

# Les traces et données d'apprentissage : comment et pourquoi les exploiter ?





# De quoi allons-nous parler...

- 💧 Objectif de l'analyse de traces en EIAH
- 💧 Ce que l'on peut trouver dans les traces d'apprentissage
- 💧 Deux communautés de recherche : EDM et LAK
- 💧 Exemples de travaux de recherche
- 💧 Ce qu'il reste à faire...
- 💧 Pour aller plus loin : références, conférences et revues



# Objectif de l'analyse de traces en EIAH

- ◆ Suivi de l'apprenant et de son parcours d'apprentissage
- ◆ Tutorat des apprenants lors des séances d'apprentissage, formation de groupes
- ◆ Prédiction des succès et/ou échecs pour offrir des interventions proactives
- ◆ Evaluation sommative, efficacité de l'apprentissage
- ◆ Adaptation / personnalisation automatique du système
- ◆ Recommandation pour donner le contrôle à l'apprenant
- ◆ Réflexivité pour faire prendre conscience à l'apprenant de son apprentissage et de ses méthodes de travail

➤ Comprendre la manière dont les humains apprennent.

➤ Comprendre la manière dont chaque apprenant singulier apprend.



# Plan du cours

- ◆ Objectif de l'analyse de traces en EIAH
- ◆ Ce que l'on peut trouver dans les traces d'apprentissage
- ◆ Deux communautés de recherche : EDM et LAK
- ◆ Exemples de travaux de recherche
- ◆ Ce qu'il reste à faire...
- ◆ Pour aller plus loin : références, conférences et revues



# Ce que l'on peut trouver dans les traces d'apprentissage

- ◆ Données qualitatives des apprenants
  - ◆ Réponses aux formulaires, aux tests...
  - ◆ Scores obtenus
  - ◆ ...
- ◆ Données collectées à partir des actions des apprenants ou des enseignants
  - ◆ Actions explicites
    - ◆ Connexion à la plateforme
    - ◆ Activités faites
    - ◆ Demande d'aide
    - ◆ Participation au forum (lecture, posts, message..)
    - ◆ Affectation d'une ressource à un apprenant
    - ◆ ...
  - ◆ Actions implicites
    - ◆ Temps de consultation d'une ressource
    - ◆ Temps d'inactivité
    - ◆ Les gestes



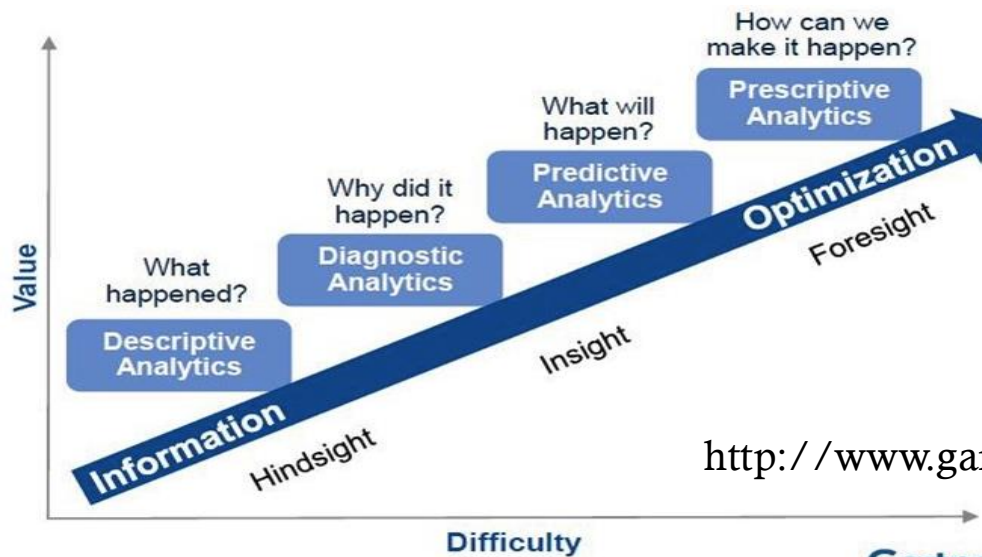
# Quels supports ?

