

## APPENDIX 2

### to Driving factors behind subjective resilience on organic dairy sheep farms

by Augustine Perrin and Guillaume Martin

#### Details on PLS, sPLS and main results

This appendix details the R code used for performing PLS and sPLS.

A complete version of the R code used for the statistical analysis is available upon request to the authors.

The following R code proposes one example for performing one PLS and one sPLS.

#### R code. Example for PLS 1.1

For this example, explanatory variables are slopes and intercepts and interest variables are the evolution of the four dimensions of farmers' satisfaction.

```
X=as.matrix(Data[,c("Ev.UAA", "Ev.AFL", "Ev.Crops", "Ev.Ewe", "Ev.EweProd", "Ev.Lambing", "Ev.Turnout", "Ev.GrazDur", "Ev.FullGrazDur", "Ev.GFeedDur", "Ev.ConcDist", "Ev.LU", "Ev.Pastures", "Ev.StockRate", "Ev.SSFod", "Ev.SSConc", "Ev.WorkLU", "Ev.WorkHa", "I.UAA", "I.AFL", "I.Crops", "I.Ewe", "I.EweProd", "I.Lambing", "I.Turnout", "I.GrazDur", "I.FullGrazDur", "I.GFeedDur", "I.ConcDist", "I.LU", "I.Pastures", "I.StockRate", "I.SSFod", "I.SSConc", "I.WorkLU", "I.WorkHa", "I.LandSat", "I.AnimSat", "I.EconSat", "I.SociSat")])
```

#X is the matrix with explanatory variables

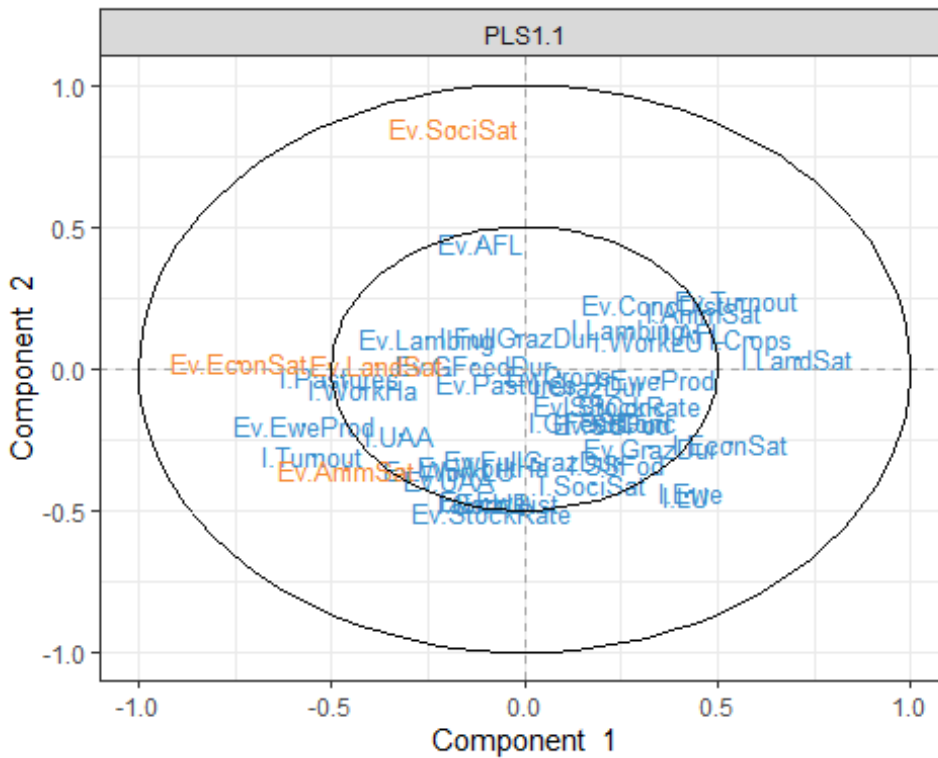
```
Y=as.matrix(Data[,c("Ev.LandSat", "Ev.AnimSat", "Ev.EconSat", "Ev.SociSat")])  
#Y is the matrix with interest variables
```

```
pls = pls(X,Y, ncomp=4, mode="regression")
```

