\pm ij)}) - f_i(\beta, x) = \sum_k \beta_k (1 - 2x_{ij}) \delta_{ij,k}(x) .$$

This shows that β_k is multiplied by the 'variable'

$$(1 - 2x_{ij}) \delta_{ij,k}(x) .$$

Define

$$\delta_{ii,k}(x) = 0 .$$

In analogy to linear regression,
we can make the values β_k comparable by considering

$$\sigma^2_{ik}(x) = \text{var}\Big\{\delta_{ij,k}(x)(1 - 2x_{ij}) \;\Big|\; i \text{ fixed } \Big\}$$
$$= \frac{1}{n}\sum_{j=1}^{n}\big(\delta_{ij,k}(x)(1 - 2x_{ij})\big)^2 - \Big(\frac{1}{n}\sum_{j=1}^{n}\delta_{ij,k}(x)(1 - 2x_{ij})\Big)^2 ,$$

the within-actor variance of this 'variable'; and

$$\sigma^2_k(x) = \frac{1}{n}\sum_i \sigma^2_{ik}(x) .$$

The product

$$\sigma_k(x)\,\beta_k$$

expresses the parameters β_k, for different k and a given x, on a common scale. The standard deviation is used here as a somewhat arbitrary, but well-known measure of scale.

Answer to Question 1

Effect	$\hat{\beta}_k$	(s.e.)	σ_k	$\sigma_k\,\hat{\beta}_k$
rate period 1	4.227	(0.268)		
rate period 2	4.047	(0.278)		
outdegree	–5.891	(0.660)	0.79	–4.67
outdegree - activity	0.637	(0.088)	5.48	4.33
indegree - popularity ($\surd$)	0.790	(0.100)	6.74	4.29
out-in degree assortativity	–0.0184	(0.0025)	331.54	–6.11
4-cycles	0.0389	(0.0057)	55.08	2.14

$\hat{\beta}_k$: estimates; s.e.: standard errors;
σ_k: mean within-ego standard deviations of change statistics;
$\sigma_k\hat{\beta}_k$: their product.

The order of magnitude of the effects is similar.

5. Question 2: Relative Importance

Effect	par.	(s.e.)
rate (period 1)	14.471	(1.511)
rate (period 2)	11.949	(1.250)
outdegree	–3.419	(0.294)
reciprocity	3.815	(0.315)
GWESP transitive	1.906	(0.125)
GWESP cyclic	0.394	(0.114)
indegree - popularity	–0.008	(0.027)
outdegree - popularity	–0.092	(0.063)
outdegree - activity	0.110	(0.047)
reciprocated degree - activity	–0.279	(0.066)
sex alter	–0.156	(0.091)
sex ego	0.057	(0.106)
same sex	0.591	(0.082)
reciprocity × GWESP transitive	–1.148	(0.167)

How strong are the contributions of the effects
to determining the actors’ choice probabilities?

Relative Importance

Indlekofer & Brandes (*Network Science*, 2013) proposed a measure for the *Relative Importance* of effects.

This is based on the following approach:
how large is the change in the probability vector π_i
if one of the effects is dropped?

Ingredients of the approach:

- Just replace β_k by 0, do not re-estimate.
- Measure change in probability vector by the ℓ_1 distance, i.e., sum of absolute differences.

Define π_i as vector of probabilities for actor i in next ministep, and $\pi_i^{(-k)}$ as the same if effect k obtains a weight of 0.

Then *importance of effect k for actor i* is

$$\left\|\pi_i - \pi_i^{(-k)}\right\|_1 = \sum_j \left|\pi_{ij} - \pi_{ij}^{(-k)}\right| .$$

Relative importance of effect k for actor i is

$$I_k(X, i) = \frac{\left\|\pi_i - \pi_i^{(-k)}\right\|_1}{\sum_{\ell=1}^K \left\|\pi_i - \pi_i^{(-\ell)}\right\|_1} ;$$

Expected relative importance (for random *i*) is

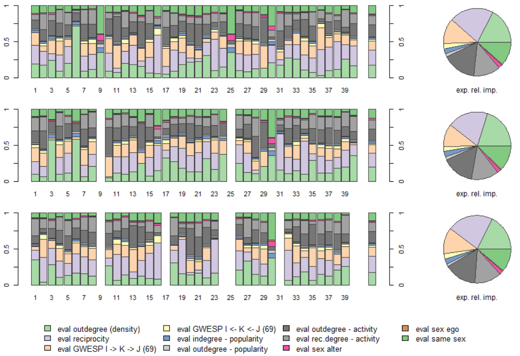
$$\frac{1}{N}\sum_{i=1}^N I_k(X, i) .$$

Expected (raw / total) importance can then be defined as

$$\frac{1}{N}\sum_{i=1}^N \left\|\pi_i - \pi_i^{(-k)}\right\|_1 .$$

Answer to Question 2

Effect	$\hat{\beta}_k$	(s.e.)	Exp. rel. importance wave 1	wave 2	wave 3
outdegree	-3.5904	(0.2161)	0.1730	0.2005	0.1749
reciprocity	3.5827	(0.2503)	0.2115	0.1952	0.2209
GWESP transitive	1.5569	(0.0997)	0.1225	0.1085	0.1267
GWESP cyclic	0.3290	(0.1178)	0.0252	0.0216	0.0253
indegree - popularity	-0.0274	(0.0247)	0.0245	0.0266	0.0232
outdegree - popularity	-0.0197	(0.0453)	0.0148	0.0155	0.0141
outdegree - activity	0.1626	(0.0323)	0.1626	0.1549	0.1534
reciprocated degree - activity	-0.3915	(0.0560)	0.1283	0.1303	0.1306
sex alter	-0.1458	(0.0886)	0.0179	0.0191	0.0172
sex ego	0.0597	(0.0966)	0.0049	0.0054	0.0047
same sex	0.6515	(0.0782)	0.1147	0.1224	0.1091



Previous page:
Plot of relative importance of effects for first 40 actors
(blank = absent) and averaged for all actors (pie-chart).