- ◆ Ordinateurs
- ◆ Tablettes
- ◆ Liseuses
- ◆ Smartphones
- ◆ Tableaux numériques
- ◆ ...
- ◆ Capteurs
- ◆ Objets connectés (montres...)



# Quelles temporalités?

	Passé	Présent	Futur
Ce que l'on regarde	Que s'est-il passé ? (Reporting)	Que se passe-t-il actuellement ? (Alertes)	Qu'est-ce qui va se passer? (Extrapolation)
Avec quel objectif	Comment et pourquoi est-ce arrivé ? (Modélisation, conception expérimentale)	Quelle est la meilleure prochaine action ? (Recommandation)	Qu'est-ce qui peut se passer de pire/mieux ? (Prédiction, optimisation, simulation)



<http://www.gartner.com/>



# Plan du cours

- ◆ Objectif de l'analyse de traces en EIAH
- ◆ Ce que l'on peut trouver dans les traces d'apprentissage
- ◆ Deux communautés de recherche : EDM et LAK
- ◆ Exemples de travaux de recherches
- ◆ Ce qu'il reste à faire...
- ◆ Pour aller plus loin : références, conférences et revues



# Historique des deux communautés de recherche : EDM et LAK

- ◆ Emmergence des sociétés savantes
  - ◆ 2007 : IEDMS - International Educational Data Mining Society
  - ◆ 2011 : SoLAR – Society for Learning Analytics Research
- ◆ Pourtant
  - ◆ Données d'apprentissage présentes et utilisées depuis les ITS...
  - ◆ Utilisées par des équipes pluridisciplinaires
    - Pour modéliser *a priori* l'apprenant, faire des analyses exploratoires des données...
    - Groupe Intelligence & Didactique (Balacheff 1994)
    - Conférences AIED et EIAO/EIAH (depuis 1989) ...
- Ces travaux ont posé les fondements des problématiques et des méthodes



# Alors pourquoi cette emmergence ?

- ◆ 4 facteurs (Baker & Siemens, 2014)
  - ◆ **Volume de données** en forte et rapide progression
    - ◆ Archives publiques comme DataShop (<https://pslcdatashop.web.cmu.edu/>)
    - ◆ Multiplication des LMS (Learning Management System), MOOCs...
  - ◆ **Données structurées**
    - ◆ Travaux sur l'interopérabilité
  - ◆ **Capacités de calcul**
  - ◆ **Nouveaux outils d'analyse sans programmation**
    - ◆ Algorithmes d'analyse inclus
    - ◆ Rapid Miner (<https://rapidminer.com/>),  
R (<https://www.r-project.org/>),  
KEEL (<http://www.keel.es/>),  
Weka (<http://www.cs.waikato.ac.nz/ml/weka/>),  
SNAPP (<http://www.snapsurveys.com>)



# Changement des environnements et des problématiques



## ◆ Dans les ITS

- ◆ Données fines, capteurs physiologiques
- ◆ Destinées à évaluer certains aspects émotionnels, comportementaux ou cognitifs
- ◆ Structuration *a priori*

## ◆ Dans les LMS et les MOOCs

- ◆ Données moins fines
- ◆ Mais massives (augmentation exponentielle)
  - ◆ Par le nombre d'utilisateurs
  - ◆ Par l'intensité des usages d'un utilisateur : monde numérique dans le monde de tous les jours avec géolocalisation, indicateurs biométriques...
- ◆ Structuration *a posteriori*



# EDM : Educational Data Mining

## ◆ Data Mining

- ◆ En recherche fondamentale, application d'algorithmes, issus de la statistique ou de l'IA, pour l'extraction de l'information (utile et inconnue) de gros volume de données non-structurées
- ◆ Etranger aux théories sur le comportement humain comme la linguistique, la psychologie ou la sociologie
- La structure des données serait consubstantielle aux données elles-mêmes



# EDM : Educational Data Mining

## ◆ Historique

- ◆ 2000 : introduction du nom EDM lors de la conf. ITS
- ◆ 2005 : 1<sup>er</sup> Workshop EDM lors de la conf. AAAI  
(Association for the Advancement of Artificial Intelligence)
  - ◆ Travaux sur les traces des ITS
- ◆ 2007 : état de l'art complet (Roméro & Ventura)
- ◆ 2007 : IEDMS - International Educational Data Mining Society
- ◆ 2008 : conf. internationale EDM, annuelle
- ◆ 2009 : Journal of Educational Data Mining