# We perform the PLS to explain and predict Y thanks to X

```
plotVar(pls,cex = c(3.5,3.5),title = "PLS1.1")
```

# We plot variables to illustrate associations. The variables X and Y are represented through their projections onto the plane defined either by X-variables or Y-variables. The variables X and Y being assumed to be of unit variance, their projections are inside a circle of radius 1 centered at the origin called correlation circle. Strongly associated (or correlated) variables are projected in the same direction from the origin. The greater the distance from the origin the stronger the association. Two circumferences of radius 1 and 0.5 are plotted to reveal the correlation structure of the variables (<http://mixomics.org/graphics/variable-plots/plotvar/>).



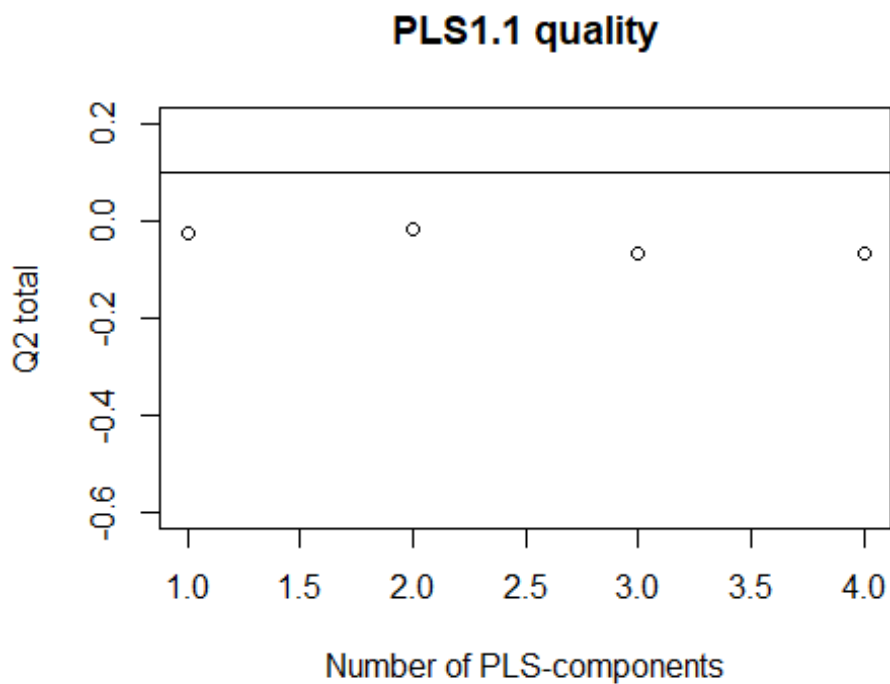
```
par(mfrow=c(1,1))
```

We aim at getting an as satisfying as possible model by performing sPLS after PLS. The quality of the pls and sPLS is assessed through graphs showing Q2 total, Q2 for each predicted variable and each component, R2 for each predicted variable and each component, and MSEP for each predicted variable and each component.

