

Non-stationary modeling of tail dependence of two subjects' concentration

Kshitij Sharma*^{†‡} and Valérie Chavez-Demoulin[†] and Pierre Dillenbourg[‡]

[†]*Faculty of Business and Economics*
University of Lausanne

[‡]*CHILI Lab*
EPFL, Switzerland

Abstract

We analyse eye-tracking data to understand how people collaborate. Our dataset consists of time series of measurements for eye movements, such as spatial entropy, calculated for each subject during an experiment when several pairs of participants collaborate to accomplish a task. We observe that pairs with high collaboration quality obtain their highest values of concentration (or equivalently lowest values of spatial entropy) occurring simultaneously. In this paper, we propose a flexible model that describes the tail dependence structure between two subjects' entropy when the pair is collaborating. More generally, we develop a generalized additive model (GAM) framework for tail dependence coefficients in the presence of covariates. As for any GAM-type model, the methodology can be used to predict collaboration quality or to explore how joint concentration depends on other cognitive operations and varies over time.

Key words: Collaborative learning; Copulas; Entropy; Generalized Additive Models; Tail dependence.

*Corresponding author: kshitij.kshitij@unil.ch.

1 Introduction

Recent developments of quantitative methodologies for education research have emerged under the field of “Learning Analytics” (see Appendix A for formal definition). In this context, eye-trackers unveil major cognitive processes when analyzing the resulting gaze traces (see Appendix A for more details). In this paper, we analyze eye-tracking data to understand how people collaborate. The data consist of gaze traces of 33 pairs of subjects who collaborate to accomplish a concept map task (see Appendix A for formal definition). The two subjects are working synchronously online, looking at the same display on two different screens. During the concept map activity the participants could not see each other, but they could talk to each other. Moreover, their screens were completely synchronised, that is, the participants could see each others’ action and when one peer scrolled the screen the partner’s viewport was updated automatically. From the gaze traces, time series of indicators regarding subject’s attention can be measured. These measures, formally defined later in Section 4, are the following. “Spatial entropy” (SE) reflecting the concentration of the subject. A very low SE value means high concentration; “Stability” indicates whether the two subjects look each at the same region of the screen over two consecutive time windows. The region of the screen is not necessarily the same for each subject. “Similarity” is a continuous measure, high if the subjects look at the same objects at the same time. At the end of the experiment, every pair was classified into a high collaboration quality pair or a low collaboration quality pair, based on the output of the accomplished task and not on the eye-tracking data. A primary question of interest is: can we use the eye-tracking data to distinguish good and bad pairs in terms of their collaboration? Figure 1 provides an answer to this question: it shows a scatterplot of the two subjects’ SE time series (grey points) for the 17 pairs with low collaboration quality (left panel) and for the 16 pairs with high collaboration quality (right panel). The black points correspond to SE values for which both stability and similarity are high. On one hand, the left panel shows no dependence structure between the two subjects with low collaboration quality. Additionally, both the high stability and similarity with high SE values seem to happen randomly for these bad collaborative pairs. On the other hand, the right panel, concerning high collaboration quality pairs, shows a lower tail dependence structure. This means that the sequence of high concentration for the two subjects happens together, whereas their average SE does not show a specific structure. The main (cognitive) reason is that “normal subjects’ behavior” includes many reactions other than that of collaborating. On the contrary, sequences of high concentration, for instance, involve responding to the collaborative aim only. Further, the highest synchronized concentration times are joined with a high level of similarity and stability as highlighted by the black points. The question we address in this paper is: How does the probability that the highest concentrations of the two subjects arise at the same

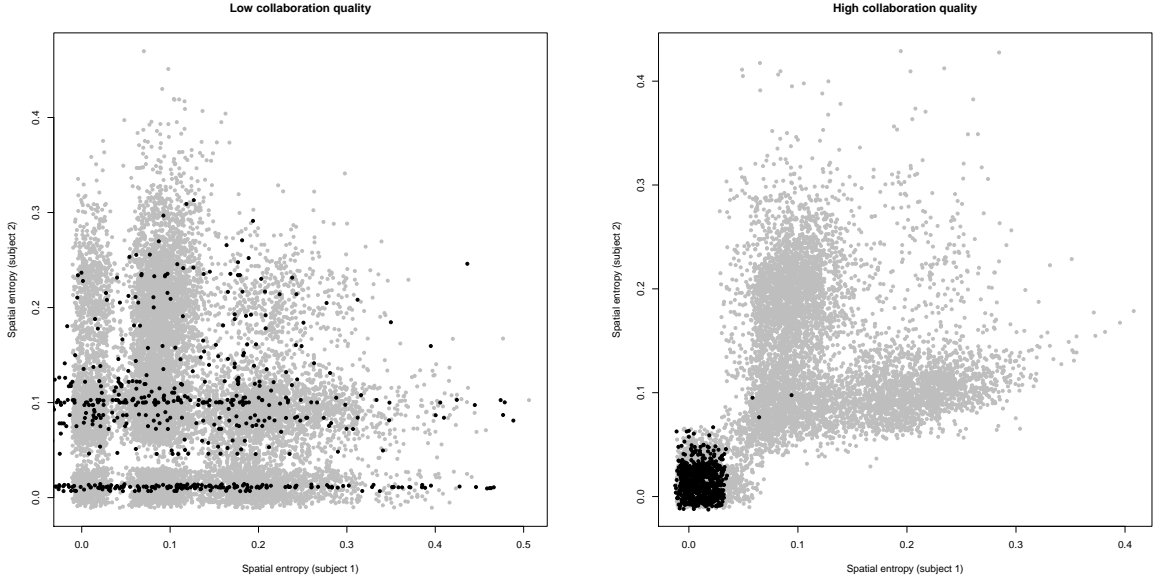


Figure 1: Scatterplot of jittered subjects’ spatial entropy (grey points) for low collaboration quality pairs (left panel) and high collaboration quality pairs (right panel). The black points indicate high level of stability and similarity of the pair subjects.

time vary over time and change with some cognitive characteristics such as similarity and stability, measured at each time from the gaze data? Answering this question might aid our understanding of the cognitive processes underlying the collaboration. Understanding cognitive processes responsible for high collaboration quality enables us to provide feedback to those pairs that are not collaborating well. Providing feedback to learners is one of the most important aims of learning analytics. The more general question we address is: How to model lower tail dependence structure in the presence of covariates and possible non-stationarity? Note that the model we develop can be used for prediction purposes. This is not addressed in this paper but it may be interesting to detect collaboration quality after a certain period of time when a pair of subjects are working together. Their coincidence of high concentration can also be predicted from the significant cognitive characteristics. Letting X_1 and X_2 being the SE of subjects 1 and 2 respectively, we use the coefficient of lower tail dependence

$$\lambda_\ell(X_1, X_2) = \lim_{q \rightarrow 0} \mathbb{P}\{X_2 \leq F_2^{-1}(q) \mid X_1 \leq F_1^{-1}(q)\}. \quad (1)$$

Dependencies of extreme events have received increasing attention in many areas from insurance and finance to environmental contexts. In practice, the tail dependence coefficient represents the most standard way to describe the amount of extremal dependence. It was first introduced in Sibuya (1960) and measures the probability of one extreme value event occurring, given that another event assumes an extreme value.