# EDM : Educational Data Mining

- ◆ Définition (Baker & al. 2009)
  - ◆ Un domaine de recherche impliquant :
    1. Le développement de méthodes dédiées à l'exploration des spécificités des données d'apprentissage
    2. Le déploiement de ces méthodes pour une meilleure compréhension des apprenants et de l'environnement d'apprentissage





# LAK : Learning Analytics & Knowledge

- Historique
  - 2011 : SoLAR – Society for Learning Analytics Research
  - 2014 : Journal of Analytics
- Définition (Siemens, 2011)
  - Un domaine de recherche pour :
    - L'évaluation, l'analyse, la collecte et la communication
    - des données relatives aux apprenants, leur contexte d'apprentissage
    - dans la perspective d'une compréhension et d'une optimisation de l'apprentissage et de son environnement





# Quelle est la différence ?

## 💧 EDM

- 💧 Priorise le développement de nouveaux outils et algorithmes pour révéler les modèles enfouis dans les données d'apprentissage.
- 💧 Objectif : **donner au logiciel** la capacité de prédire les résultats d'un apprenant et de personnaliser sa stratégie d'apprentissage

## 💧 LAK

- 💧 Modélisation et visualisation des données transmises aux acteurs de l'apprentissage
  - 💧 Apprenants, parents, personnels d'éducation, enseignants, gestionnaires, administrateurs...
- 💧 Objectif : amplifier le **rôle décisionnel des acteurs** de l'apprentissage en leur donnant les moyens d'explorer les données d'apprentissage, sans les noyer, pour améliorer l'expérience d'apprentissage



# Caractéristiques d'EDM

- ◆ Méthode privilégiée
  - ◆ Fouille de données automatique
- ◆ Approche
  - ◆ Réduire les systèmes d'apprentissage à ses composantes principales
  - ◆ Modélisation séparée des apprenants, des tuteurs, du domaine enseigné...
  - ◆ Repose sur les théories du DM
- ◆ Objectif
  - ◆ Concevoir des systèmes automatiques
  - ◆ Avec des objectifs de suivi, d'adaptation et de personnalisation



# Caractéristiques de LAK

- ◆ Méthode privilégiée
  - ◆ Outil d'aide à la décision : découverte des données par les acteurs *via* des outils d'analyse et de visualisation
- ◆ Approche
  - ◆ Systémique : appréhender la situation pédagogique comme un ensemble
  - ◆ Ne repose pas sur des modèles SHS
- ◆ Objectif
  - ◆ Donner de l'autonomie aux acteurs (enseignants, apprenants...)



# Synthèse

- ◆ Deux approches souvent confondues
- ◆ Deux approches différentes
  - ◆ EDM : méthode automatique et résultat pour le système
  - ◆ LAK : méthode (semi-)manuelle et résultat pour l'humain
- Mais deux approches complémentaires avec des objectifs communs !!!



# Objectif commun : prédiction entre variables

- ◆ Objectif du traitement

- ◆ Développer un modèle permettant de prédire une variable (dépendante) à partir de la combinaison d'autres (indépendantes)

- ◆ Méthodes

- ◆ Classification : arbre de décision, règles de décision, régression séquentielle ou logistique

- ◆ Exemple

- ◆ Algo. d'estimation des connaissances dans les ITS



# Objectif commun : découverte de structure

## ◆ Objectif du traitement

- ◆ Etablir la structure des données sans idées préconçue sur l'objet recherché, sans intérêt *a priori* pour aucune variable

## ◆ Méthodes

- ◆ Clustering ou partitionnement : trouver les données qui se regroupent entre elles, en séparant le jeu en plusieurs sous-ensemble
- ◆ Analyse factorielle pour trouver des sous-ensembles homogènes
- ◆ Analyse de réseaux sociaux : analyse à partir des relations et interactions entre individus