*# Model quality through Q2, R2 and MSEP:*

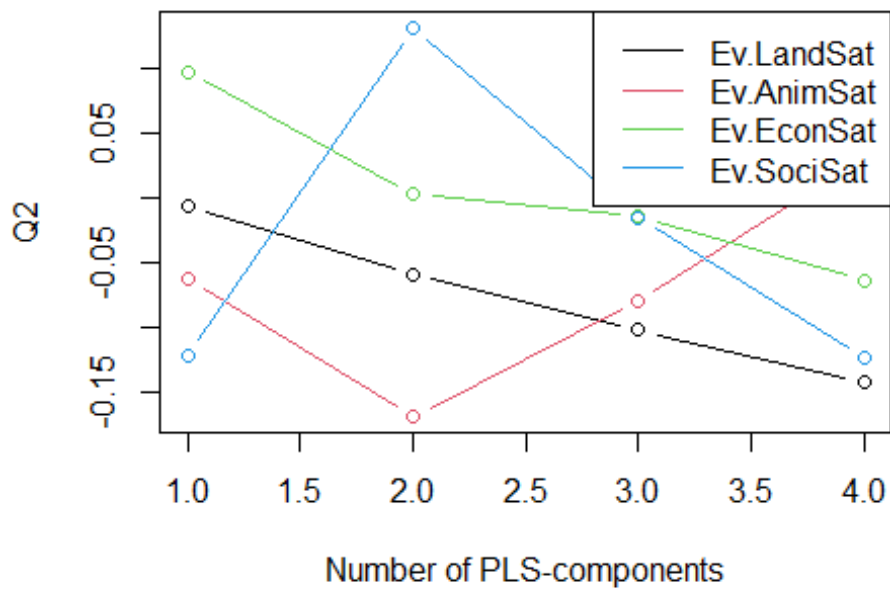
```
tune.pls = perf(pls, validation="Mfold", folds=10, criterion="all", progressBar=F, nrepeat=50000)
```

```
plot(tune.pls$Q2.total, xlab="Number of PLS-components", ylab="Q2 total", main="PLS1.1 quality", ylim=c(-0.6,0.2))
abline(h=0.0975)
```



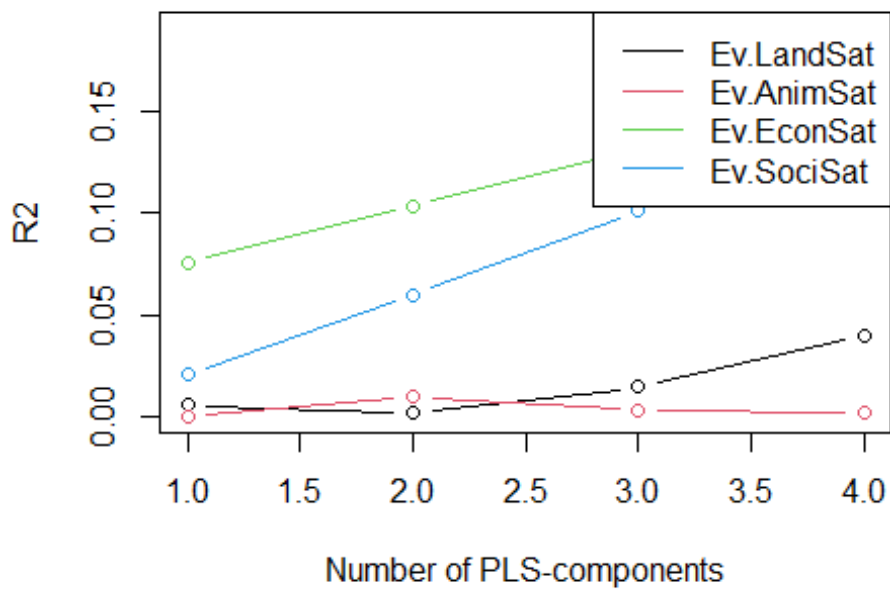
Overall quality of the model is below the ad hoc threshold of 0.0975.

```
monplot <- function(X, nom){
  rang <- range(as.vector(X))
  plot(X[1,], ylim=rang, xlab="Number of PLS-components", ylab=nom, col=1,
  type="b")
  for (i in 2:nrow(X)) {
    lines(X[i,], col=i, type="b")
  }
  legend("topright", rownames(X), col=1:nrow(X), lty=rep(1,nrow(X))) }
monplot(tune.pls$Q2, "Q2")
```