Works on estimating the tail coefficient include Schmidt and Stadtmueller (2006), Ferreira (2013), Embrechts et al. (2003), Joe (1997), and references therein. Li (2016) proposes a general Bayesian approach for directly modeling tail dependence as explicit functions of covariates. Gardes and Girard (2015) introduces a nonparametric estimator of the conditional tail copula. In many applications, the tail dependence coefficient is likely to be time-varying and to depend on covariates. In an eye-tracking context, explaining how the probability of the coinciding extremes of subjects' SE varies over time until the task ends and changes with some characteristics measured from the gaze data is of primary importance. It not only provides a tool to classify low and high collaboration quality but it also serves in exploring and understanding cognitive aspects of collaboration.

In this paper, we describe a flexible, non-stationary regression-type approach under which the tail dependence coefficient depends on covariates in a parametric or non-parametric way, letting the data decide for the functional form. The method is based on the generalized additive models (GAMs) for dependence structures of Vatter and Chavez-Demoulin (2015), adapted here to tail dependence coefficients. The tail dependence (1) is not observable and to construct a flexible GAM-based framework for the tail dependence, we need to rely on a (penalized) likelihood. To do so, we use the link between the tail dependence coefficient (1) and the copula C of X_1 and X_2 , that is

$$\lambda_\ell = \lim_{q \rightarrow 0} \frac{C(q, q)}{q}. \quad (2)$$

More precisely, there exists a mapping between the tail dependence coefficient and the parameter of the copula. This mapping is convenient for two main reasons: first, the copula provides a density from which we can penalize the likelihood to estimate non parametric smoothed functions; second, it uses the information provided from the entire data to fit a covariate-dependent tail coefficient. In practice the copula can be selected by AIC and in our collaborative data application, not surprisingly when looking at the right panel of Figure 1, the Clayton copula is chosen. The proposed construction benefits from the flexible GAM framework.

The rest of the paper is organized as follows: in Section 2 we describe the GAM for tail dependence coefficients and provide details of the estimation procedure of a penalized log-likelihood based on copula density. Section 3 performs an illustrative methodology assessment in a simulation study. In Section 4 we describe the data collection experiment and apply the method to the gaze data where a model for the tail dependence of pairs' SE is fitted and allows an exploratory dependence structure based on time and pairs' characteristics. We conclude in Section 5.

2 Generalized additive modeling of tail dependence

The tail dependence coefficients are measures of extremal dependence that quantify the dependence in the upper and lower tails of a bivariate distribution of two random variables X_1 and X_2 with continuous marginal distributions F_1 and F_2 . The coefficients are defined in terms of quantile exceedences, and when the limit exists, the coefficient of lower tail dependence is defined in (1) above and analogously, the coefficient of upper tail dependence is

$$\lambda_u(X_1, X_2) = \lim_{q \rightarrow 1} \mathbb{P}\{X_2 > F_2^{-1}(q) \mid X_1 > F_1^{-1}(q)\}. \quad (3)$$

The tail dependence coefficients do not depend on the margins F_1 and F_2 but solely on the dependence structure of the bivariate random vector, that is, the copula. The copula theory addresses the dependence structure of multidimensional random vectors. More precisely, the copulas are functions that “couple” multivariate distribution functions to their corresponding margins. In the next section, we see how we benefit from the link between copula and tail dependence to model flexible non-stationary tail dependence using the entire dataset, in the bivariate context.

2.1 From copula to tail dependence

A copula is a multivariate distribution function $C : [0, 1]^d \rightarrow [0, 1]$ with standard uniform margins. In this paper, considering pairs of subjects collaborating together, we restrict ourselves to the bivariate case ($d = 2$). Copulas allow the construction of joint distributions with arbitrary margins. Textbooks on copulas include Joe (1997) and Nelsen (1999), for instance, and McNeil et al. (2005), Chapter 5. An important theorem is given by Sklar (1959), written below for the 2D case.

Sklar’s Theorem: *Let F be a joint distribution function with margins F_1, F_2 . There exists a copula such that for all x_1, x_2 in $[-\infty, \infty]$*

$$F(x_1, x_2) = C\{F_1(x_1), F_2(x_2)\}$$

If the margins are continuous, then C is unique. Conversely, if C is a copula and F_1, F_2 are univariate distribution functions, then F defined above is a bivariate distribution function with margins F_1, F_2 .

Most of the known copulas are members of parametric families. Implicit copulas, such as the Gaussian copula or t copula, are copulas in well-known parametric distributions; the Gaussian copula supposes (X_1, X_2) being the standard bivariate normal with correlation matrix Σ and for $(u_1, u_2) \in [0, 1]^2$, it is defined as

$$C_{\Sigma}^{\text{Ga}}(u_1, u_2) = \mathbb{P}\{X_1 \leq \Phi^{-1}(u_1), X_2 \leq \Phi^{-1}(u_2)\}. \quad (4)$$

The t copula is implicitly built in a similar way. Archimedean copulas have simpler closed forms and are therefore useful in practice. The Gumbel copula is

$$C_{\beta}^{Gu}(u_1, u_2) = \exp \left[- \left\{ (-\log u_1)^{\beta} + (-\log u_2)^{\beta} \right\}^{1/\beta} \right],$$

with $\beta \geq 1$. When $\beta = 1$, we get independence and $\beta \rightarrow \infty$ provides perfect positive dependence. Another Archimedean copula useful in our eye-tracking data context is the Clayton Copula

$$C_{\beta}^{Cl}(u_1, u_2) = \left(u_1^{-\beta} + u_2^{-\beta} - 1 \right)^{-1/\beta}, \quad (5)$$

with $\beta > 0$. When $\beta \rightarrow 0$, it tends to independence, whereas $\beta \rightarrow \infty$ gives perfect positive dependence. Other Archimedean copulas and other families of copulas exist as in McNeil et al. (2005), Chapter 7 and the references therein.