## ◆ Exemple

- ◆ Grouper des étudiants (Beal et al. 2006)
- ◆ Evaluer l'engagement d'un apprenant dans sa scolarité (MacFadyen et al. 2010)
- ◆ Identifier les choix de design généralement faits par les concepteurs d'ITS (Baker et al. 2009)



# Objectif commun : fouille de relations

- ◆ Objectif du traitement
  - ◆ Découvrir les relations entre variables au sein d'un jeu de données
- ◆ Méthodes
  - ◆ Fouille de règles d'association
  - ◆ Fouille de corrélations linéaires positives ou négatives entre variables
- ◆ Exemple
  - ◆ Trouver des patterns d'étudiants obtenant de bonnes performances pour pouvoir faire de meilleures suggestions aux étudiants qui éprouvent des difficultés (Cen et al. 2006)



# Objectif commun : traitement de données pour l'évaluation humaine

- ◆ Objectif du traitement
  - ◆ Représenter les données d'apprentissage de façon efficace pour agir sur la pédagogie
- ◆ Méthodes
  - ◆ Méthodes de visualisation et d'aide à la décision
    - ◆ Nuages de points, courbes et diagrammes
- ◆ Exemple
  - ◆ Visualisation de trajectoire des étudiants pour identifier les étudiants à risque (Bowers 2010)
  - ◆ Evaluation de la motivation (HersHKovitz et al. 2008)




# Plan du cours

- ◆ Objectif de l'analyse de traces en EIAH
- ◆ Ce que l'on peut trouver dans les traces d'apprentissage
- ◆ Deux communautés de recherche : EDM et LAK
- ◆ Exemples de travaux de recherche
- ◆ Ce qu'il reste à faire...
- ◆ Pour aller plus loin : références, conférences et revues



# Prédire la progression de l'apprenant

- ◆ Prédire le parcours comportemental ou cognitif
  - ◆ Un des plus anciens problèmes des EIAH...
- ◆ Objectifs divers :
  - ◆ Classer les apprenants
  - ◆ Les orienter en fonction des capacités estimées
  - ◆ Adapter les contenus
  - ◆ Déployer des stratégies d'engagement
  - ◆ Lutter contre le décrochage...
- ◆ Analyses menées à différents niveaux
  - ◆  Microgenèse : analyse de la simple interaction en temps réel d'un individu
  - ◆ Analyse *a posteriori* de l'apprentissage d'une cohorte sur une période donnée



# Prédire la progression de l'apprenant

## *Zimmermann et al. 2015*

### Objectif :

- Évaluer la puissance de prédiction des résultats scolaires et leur agrégation comme indicateurs de performance

### Données analysées :

- 81 variables pour une population de 171 étudiants
- Sexe, âge, notes des 3 années de licence, notes de master, nombres de crédits validés par an, temps pour obtenir une UE...

First year	Second year	Third year
<i>Basic courses (9)</i> Calculus Data Structures and Algorithms Digital Design Discrete Mathematics Introduction to Programming Linear Algebra Logic Probability and Statistics Physics	<i>Compulsory courses (11)</i> Advanced Computational Science Computer Architecture Computer Networks Electrical Engineering Formal Methods and Functional Programming Information Theory Introduction to Computational Science Introduction to Databases Operating Systems Software Architecture Systems Programming Theory of Computing	<i>Core courses (4)</i> Algorithms, Probability and Computing Distributed Systems Information Security Information Systems Modeling and Simulation Visual Computing Software Engineering  <i>Elective courses (3-4)</i> ~ 100 courses in various fields of Computer Science  <i>Compulsory major courses (2)</i> ~ 30 courses and seminars  <i>Compulsory elective courses (3)</i> ~ 100 courses in humanities, social and political sciences

Curriculum		Focus areas
<i>Focus courses (4-6)</i> ~ 10 courses in focus area	<i>Multidisciplinary courses (2-3)</i> > 1000 courses offered by 3 universities	Computational Science Distributes Systems Information Security Information Systems Software Engineering Theoretical Computer Science Visual Computing
<i>Elective courses (4-5)</i> ~ 100 courses in various fields of Comp. Sci.	<i>Compulsory elective courses (1)</i> ~ 100 courses in humanities, social and political sciences	
<i>Foundations of Comp. Sci. (4)</i> 5 courses in total	<i>Master's thesis</i>	