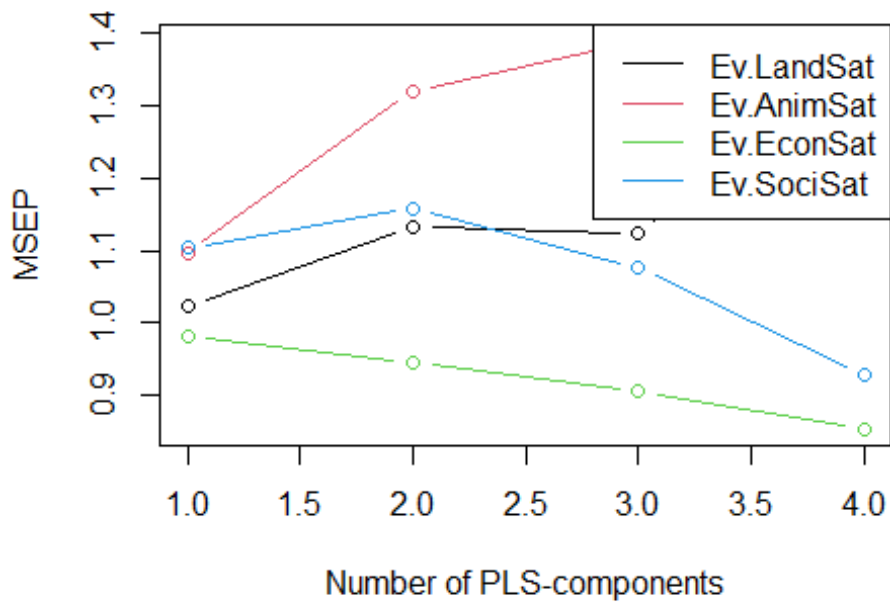
Low prediction quality for each dimension of satisfaction

```
monplot(tune.pls$R2, "R2")
```



Low explanatory quality for each dimension of satisfaction

```
monplot(tune.pls$MSEP, "MSEP")
```



High Mean Squared Error of Prediction.  
 The quality of this PLS is too low.  
 We use the sPLS to improve model quality.

### R code. Example for sPLS 1.1

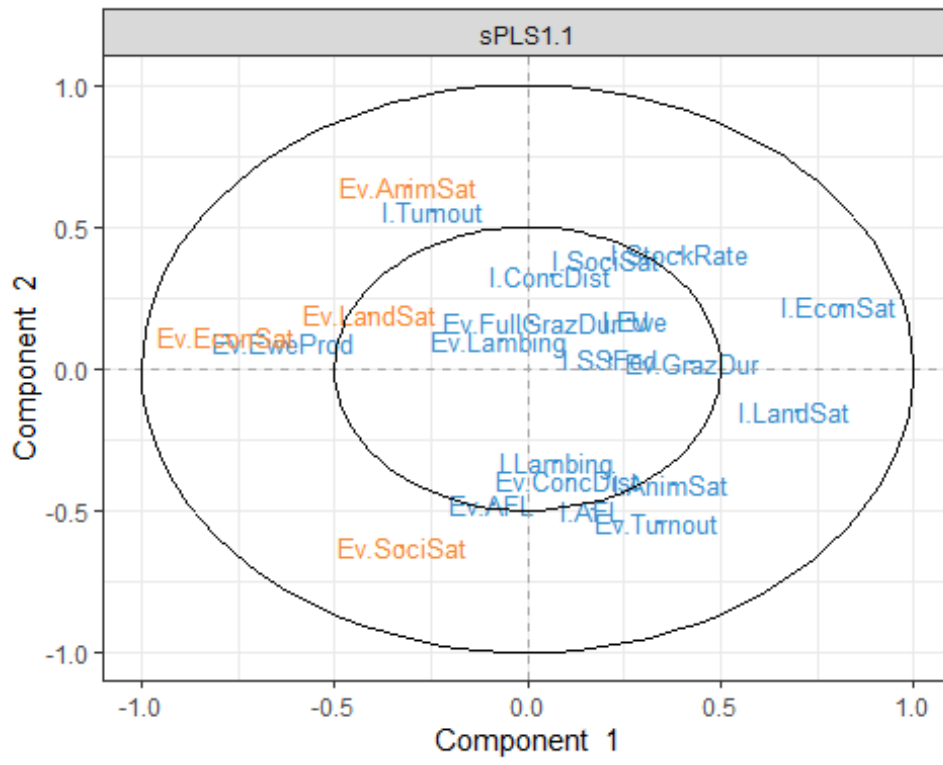
```
#Selection of variables to be kept in the sPLS, from PLS 1.1:
```

```
Tokeep<-tune.spls(X, Y, ncomp = 2)
select.keepX =Tokeep$choice.keepX
```