One drawback considering Pearson's correlation is that it depends on the marginals (F_1 and F_2) as well as on the copula. This explains why it is not invariant on strictly increasing transformations of X_1 and X_2 , an undesirable property for a dependence measure. There exist standard and useful dependence measures solely related to the copula and not to the margins. These are, for instance, rank correlations like Kendall's tau $\tau(X_1, X_2)$ and Spearman's rho $\rho_S(X_1, X_2)$ or tail dependence coefficients like (3) and (1). In the rest of the paper, we simplify the notations using $\eta = \eta(X_1, X_2)$ whenever η is a (tail) dependence measure between the two variables X_1 and X_2 . Convenient mappings between such measures and the parameters of common copulas often exist. For the lower tail dependence coefficient (1) the link is expressed in (2). The upper tail dependence coefficient (3) is a function of the copula through

$$\lambda_u = \lim_{q \rightarrow 1} \frac{1 - 2q + C(q, q)}{1 - q}.$$

Therefore, considering the limit (2) and the one above, if the limits exist $\lambda_u \in [0; 1]$ and $\lambda_l \in [0; 1]$. When $\lambda_u > 0$, we say that we have upper tail dependence. When $\lambda_u = 0$, we say that we have asymptotic independence in the upper tail. When $\lambda_l > 0$, we say that we have lower tail dependence and if $\lambda_l = 0$, we say that we have asymptotic independence in the lower tail. For elliptical copulas, $\lambda_u = \lambda_l$ and the Gaussian copula is asymptotically independent for $|\rho| < 1$, so that for the Gaussian copula $\lambda_u = \lambda_l = 0$. When the copula shows tail dependence (lower or upper or both), there is an explicit mapping between the tail dependence and the copula parameters, provided that we assume a parametric copula family. Let us denote the mapping

$$\lambda = \nu(\beta). \quad (6)$$

For instance, the t copula is tail dependent when $\rho > -1$ and the mapping is

$$\lambda = 2\bar{t}_{df+1} \left(\sqrt{df+1} \sqrt{1-\rho} / \sqrt{1+\rho} \right), \quad (7)$$

where \bar{t} denotes the survival of the t distribution and df its degrees of freedom. The Gumbel copula is upper tail dependent for $\beta > 1$ and the mapping is

$$\lambda_u = 2 - 2^{1/\beta}. \quad (8)$$

The Clayton copula is lower tail dependent for $\beta > 0$ and the mapping is

$$\lambda_\ell = 2^{-1/\beta}. \quad (9)$$

The Clayton copula is specifically the one selected in our application for subjects' extreme concentration, measured by very low SE. In this case, to get a model for λ_ℓ depending on covariates, we use its link (9) with the Clayton copula. By writing the copula likelihood in terms of λ_ℓ we get a likelihood-based approach for λ_ℓ on which formal inference follows. Note that the inference could have been carried out directly on the copula parameter but it is preferable to model the lower tail dependence (1) because it is a more commonly used quantity in practice and its interpretation is simpler. In the next section, we describe a new methodology to model such conditional tail dependence in a very flexible way.

2.2 A flexible model for conditional tail dependence

As a natural extension of generalized linear models (GLM), the popular generalized additive models (Hastie and Tibshirani, 1990) link the mean behaviour of a random variable X with a set of covariates $\mathbf{W} \in \mathbb{R}^q$ through

$$\mathbb{E}(X \mid \mathbf{W} = \mathbf{w}; \theta) = g \left\{ \mathbf{y}^\top \boldsymbol{\beta} + \sum_{k=1}^K h_k(t_k) \right\}, \quad (10)$$

where

- g is a link function,
- (y_1, \dots, y_p) and (t_1, \dots, t_K) are (not necessarily disjoint) subsets (of respective size p and K) of $\{w_1, \dots, w_q\}$ or products if we consider interactions of the variables $\{w_1, \dots, w_q\}$,
- $\boldsymbol{\beta} \in \mathbb{R}^s$ is a vector of parameters, and
- $h_k : \mathbb{H}_k \rightarrow \mathbb{R}$ are smooth functions supported on closed $\mathbb{H}_k \subset \mathbb{R}$, for all k .
- $\theta \in \Theta$ is the column vector of stacked parameters containing both $\boldsymbol{\beta}$ and the information encoding h_k .

Built on roughness penalty smoothing, a generalized additive model provides a flexible model for a univariate response. In a bivariate context, if we are interested in the dependence between X_1 and X_2 that both can be individually explained by covariates of \mathbf{W} , it is very likely that their dependence structure also depends on some of the

covariates in \mathbf{W} . Vatter and Chavez-Demoulin (2015) address this problem and extend GAMs to the dependence structure between random variables. In this paper we adapt the GAM methodology to a tail dependence structure. The tail dependence can be either λ_u or λ_ℓ , so we denote it by λ in what follows. Using the same notations as for the classical model (10), we suppose that λ depends on a set $\mathbf{W} \in \mathbb{R}^q$ of covariates through the semi-parametric form

$$\lambda(X_1, X_2 \mid \mathbf{W} = \mathbf{w}; \theta) = g \left\{ \mathbf{y}^T \boldsymbol{\beta} + \sum_{k=1}^K h_k(t_k) \right\}, \quad (11)$$

where $g(x) = e^x / (e^x + 1)$ is the inverse logit function forcing λ to be in $[0, 1]$, as required. In our collaboration application, typically the factor stability with two levels (high and low) will be one of the w 's with full parametric form and the covariates similarity and time may be each one of the t 's. We assume that each smooth function $h_k \in \mathcal{C}^2(\mathbb{H}_k)$ admits a finite s_k -dimensional basis parametrized by $\mathbf{h}_k = (h_{k,1}, \dots, h_{k,s_k})^\top \in \mathbb{R}^{s_k}$ and a quadratic penalty representation $\int_{\mathbb{H}_k} h_k''(t)^2 dt = \mathbf{h}_k^\top S_k \mathbf{h}_k$, where S_k is a uniquely determined symmetric matrix (Green and Silverman, 1994; Hastie and Tibshirani, 1990; Wood, 2006). The class of C^2 smoothers is broad and allows many flexible smoothers with a finite quadratic penalty representation such as cyclic cubic splines, natural cubic splines and tensor product splines (Wood, 2006). All are implemented in the R package `mgcv`.