# Prédire la progression de l'apprenant

## *Zimmermann et al. 2015*

### ◆ Technique :

- ◆ *A posteriori*
- ◆ Modèle de régression

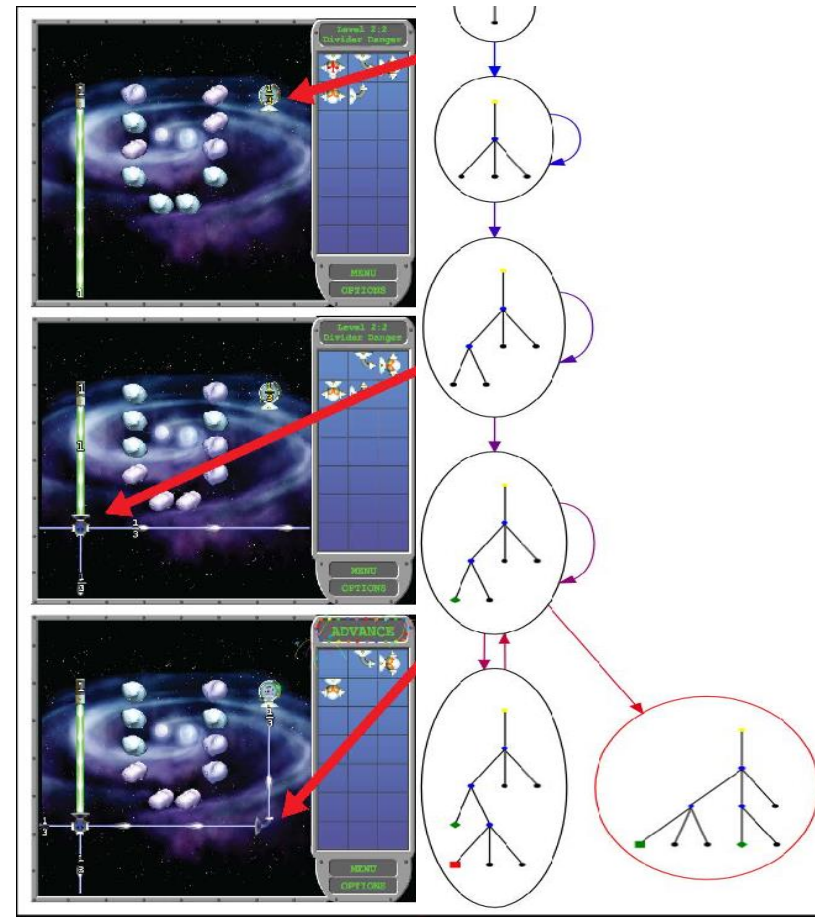
### ◆ Résultats :

- ◆ Les résultats de licence peuvent expliquer 54% de la variance dans les résultats de cycles supérieurs
- ◆ La moyenne de notation globale de la 3<sup>ème</sup> année est la variable la plus significative
- Proposition d'une méthodologie pour dresser des lignes générales pour les comités d'admission



# Prédire la progression de l'apprenant *Martin et al. 2013*

- Objectif :
  - Etablir les différents chemins dans l'apprentissage des fractions avec le jeu en ligne Refaction
- Données analysées :
  - Codage *a posteriori* des transitions entre états du jeu
- Technique :
  - Fouille de données et visualisation de graphes
- Résultat :
  - Algorithme de classification permettant de regrouper les apprenants selon des profils d'apprenants (ceux allant droit à la solution, ceux qui expérimentent...)





# Prédire la progression

## *Item Response Theory*

### *Baker, 2001*

- ◆ Constat :
  - ◆ l'évaluation des propriétés techniques d'un item (par le calcul de certains indices : de difficulté ou de discrimination) fournit des résultats qui sont toujours relatifs à l'échantillon particulier d'individus auquel l'item a été administré (plus généralement, relatifs aux conditions dans lesquelles l'opération a été réalisée: moment, contexte, etc.).
  - ◆ un item jugé facile ou difficile au sein d'un échantillon d'individus, peut ne plus l'être (ou ne plus l'être autant) s'il était appliqué à un échantillon différent
- ◆ Modèle statistique pour produire une évaluation des propriétés de l'item qui soit indépendante d'un groupe particulier d'individus
- ◆ A l'origine, elles reposaient sur un modèle (conceptuellement difficile à justifier) de nature entièrement déterministe
- ◆ Par la suite, le modèle a été remplacé par des modèles beaucoup plus réalistes, de type probabiliste.