The graph was produced by

```
plot(RI, actors=1:40, addPieChart = TRUE, legendColumns=4)
```

where RI was the object produced by `sienaRI`.

6. Question 3: size of separate contributions

In a model for network dynamics,
how large are the separate contributions of each included effect?

Entropy-based measures

The uncertainty / variability
in the outcomes of categorical random variables
can be expressed by the *entropy* (Shannon, 1948).

For a probability vector $p = (p_1, \ldots, p_K)$, entropy is defined as

$$H(p) = -\sum_{k=1}^{K} p_k \, {}^2\log(p_k) \ . \qquad (1)$$

Minimum 0 (if one category has $p_k = 1$, outcome is certain);
maximum ${}^2\log(K)$ (if $p_k = 1/K$ for all k, totally random).

This can be transformed to the range [0, 1] by

$$1 - \frac{H(p)}{{}^2\log(K)}$$

so that 0 means total uncertainty and 1 means total certainty.

This can be applied to the probability vector

$$\pi_i(\beta, x) = \Big(\pi_{i1}(\beta, x), \ldots, \pi_{in}(\beta, x)\Big)$$

of choices in the ministep for current network state x.

The degree of determination (certainty), or amount of information,
in the outcome of the ministep for actor i,
for current network state x, can be expressed by

$$R_H(i, \beta, x) = 1 - \frac{H(\pi_i(\beta, x))}{{}^2\log(K)} \ . \qquad (2)$$

For models with constant rate function, this can be averaged:

$$R_H(\beta, x) = \frac{1}{n} \sum_i R_H(i, \beta, x) \ . \qquad (3)$$

For network panel data we may average over waves:

$$R_H(\beta) = \frac{1}{M}\sum_m R_H\big(\beta, x(t_m)\big)\ . \tag{4}$$

This is a measure between 0 and 1:
1 if the outcome of the ministep for a given actor is certain;
0 if all actors choose a random change (all probabilities $1/n$).

This measure was proposed in Snijders (*Mathématiques et Sciences Humaines*, 2004).

Values generally will be low!

Contributions to R_H

For the contribution of an effect, or set of effects,
to the information about the outcome:
estimate the model twice,
giving parameter estimates $\hat\beta_1$ for full model
and $\hat\beta_2$ for restricted model, and consider the difference

$$R_H(\hat\beta_1) - R_H(\hat\beta_2)\ . \tag{5}$$

Note that this is not necessarily positive,
because the estimation method does not maximize $R_H(\hat\beta)$.

Answer to Question 3

As an example we consider the friendly relation for Van de Bunt's data set, waves $t_1 - t_4$.

Effect	par.	(s.e.)
Rate 1	3.94	(0.64)
Rate 2	4.71	(0.81)
Rate 3	3.56	(0.59)
outdegree (density)	–1.545***	(0.208)
reciprocity	2.111***	(0.272)
outdegree-popularity	–0.192***	(0.039)
transitive triplets	0.456***	(0.049)
sex alter	0.384*	(0.191)
sex ego	–0.535*	(0.209)
same sex	0.156	(0.181)
program similarity	0.727***	(0.207)

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$;
Overall maximum convergence ratio 0.07.

The following table gives semi-standardized parameter estimates and relative importance of the effects.

Effect	$\hat{\beta}_k$	(s.e.$_k$)	σ_k	$\sigma_k \hat{\beta}_k$	RImp$_k$	$t_k = \hat{\beta}_k/\text{s.e.}_k$
outdegree (density)	–1.545***	(0.208)	0.58	–0.90	0.22	–7.44
reciprocity	2.111***	(0.272)	0.32	0.68	0.20	7.76
outdegree-popularity	–0.192***	(0.039)	4.41	–0.84	0.08	–4.86
transitive triplets	0.456***	(0.049)	1.55	0.70	0.24	9.21
sex alter	0.384*	(0.191)	0.42	0.16	0.15	2.01
sex ego	–0.535*	(0.209)	0.22	–0.12	0.05	–2.56
same sex	0.156	(0.181)	0.59	0.09	0.04	0.86
program similarity	0.727***	(0.207)	0.34	0.24	0.03	3.52

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$;

Effect sizes for friendly relation between van de Bunt's students, waves $t_1 - t_4$.
$\hat{\beta}_k$: parameter estimate; s.e.: standard error;
σ_k: within-ego standard-deviation of change statistic;
$\sigma_k \hat{\beta}_k$: semi-standardized beta; RImp$_k$: relative importance;
the last three averaged over actors and waves.

The following table shows the degrees of determination (certainty), by wave, for four models:

1. the empty model, with only the outdegree effect;
2. the purely structural model, with the effects of reciprocity, outdegree popularity, and transitive triplets;
3. the model with only covariates, with the three effects of sex and program similarity;
4. the full model.

Model	t_1	t_2	t_3	t_4
Empty model	0.02	0.02	0.02	0.02
Structural model	0.19	0.18	0.17	0.15
Covariate model	0.05	0.05	0.05	0.04
Full model	0.21	0.19	0.20	0.18

The average for the full model is 0.20; the totally structural model comes quite close, whereas the covariate model explains only about one quarter of this value. We conclude that the structural influences are of much greater importance in determining the network dynamics than the covariate influences.

7. sienaRI

These measures are implemented in the function `sienaRI`.

Consult the help page!

Also see Section 13.5 of the RSiena manual.


The horse race of Siena,
il Palio, was originated in year