Using the mapping (6) between the copula parameter β and the tail dependence coefficient λ as in (7), (8), and (9), for \mathbf{w} and $\theta \in \Theta$, the copula parameter is

$$\beta(\mathbf{w}; \theta) = \nu^{-1} \{ \lambda(\mathbf{w}; \theta) \},$$

where ν^{-1} is continuous and strictly increasing. The model (11) is estimated using the penalized log-likelihood GAM framework based on the copula density c and the log-likelihood function

$$\ell_0(u_1, u_2, \mathbf{w}; \theta) = \log \left(c \left[u_1, u_2; \nu^{-1} \{ \lambda(\mathbf{w}; \theta) \} \right] \right). \quad (12)$$

Considering a sample of n observations $\{u_{i1}, u_{i2}, \mathbf{w}_i\}_{i=1}^n$, the parameters can be estimated by maximizing the penalized log-likelihood

$$\ell(\theta, \boldsymbol{\gamma}) = \ell(\theta) - \frac{1}{2} \sum_{k=1}^K \gamma_k \int_{\mathbb{H}_k} h_k''(t_k)^2 dt_k, \quad (13)$$

with $\ell(\theta) = \sum_{i=1}^n \ell_0(u_{i1}, u_{i2}, \mathbf{w}_i; \theta)$, $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_K)^\top$ and $\gamma_k \in \mathbb{R}_+ \cup \{0\}$ for all k . On each component, the integral term is a roughness penalty and γ_k is a smoothing parameter controlling the roughness of the smoothing function with higher values yielding smoother curve. A common related quantity is the effective degrees of freedom

(edf) defined as $\text{tr}(\mathbf{I} + \gamma_k \mathbf{S}_k)$ for each smooth function h_k (Hastie and Tibshirani, 1990). The penalized maximum log-likelihood estimator is defined as

$$\hat{\theta}_n = \underset{\theta \in \Theta}{\text{argmax}} \ell(\theta, \gamma).$$

The estimation procedure is an iteratively reweighted ridge regression problem (Green, 1987). The mapping with the copula parameters for both the dependence coefficients and tail dependence coefficients being functions with similar properties we refer the interested reader to Vatter and Chavez-Demoulin (2015) for more details on the asymptotic properties of the penalized maximum likelihood estimator. Confidence intervals are based on the \sqrt{n} -consistency and the asymptotic normality of the penalized maximum likelihood estimator can also be applied to compare nested models using likelihood-ratio statistics.

3 Simulation study

In this section, we illustrate the methodology through a simulation study using a Clayton copula used in our eye-tracking application. For a more extensive simulation study in the context of dependence coefficients see Vatter and Chavez-Demoulin (2015). We assume that the true underlying model has a time-varying lower tail dependence coefficient of three different forms, as shown in the top panel of Figure 3. We generate three random covariates from a Gaussian copula of dimension 3 and equicorrelation 0.5 and compute the lower tail dependence using the three deterministic forms. Using the inverse of the mapping (9), we generate pseudo-samples from the corresponding Clayton copulas and apply our methodology. The bottom panels of Figure 3 show one sample of pseudo-observations. The pseudo-observations graph may look useless at first glance, but it importantly illustrates the need for a flexible method when no knowledge of the data-generating process is available. Indeed, although some linear dependence is visible, it is impossible to guess the underlying existing time-varying feature. We simulate 500 datasets of 1000 observations for each form and apply our methodology to estimate lower tail dependence coefficients. The results are shown in Figure 3. The 95% bootstrap confidence intervals (dashed lines) are rather narrow and it is difficult to distinguish the estimated curves (dotted lines) from the true ones (straight lines).

4 Models for eye-tracking data

The description of the data collection experiment and the definition of the variable (entropy, to appear later in this section) contained in the next two paragraphs are borrowed from Sharma et al. (2017).

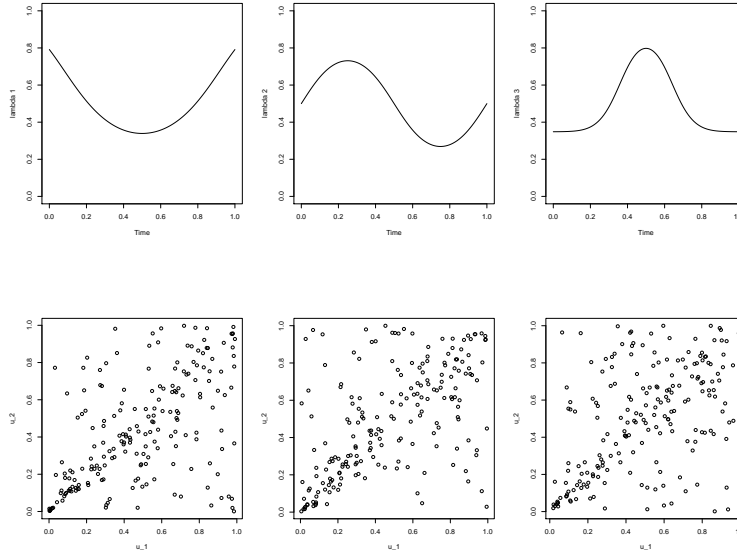


Figure 2: Clayton copula with time-varying λ_ℓ (top panels) with a quadratic form (left), sinusoidal form (middle), and exponential form (right) and a corresponding sample of pseudo observations (bottom panels).

Sixty-six master students (20 female) from the École Polytechnique Fédérale de Lausanne participated in the present study. The participants were rewarded with CHF 30 for their participation. The participants came to the laboratory in pairs. Once the participants signed the consent form, they took an individual pretest (True/False questions). They then watched a video ^{1 2} (duration, 17 minutes and 5 seconds) and collaborated on a concept-map about the video content (duration, 10–12 minutes). A concept map is a graphical representation (organisation) of a given body of knowledge, where the nodes are the key concepts, which constitute the knowledge and the edges are the conceptual link between them. Concept maps are often used in education as a method to summarise, retain, and communicate knowledge and are also used for collaborative knowledge modelling. An example of a concept map is shown in Figure 4. To create the concept maps, the master students used IHMC CMap tools ³. The IHMC CMap tool is a web platform to “*construct navigate share and criticize knowledge models represented as concept maps*”. This tool enables the teams to collaboratively create and edit concept maps remotely. All the actions of a participant are made visible to the peers. This helps in keeping the two screens synchronised without having

¹‘Resting Membrane Potential-Part 1’

²‘Resting Membrane Potential-Part 2’