# Prédire la progression

## *Item Response Theory*

### *Baker, 2001*

- ◆ Ces modèles sont fondés sur le postulat que la réponse d'un individu à l'item (et notamment sa probabilité de fournir une réponse correcte) est déterminée - ou peut être expliquée - par deux sortes de facteurs :
  - ◆ certains attributs du sujet (sa compétence par exemple), qui, n'étant pas directement accessibles à l'observation et à la mesure, sont généralement qualifiés de traits latents ;
  - ◆ les propriétés de l'item lui-même, notamment, sa difficulté, son pouvoir de discrimination, sans oublier le rôle que la "chance" (réponses "au hasard") peut jouer dans certains cas.
- ◆ La réponse fournie à l'item est donc considérée comme une fonction des caractéristiques de l'individu et des caractéristiques de l'item
- ◆ La théorie des réponses aux items utilise des modèles à un, deux ou trois paramètre(s), qui établissent la relation entre le trait latent de l'individu (son niveau de compétence par exemple) et la probabilité pour cet individu de réussir un item donné



# Prédire la progression

## *Item Response Theory*

### *Hambleton & Jones 2005*

1- Probabilité de répondre correctement:

$$P(\theta) = c + (1 - c) \frac{1}{1 + e^{-a(\theta - b)}}$$

2- Indépendance des items du test

$$P(S|\theta) = \prod_i P_i(v|\theta)$$

3- Le score de l'apprenant est la valeur de  $\theta$  qui maximise  $P(\theta)$

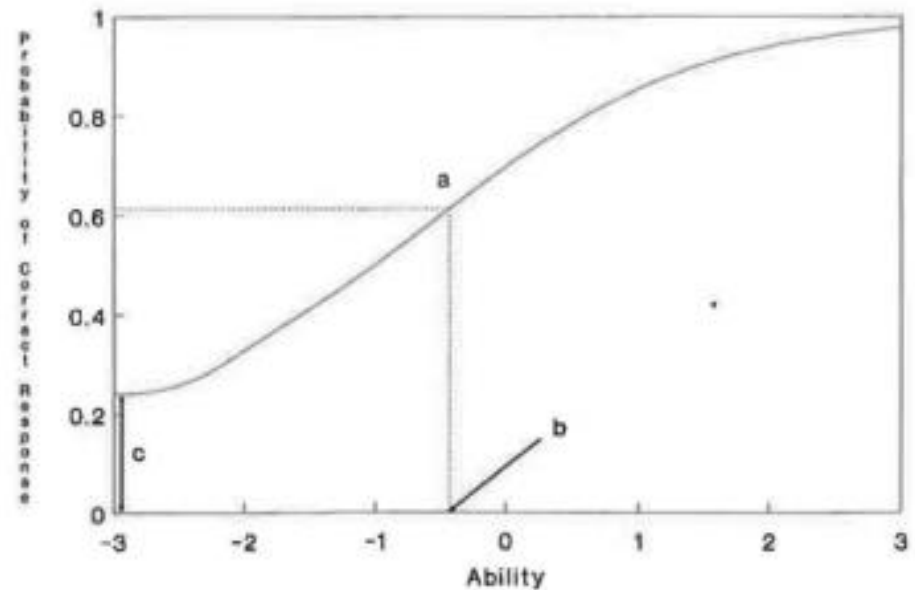


FIGURE 1. A three-parameter logistic model item characteristic curve

Courbe caractéristique d'item dans le modèle IRT à 3 paramètres. Avec  $a$  le paramètre de discrimination,  $b$  le paramètre de difficulté et  $c$  le paramètre de correction du hasard. (Source : [Hambleton and Jones, 2005])