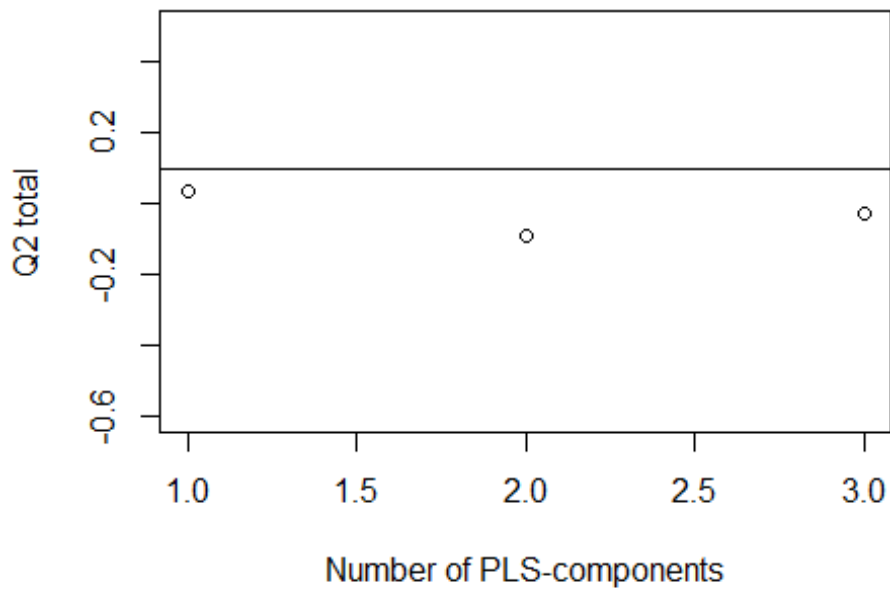
```
#sPLS 1.1
```

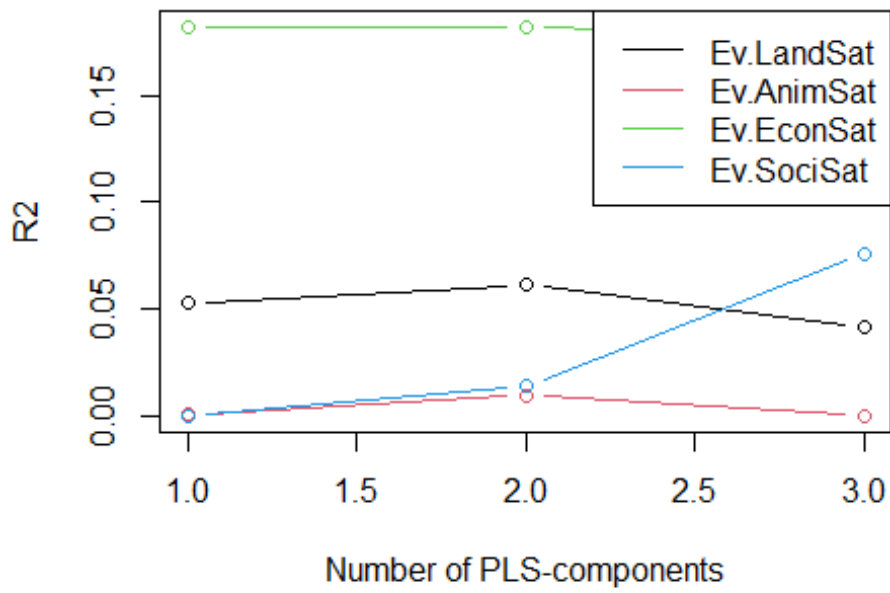
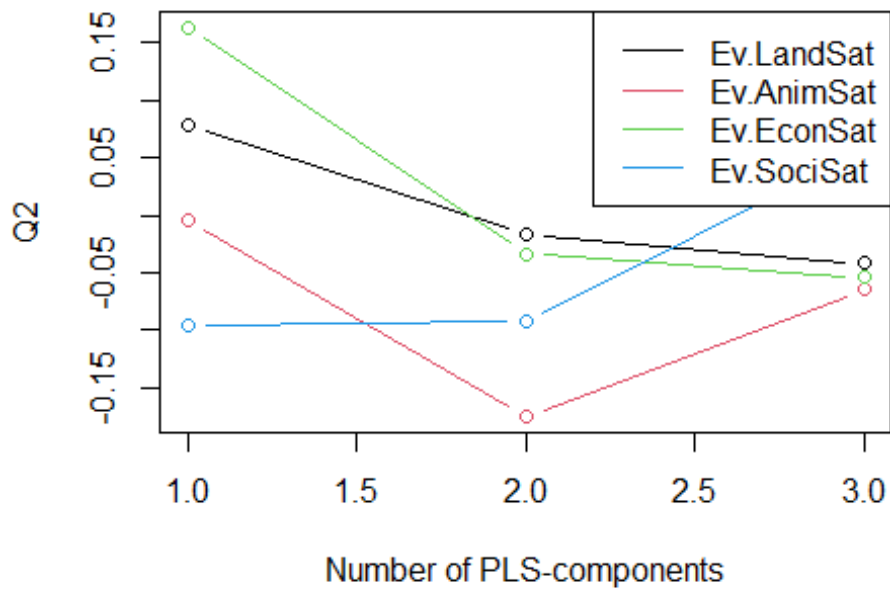
```
Spls = spls(X,Y, ncomp=3, mode="regression",keepX = select.keepX)
```

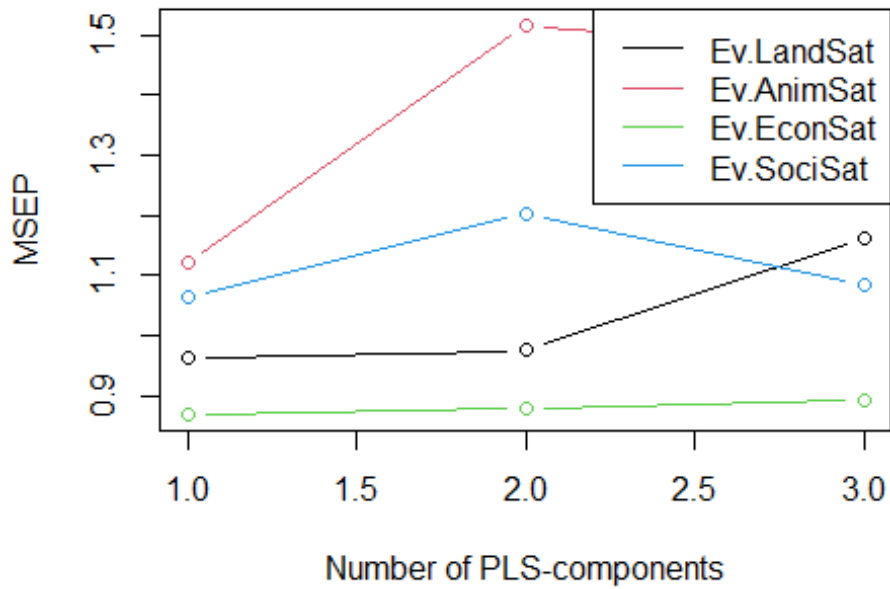
We draw the same graphs as for the PLS (i.e. PlotVar, Q2.total, Q2, R2, MSEP)



**sPLS1.1 quality**







For the sPLS, Q2 is near the threshold on component 1 but often below and variable (the value slightly varies with each execution of the `tune.pls()` line of the code).