³<http://cmap.ihmc.us>

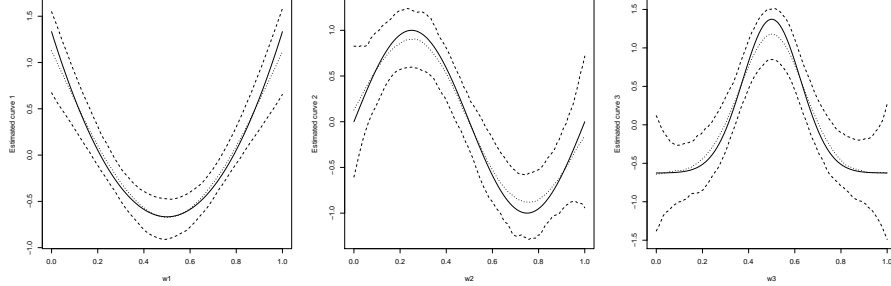


Figure 3: Simulation results for the Clayton copula with with three additive components: quadratic form (left), sinusoidal form (middle), and exponential form (right). True curve (straight line), mean estimated (dotted line), and 95%-confidence intervals (dashed lines).

to additionally build such a concept map tool. Finally, the master students took an individual posttest (True/False questions). During the concept-map phase, the participants were allowed to talk and their screens were therefore synchronized, that is, they could see each others' actions. The final concept maps were then compared against an expert-map (made by two experts) and were scored as follows: one mark each for the correct connection between the concepts and for the correct label for the connections; and, one half-mark for the partially correct label for the connections. The pairs were then divided into two levels based on their concept map score using a median split. We tested the bimodality for the distribution of the concept map scores through a mixture of two Gaussian distributions, which confirms that the median cut divides the pairs into two correct categories. For more details about the data see Sharma et al. (2017).

There are two options to measure how much a person is concentrating during a given task. One is fixation duration and the second is spatial entropy. High fixation duration tells us how much the person is paying attention to the content displayed on the screen. This measure has been extensively used in the eye-tracking research (Jacob and Karn, 2003; Just and Carpenter, 1976). However, it does not provide a spatial measure of concentration; the fixation duration is computed for a specific point on the screen. For having a spatial measure, we follow Sharma et al. (2013) where SE is used to define concentration/attention over a bigger area on the screen. Recall that X_1 and X_2 denote the spatial entropy of subject 1 and 2 respectively as represented in Figure 1.

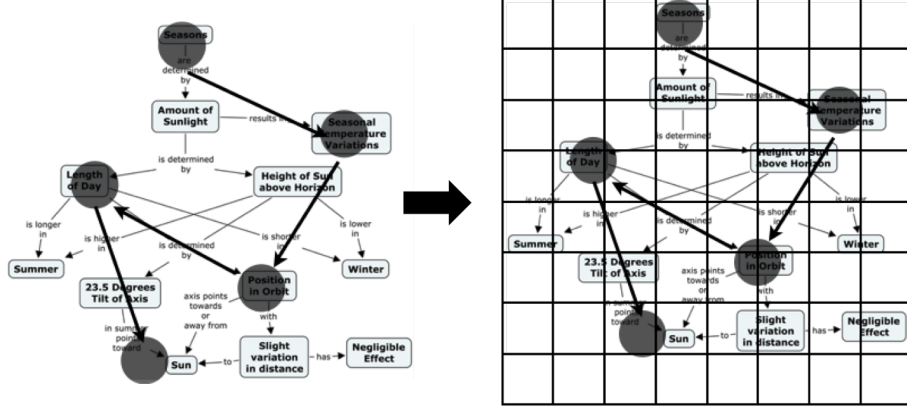


Figure 4: A typical example of a concept map (left) with the hypothetical gaze pattern overlaid. The circles denote the fixations and the arrows depict the saccades. On the right hand side, we show how we divide the whole screen into grids to compute the proportionality vector. The Figure is borrowed from Sharma et al. (2017).

More precisely, we observe two times series of entropy $x_{k,t}$ for $k = 1, 2$ measured at 10 second intervals as follows. We define a grid of $n=100$ -pixel by $m=100$ -pixel over the screen. An exmple of a grid is shown in the right panel of Figure 4. For each time t , denote by $p_{ij}^{t,k}$ the proportion of time spent by the subject on the cell (i, j) during the time window between $t - 1$ and t . The time series of the Shannon entropy of subject k at time t is

$$x_{k,t} = - \sum_{i=1}^n \sum_{j=1}^m p_{ij}^{t,k} \log(p_{ij}^{t,k}), \quad k = 1, 2,$$

Let us further define the gaze characteristics which will form \mathbf{W} , the set of covariates,

- Activity is the total number of actions done by the pair between $t - 1$ and t

$$\text{Activity}_t = \sum_{k=1}^2 \sum_{i=1}^n \sum_{j=1}^m a_{ij}^{t,k},$$

where $a_{ij}^{t,k}$ is the number of actions done by subject k at time t on cell (i, j) .

- Similarity denotes the pair similarity measured at time t

$$\text{Similarity}_t = \frac{\sum_{i=1}^n \sum_{j=1}^m p_{ij}^{t,1} p_{ij}^{t,2}}{\sqrt{\sum_{i=1}^n \sum_{j=1}^m (p_{ij}^{t,1})^2} \sqrt{\sum_{i=1}^n \sum_{j=1}^m (p_{ij}^{t,2})^2}} = \cos \left\{ \angle(p^{t,1}, p^{t,2}) \right\},$$

where $p^{t,k}$ denotes the $n \times m$ -vector composed of the $n \times m$ components $p_{ij}^{t,k}$. The vectors $p^{t,1}$ and $p^{t,2}$ are not collinear so they span a two-dimensional plane on

which the angle $\angle(p^{t,1}, p^{t,2})$ is defined. Having a high value of similarity means that the two subjects are often looking at similar objects at the same time.

- Stability is a factor with two levels: H for high stability and L for low stability. The stability factor uses individual stability measured at the individual level at time t

$$s^{t,k} = \frac{\sum_{i=1}^n \sum_{j=1}^m p_{ij}^{t,k} p_{ij}^{t-1,k}}{\sqrt{\sum_{i=1}^n \sum_{j=1}^m (p_{ij}^{t,k})^2} \sqrt{\sum_{i=1}^n \sum_{j=1}^m (p_{ij}^{t-1,k})^2}} = \cos \left\{ \angle(p^{t,k}, p^{t-1,k}) \right\}, \quad k = 1, 2.$$

We chose two consecutive times and not more because the time interval between two measures already represents 10 seconds. More consecutive times would lead to negligible values for stability. Then

$$\text{Stability}_t = \begin{cases} \text{H, if } s^{t,1} + s^{t,2} > \text{median}(s^{t,1} + s^{t,2}) \\ \text{L, otherwise} \end{cases}$$

High stability means that each participant was looking at the same set of objects on the screen for the two consecutive time windows.