# Prédire la progression

## *Item Response Theory*

- ◆ Classiquement utilisée pour le *Computerized Adaptive Testing* (CAT)
  - ◆ Tests adaptatifs permettent de poser une question à un apprenant, de traiter sa réponse à la volée, et de choisir la question suivante à lui poser en fonction de ses réponses précédentes.
  - ◆ Permet de réduire le nombre de questions à poser à un apprenant tout en conservant une mesure précise de son niveau.
- ◆ Multidimensional Item Response Theory (MIRT) (Reckase, 2009)
- ◆ Modèle hybride de diagnostic de connaissances adaptatif (Jill Jenn Vie, 2016)

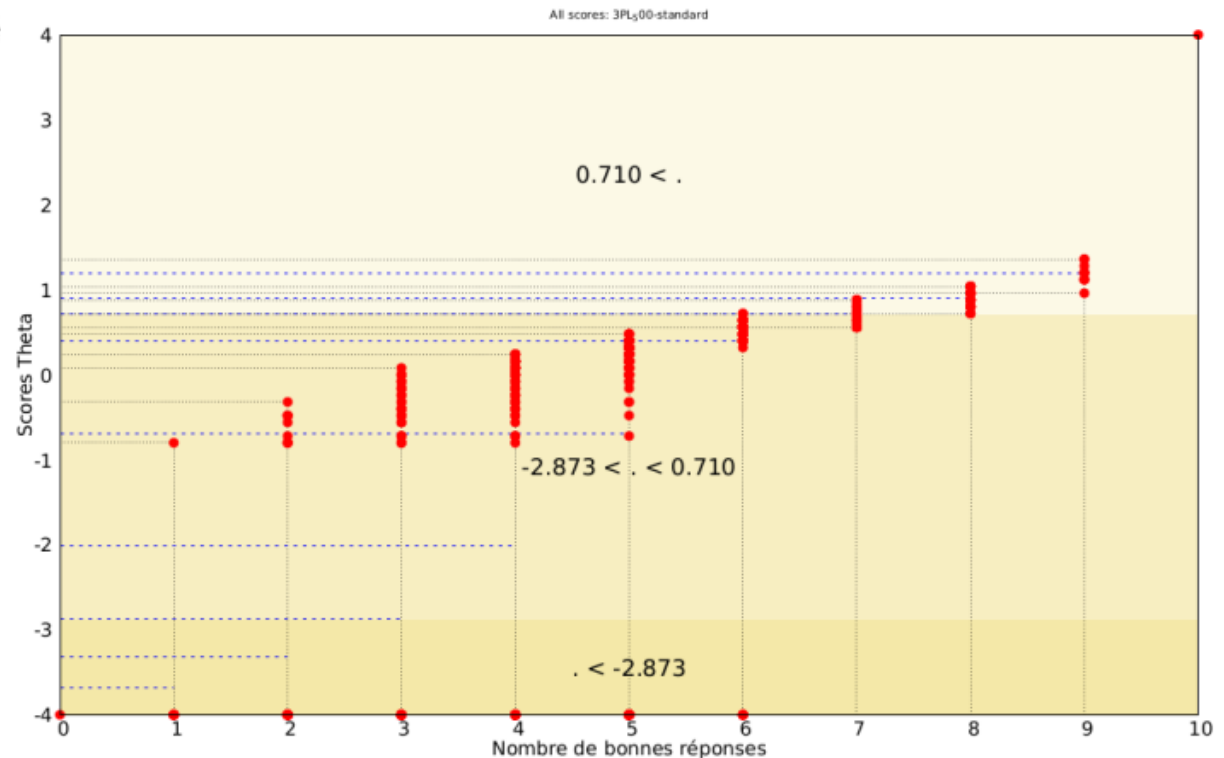


# Prédire la progression

## *Item Response Theory*

### *Brochard & Badier 2019*

- ❖ Détournée pour caractériser un niveau de maîtrise de l'apprenant sur les quiz du chapitre étudié
- ❖ Performance sur le test =  $f(\text{niveau de capacité } \theta, \text{items du test})$



Répartition des scores  $\theta$  possibles selon le nombre de réponses correctes pour le quiz Puissances 3<sup>ème</sup> et mise en évidence des seuils IRT. Chaque point rouge correspond à une valeur de  $\theta$  pour une séquence S. Les pointillés bleus correspondent à la moyenne des  $\theta$  observés pour les séquences groupées par nombre de bonnes réponses. (Source : P.Brochard, Nomad Education, 2019)