With the variables selected, the quality of the PLS and of the sPLS is too low to conclude.



Using the same code as the one provided above, we performed a series of PLS and sPLS.  
The following table summarizes the results.

Model_Name	All_Var	Var_Selected	I_and_Ev	Ev_Only	Four_dim_Sat	Overall_Sat	Quality
PLS1.1	Yes		Yes		Yes		too low
sPLS1.1		Yes	Yes		Yes		low
<b>PLS1.2</b>	<b>Yes</b>		<b>Yes</b>			<b>Yes</b>	<b>Q2 acceptable but variable comp1</b>
<b>sPLS1.2</b>		<b>Yes</b>	<b>Yes</b>			<b>Yes</b>	<b>Q2 high and stable (comp1), unstable (comp2)</b>
PLS1.3	Yes			Yes	Yes		too low
sPLS1.3		Yes		Yes	Yes		too low
PLS1.4	Yes			Yes		Yes	too low
<b>sPLS1.4</b>		<b>Yes</b>		<b>Yes</b>		<b>Yes</b>	<b>Q2 acceptable but variable</b>

**All\_Var:** PLS performed with all variables from the X matrix

**Var\_Selected:** Variables from the X matrix are selected via sPLS

**I\_and\_Ev:** X matrix contains intercept values (I.) and slope values (Ev. for "Evolution")

**Ev\_Only:** X matrix contains slopes only

**Four\_dim\_Sat:** Y contains the values of evolution for the four dimensions of satisfaction (Ev.LandSat; Ev.AnimSat; Ev.EconSat; Ev.Soci.Sat)

**Overall\_Sat:** Y contains the variable "Ev.OverallSat", an aggregation of the four dimensions of satisfaction

**Quality:** quality of the model for the different components (comp) according to obtained quality graphs

sPLS 1.2 presented a high and stable value for Q2 on component 1. On this component, farmers' overall satisfaction increases when:

- 1-Estimated Overall satisfaction is lower at the conversion period
- 2-Ewe productivity increases
- 3-The duration of the grazing period decreases