- The covariate time $\in [0, T]$ where T is the length (in seconds) of the task, allow for non-stationarity.

In Sharma et al. (2017), a comparison of the effectiveness of extreme value theory (EVT)–based modeling over traditional methods (ANOVA, linear regression, correlation tests) to classify the collaboration quality of the pairs is provided. The competitive performance of EVT approaches in this data context is explained by the fact that the tail is more informative than the body distribution, as shown in Figure 1. The lower tail actually represents the episodes during which the subjects are most likely to be together focused in a high level of collaborative quality whereas the average joint values are less informative, probably containing other effects than collaboration. The fact that the average joint values are less informative to explain the quality of collaboration is confirmed by a machine learning based approach. We use a recurrent neural network setup with the seven variables: spatial entropy for subject 1 (X_1); stability for subject 1 (S_1); number of actions performed by subject 1 (A_1); spatial entropy for subject 2 (X_2); stability for subject 2 (S_2); number of actions performed by subject 2 (A_2); similarity between the two subjects S . Table 1 contains the accuracy for predicting the quality of collaboration using different settings for the learning rate and the number of hidden layers. The low predicting accuracy compared to the 75.8% provided by a simple classification based on copula selection from AIC confirms what we formulated in Section 1 on the basis of Figure 1. The average values do not bring information and a different point of view to study the gaze data is required when it comes to analyzing the coinciding concentrations. EVT offers the (correct) pair of glasses to look at the data here. We extend the work of Sharma et al. (2017) in a non-stationary context of

Table 1: Recurrent neural network at different learning rates, method based on copula selection (last row) and prediction accuracy (last column).

| Learning Rate | Hidden Layers with 10 neurons each | Accuracy (%) |
|----------------------------------|---------------------------------------|--------------|
| 0.01 | 1 | 40 |
| 0.01 | 2 | 52.5 |
| 0.01 | 3 | 50 |
| 0.1 | 1 | 47.5 |
| 0.1 | 2 | 45 |
| 0.1 | 3 | 40 |
| 1.0 | 1 | 46.6 |
| 1.0 | 2 | 42.5 |
| 1.0 | 3 | 40 |
| Method based on copula selection | | 75.8 |

bivariate tail dependence. We use our methodology described in Section 2.2 to extract information from the joint lower tail of the SE series of high collaboration quality pairs. More precisely, we are interested in modeling the coinciding lower extremes of subjects' SE in terms of the significant covariates of \mathbf{W} .

As discussed before the pairs with low collaboration quality show no dependence structure. This observation itself is a result meaning that these pairs do not coordinate themselves. Therefore our exploration focuses only on pairs with a high quality of collaboration. Before attempting to explain the tail dependence coefficient (1) of the subjects' SE with the characteristic given by the set of covariates \mathbf{W} , we first need to remove the possible effect of the covariates on X_1 and X_2 individually so that the effect of the covariates on the lower tail dependence between X_1 and X_2 does not originate from their effect on the margins. To do so, we fit a GAM on the $N = 10712$ data $x_{1,t}$ and $x_{2,t}$ separately, leading to the two significant models

$$\hat{\mu}_{X_1} = \hat{\alpha}_{X_1} + \hat{h}_{X_1}(\text{stability, edf} = 4.46) + \hat{h}_{X_1}(\text{time, edf}=8.57) \quad (14)$$

$$\hat{\mu}_{X_2} = \hat{\alpha}_{X_2} + \hat{h}_{X_2}(\text{stability, edf} = 4.64) + \hat{h}_{X_2}(\text{time, edf} = 8.48), \quad (15)$$

where μ_{X_k} refers to the mean of X_k and the functions h_{X_k} are smoothed non-parametric functions. The selected equivalent degrees of freedom (edf) are rather similar for the two subjects which is not surprising, the subjects being interchangeable. Having removed the covariates' effect on the marginals, we apply the new methodology to the GAM's residuals.

We first choose a copula family for the residuals shown in the scatterplot of Figure 5. Given the set of $N = 10712$ pairs of residuals (r_{i1}, r_{i2}) ($i = 1, \dots, N$) coming from the

| | Gaussian | t | Gumbel | Clayton |
|-----|----------|----------|----------|-----------------|
| AIC | -6632.52 | -6662.37 | -3387.25 | -13425.5 |

Table 2: AIC for different copula families for high collaboration quality pairs.

| | Estimate | s.e. | p-value |
|--------------------------|----------------|---------------|--------------|
| $\hat{\alpha}_0$ | $1.348e + 00$ | $7.852e - 03$ | $< 2e - 16$ |
| $\hat{\beta}_H$ | $2.987e + 03$ | $4.431e + 02$ | $1.66e - 11$ |
| $\hat{\beta}_L$ | $1.844e + 03$ | $3.012e + 02$ | $9.54e - 10$ |
| $\hat{\beta}_{Activity}$ | $-1.790e - 03$ | $3.003e - 03$ | 0.551 |
| | edf | | p-value |
| \hat{h} | 8.826 | | $< 2e - 16$ |

Table 3: Results for the high quality of collaboration model.

16 pairs, we fit several copulas on (u_{i1}, u_{i2}) , where $\{u_{i1} = \hat{F}_1(r_{i1}), u_{i2} = \hat{F}_2(r_{i2})\}$ form the pseudo sample obtained by estimating the margins F_1 and F_2 by their empirical distributions. Table 2 shows the AICs for the different fitted copulas. Not surprisingly, the Clayton copula minimises the AIC for the high quality collaboration pairs.

Having chosen the Clayton copula, we fit and compare different models including different sets of covariates of \mathbf{W} . The last column of Table 3 provides the p-value for each covariate. This shows that the covariate activity is not significant to explain the lower tail dependence λ_ℓ . The resulting significant conditional model is

$$\hat{\lambda}_\ell = g(\hat{\alpha}_0 + \hat{h}(\text{time}, \text{edf} = 8.826) + \mathbb{1}_{\{\text{stability}=\text{“H”}\}} \left\{ \hat{\beta}_H \text{similarity} \right\} + \mathbb{1}_{\{\text{stability}=\text{“L”}\}} \left\{ \hat{\beta}_L \text{similarity} \right\}), \quad (16)$$

where $g(x) = \nu\{exp(x)\}$, with ν as defined in (9), and the estimated values are provided in Table 3. The non-parametric function h for explaining time variation is estimated from a smoothing spline with around 9 degrees of freedom, highlighting the non-stationarity of the dependence coefficient. The effects of stability, similarity, and time are shown in Figure 6. The left panel shows the lower tail dependence fitted values as a function of similarity. The non-linearity of the curves is due to the link function and the effect of time. The lower tail dependence increases with stability. The right panel shows the time-varying curve of the lower tail dependence estimate for two fixed values of similarity. For the highest value of similarity, the effect of time is smoothed by the biggest effect of similarity on the estimated value. The time varying curve does not bring relevant information for a predictive purpose but its exploratory feature is useful to compare pairs, for instance.

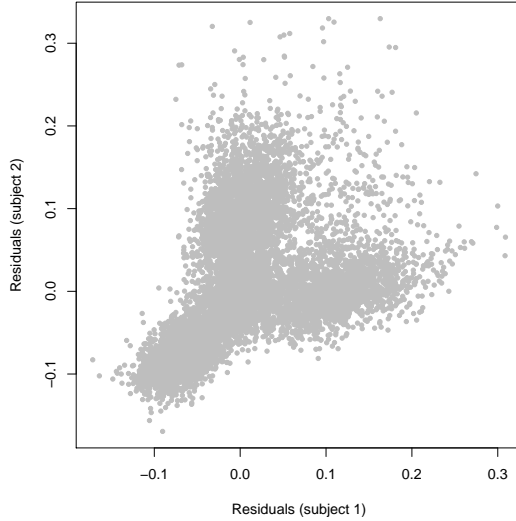


Figure 5: Scatterplot of the residuals of the individual model for subject 1' SE and of the individual model for subject 2' SE.

In terms of behavior, intuitively, we can explain the relation among entropy, stability, and similarity as follows. If the pair has a high similarity and both the subjects are highly focused (low entropy) over a long period of time, they both have high values of stability. However, the inverse of this might or might not be true. Our analysis brings forward the key difference between the high and the low quality pairs. The key difference is that for the high quality pairs the inverse of the intuition is also true. We observe that the lower tail dependence increases with pairs' similarity values and it is significantly higher in the episodes with high stability (Table 3 and Figure 6). This, however, cannot be said for the pairs with low collaboration quality.

5 Conclusion

With a huge amount of eye-tracking data available these days, the model discussed here is a step toward modeling collaborative learning. It is based on the most recent developments of dependence modeling beyond correlation (through copulas), as well as on the semi-parametric conditional model of Vatter and Chavez-Demoulin (2015). The methodology conveniently uses the information provided from the entire data to fit a covariate-dependent tail coefficient (which is standardly calculated from extreme values only). Many other explanatory variables of any type can be incorporated, because the framework benefits directly from the flexibility of generalized additive models. In that sense, the proposed method opens the door to other explorations in learning

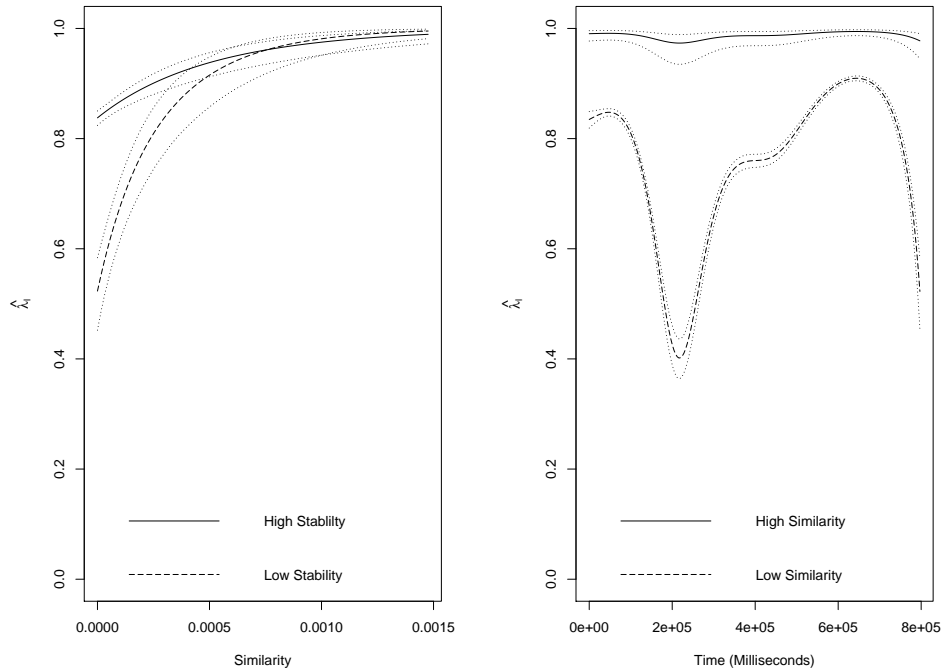


Figure 6: High collaboration quality model: fitted values of the lower tail dependence coefficient as a function of similarity (left) for the two levels of stability and different times and as a function of time (right) for two different levels of similarity.

education. An improvement might involve considering the multivariate dimension $d > 2$, for instance groups of more than two subjects working together to accomplish shared goals. That can be done using pair-copula constructions as in Aas et al. (2009) although the mapping with a high dimensional tail dependence might not always exist. An alternative extension for $d > 2$, and following the relevance of the extremal coefficient in the bivariate context in Sharma et al. (2017), would be the non-stationary extremal coefficient that Smith (1990), Coles and Tawn (1996) defined for max-stable processes. This promising approach is currently under investigation.

References

- K. Aas, C. Czado, A. Frigessi, and H. Bakken. Pair-copula constructions of multiple dependence. *InsuranceMath.Econom.*, 44(2):182–198, 2009.
- P. D. Allopenna, J. S. Magnuson, and M. K. Tanenhaus. Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of memory and language*, 38(4):419–439, 1998.
- D. H. Ballard, M. M. Hayhoe, F. Li, S. D. Whitehead, J. P. Frisby, J. G. Taylor, and

- R. B. Fisher. Hand-eye coordination during sequential tasks [and discussion]. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 337(1281):331–339, 1992.
- W. G. Chase and H. A. Simon. Perception in chess. *Cognitive psychology*, 4(1):55–81, 1973.
- S.G. Coles and J.A. Tawn. Modelling extremes of the areal rainfall process. *Applied Statistics*, 58:329–347, 1996.
- P. Embrechts, F. Lindskog, and A. McNeil. Modelling dependence with copulas and applications to risk management. In S. Rachev, editor, *Handbook of Heavy Tailed Distributions in Finance*, pages 329–384. Elsevier, 2003.
- M. Ferreira. Nonparametric estimation of the tail-dependence coefficient. *REVSTAT - Statistical Journal*, 11:1–16, 2013.
- L. Gardes and S. Girard. Nonparametric estimation of the conditional tail copula. *Journal of Multivariate Analysis*, 137:1–16, 2015.
- E. R. Grant and M. J. Spivey. Eye movements and problem solving guiding attention guides thought. *Psychological Science*, 14(5):462–466, 2003.
- P. J. Green. Penalized likelihood for general semi-parametric regression models. *International Statistical Review*, 55 (3), 1987.
- P.J. Green and B.W. Silverman. *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis, 1994.
- T. J. Hastie and R. J. Tibshirani. *Generalized Additive Models*. London: Chapman and Hall, 1990.
- C. E. Hmelo-Silver. Analyzing collaborative learning: Multiple approaches to understanding processes and outcomes. In *Proceedings of the 7th international conference on Learning sciences*, pages 1059–1065. International Society of the Learning Sciences, 2006.
- R. J. Jacob and K. S. Karn. Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. *Mind*, 2(3):4, 2003.
- H. Joe. *Multivariate Models and Dependence Concepts*. Chapman and Hall, London, 1997.