Q2 was sometimes over the threshold on component 2 but was unstable

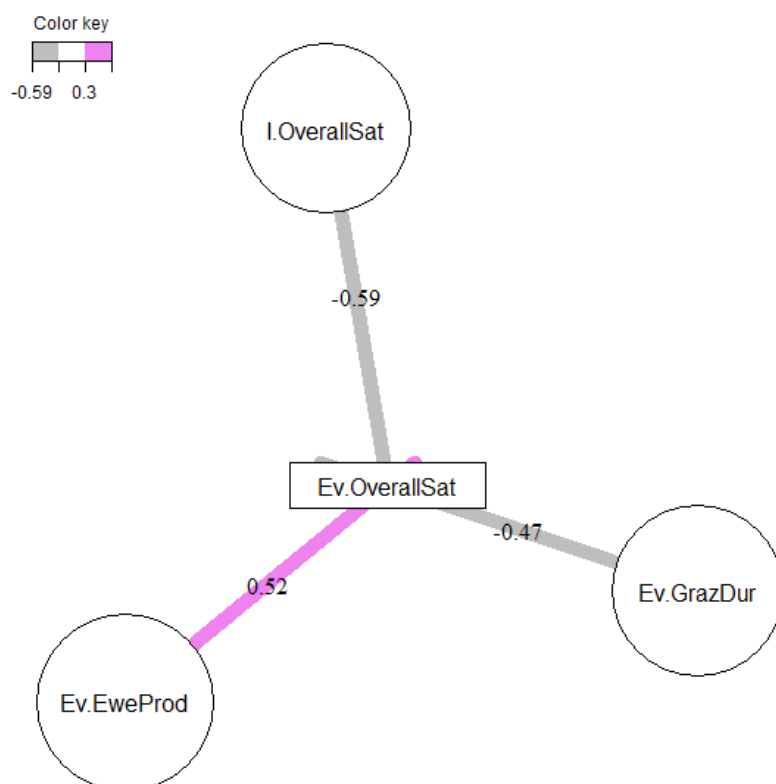
## Networks for association scores

We draw networks to know association scores between response and explanatory variables.

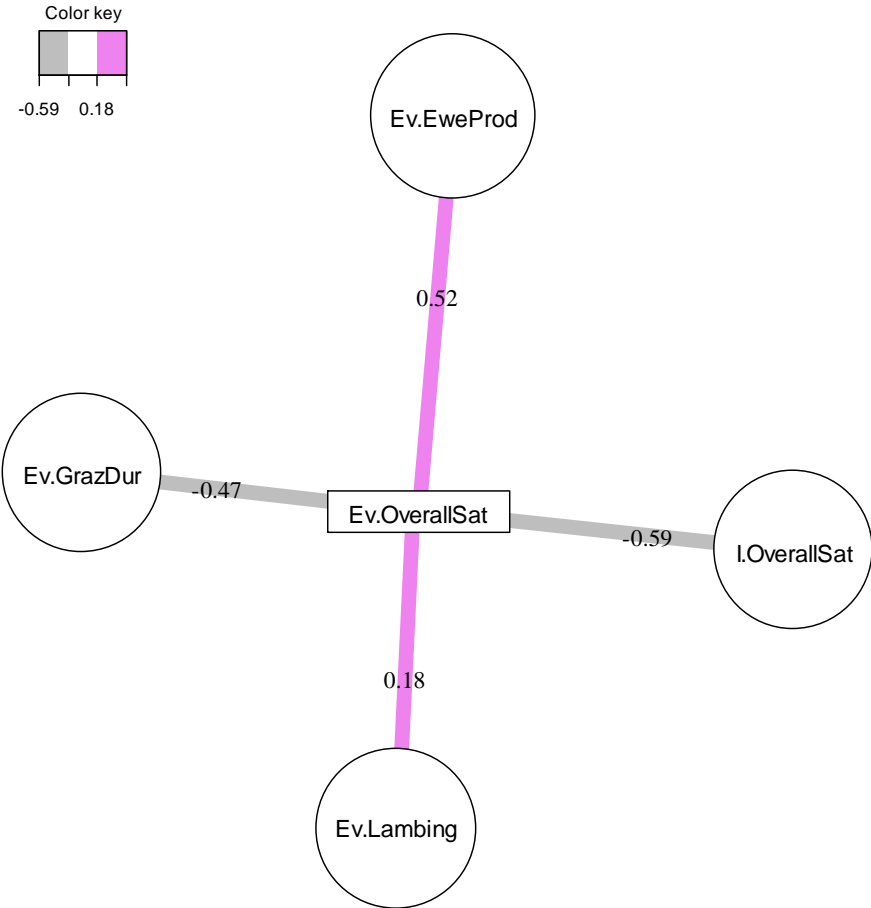
R code for networks:

```
network(Spls,  
        comp=1,  
        color.node = c("dodgerblue2","orange1"),  
        show.edge.labels=TRUE,  
        cutoff = 0.3,  
        lwd.edge=6,  
        interactive = TRUE,  
        color.edge = rep(c("gray","violet"))))
```

Corresponding graph:



If we decrease the association score (cutoff=0.178), we obtain the following graph:



"Ev.Lambing" appears as a fourth explanatory variable.