- G. Jones. Testing two cognitive theories of insight. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(5):1017, 2003.
- M. A. Just and P. A. Carpenter. Eye fixations and cognitive processes. *Cognitive psychology*, 8(4):441–480, 1976.
- G. Knoblich, S. Ohlsson, and G. E. Raney. An eye movement study of insight problem solving. *Memory & Cognition*, 29(7):1000–1009, 2001.
- F. Li. Modeling covariate-contingent correlation and tail-dependence with copulas. <https://arxiv.org/abs/1401.0100>, 2016.
- A. J. McNeil, R. Frey, and P. Embrechts. *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press, Princeton, New Jersey, 2005.
- Nelsen. *An Introduction to Copulas*. Springer, 1999.
- M.-A. Nüssli. Dual eye-tracking methods for the study of remote collaborative problem solving. 2011.
- S. Pietinen, R. Bednarik, and M. Tukiainen. Shared visual attention in collaborative programming: a descriptive analysis. In *proceedings of the 2010 ICSE workshop on cooperative and human aspects of software engineering*, pages 21–24. ACM, 2010.
- D. C. Richardson, R. Dale, and N. Z. Kirkham. The art of conversation is coordination common ground and the coupling of eye movements during dialogue. *Psychological science*, 18(5):407–413, 2007.
- M. Sangin, G. Molinari, M.-A. Nüssli, and P. Dillenbourg. How learners use awareness cues about their peer’s knowledge?: insights from synchronized eye-tracking data. In *Proceedings of the 8th international conference on International conference for the learning sciences-Volume 2*, pages 287–294. International Society of the Learning Sciences, 2008.
- R. Schmidt and U. Stadtmueller. Non-parametric estimation of tail dependence. *Scandinavian Journal of Statistics*, 33(2):307–335, 2006.
- B. Schneider, K. Sharma, S. Cuendet, G. Zufferey, P. Dillenbourg, and R. D. Pea. 3d tangibles facilitate joint visual attention in dyads. In *Proceedings of 11th International Conference of Computer Supported Collaborative Learning*, volume 1, pages 156–165, 2015.
- K. Sharma, P. Jermann, M.-A. Nüssli, and P. Dillenbourg. Gaze evidence for different activities in program understanding. In *24th Annual conference of Psychology of Programming Interest Group*, number EPFL-CONF-184006, 2012.

- K. Sharma, P. Jermann, M.-A. Nüssli, and P. Dillenbourg. Understanding collaborative program comprehension: Interlacing gaze and dialogues. In *Computer Supported Collaborative Learning (CSCL 2013)*, 2013.
- K. Sharma, V. Chavez-Demoulin, and P. Dillenbourg. An application of extreme value theory to learning analytics: Predicting collaboration quality from eye-tracking data. *Journal of Learning*, 2017. To appear.
- M. Sibuya. Bivariate extreme statistics. *Annals of the Institute of Statistical Mathematics*, 11:195–210, 1960.
- M Sklar. *Fonctions de répartition à n dimensions et leurs marges*. Université Paris 8, 1959.
- R.L. Smith. Max-stable processes and spatial extremes. *Unpublished Manuscript*, 1990.
- T. Vatter and V. Chavez-Demoulin. Generalized additive models for conditional dependence structures. *Journal of Multivariate Analysis*, 141:147–167, 2015.
- S. Wood. *Generalized Additive Models: An Introduction with R*. Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis, 2006.
- G. J. Zelinsky and G. L. Murphy. Synchronizing visual and language processing: An effect of object name length on eye movements. *Psychological Science*, 11(2):125–131, 2000.

A Definitions and details

Learning Analytics: “*Learning analytics is the use of intelligent data, learner-produced data, and analysis models to discover information and social connections, and to predict and advise on learning.*”- George Siemens (Source: “What are learning analytics?”).

Eye-tracking and Computer Supported Collaborative Learning: From Wikipedia, “*Computer-supported collaborative learning (CSCL) is a pedagogical approach where in learning takes place via social interaction using a computer or through the Internet. This kind of learning is characterized by the sharing and construction of knowledge among participants using technology as their primary means of communication or as a common resource.*”

Hmelo-Silver (2006) emphasizes the need to analyse the collaborative processes from multiple perspectives. Moreover, Hmelo-Silver (2006) addresses the needs of multiple methodological tools to be used to understand the underlying cognitive processes (how learning occurs, learning outcomes). These different tools require that multiple data

sources be used, for example dialogs (to analyse conflicts, leadership, perception of technology and peers), eye-tracking (joint attention, referential mechanisms), online actions (division of labour).

In this contribution we have focused on analysing the interaction between collaborating learners based on their eye-tracking data. Eye-tracking provides researchers with an unprecedented access to the users' attention. The eye-tracking data is rich in terms of the temporal resolution. Eye-tracking had previously been used in the differentiate problem solving skills and/or levels of expertise (Knoblich et al., 2001; Jones, 2003; Grant and Spivey, 2003) to explain cognition underlying problem solving (Ballard et al., 1992; Chase and Simon, 1973), to understand how the gaze is related to dialogues (Allopenna et al., 1998; Zelinsky and Murphy, 2000; Richardson et al., 2007), and to explain cognition underlying collaborative learning (Pietinen et al., 2010; Nüssli, 2011; Sangin et al., 2008; Sharma et al., 2013, 2012; Schneider et al., 2015).

Concept Map: A concept map is a graphical representation (organisation) of a given body of knowledge, where the nodes are the key concepts, which constitute the knowledge and the edges are the conceptual link between the concepts. Concept maps are often used in education as a method to summarise, retain, and communicate knowledge and are also used for collaborative knowledge modelling.