# Équité algorithmique

## *Verger et al. 2023*

- ◆ La prédiction de la réussite à des cours en ligne a connu un intérêt important depuis 15 ans
  - ◆ 357 articles publiés entre 2010 et mi-2018 sur le sujet
- ◆ Mais les modèles prédictifs utilisés en EIAH peuvent produire des résultats biaisés et discriminants !
- ◆ Pour évaluer les modèles
  - ◆ identifier au préalable les groupes de personnes potentiellement concernés
  - ◆ comparer les performances prédictives (e.g. précision, F1-score) du modèle en fonction de ces groupes distincts
- Un modèle est considéré comme ayant un comportement discriminant envers un groupe s'il ne produit pas des performances prédictives similaires à celles obtenues pour les autres groupes
- ◆ MAIS performances prédictives similaires entre différents groupes (i.e. le même nombre d'erreurs)  $\neq$  erreurs plus sévères (et dont les implications seraient plus néfastes) pour un groupe que pour un autre, en dépit d'un nombre d'erreurs identique
- Détaillé dans le cours « Équité algorithmique »



# Mesurer les interactions sociales

- ◆ Communauté CSCL, née en 1995
  - ◆ S'appuie sur les travaux de Vitgovsky et du socio-constructivisme
  - ◆ Cherche à évaluer le rôle des interactions sociales au sein des EIAH
- ◆ Prend de l'ampleur depuis 2011 avec les MOOCS
  - ◆ Notamment les cMOOC (connectivites) qui nécessitent de la collaboration



# Mesurer les interactions sociales

## *Schneider et al. 2013*

### Objectif :

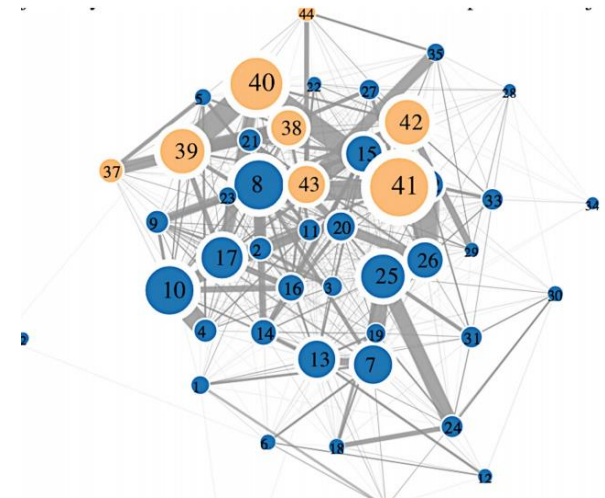
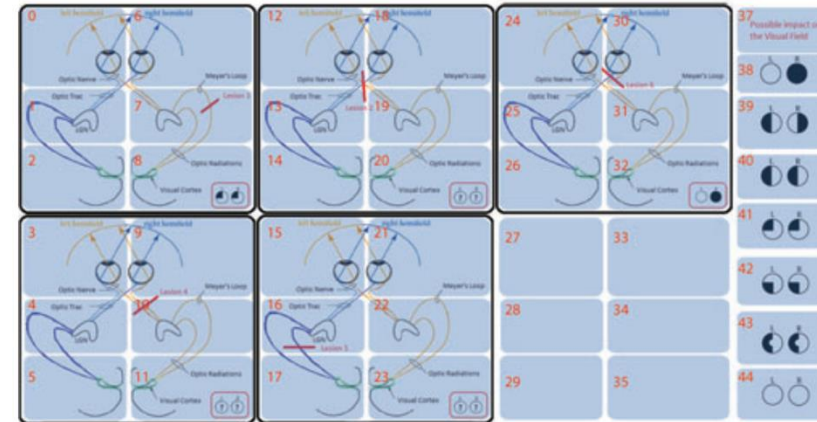
- Tester une analyse séquentielle des regards et évaluer la robustesse de cette métrique pour prédire l'apprentissage

### Données :

- 22 dyades isolées communiquent *via* un canal audio seul ou *via* un canal audio associé à un indicateur du regard du collaborateur

### Techniques :

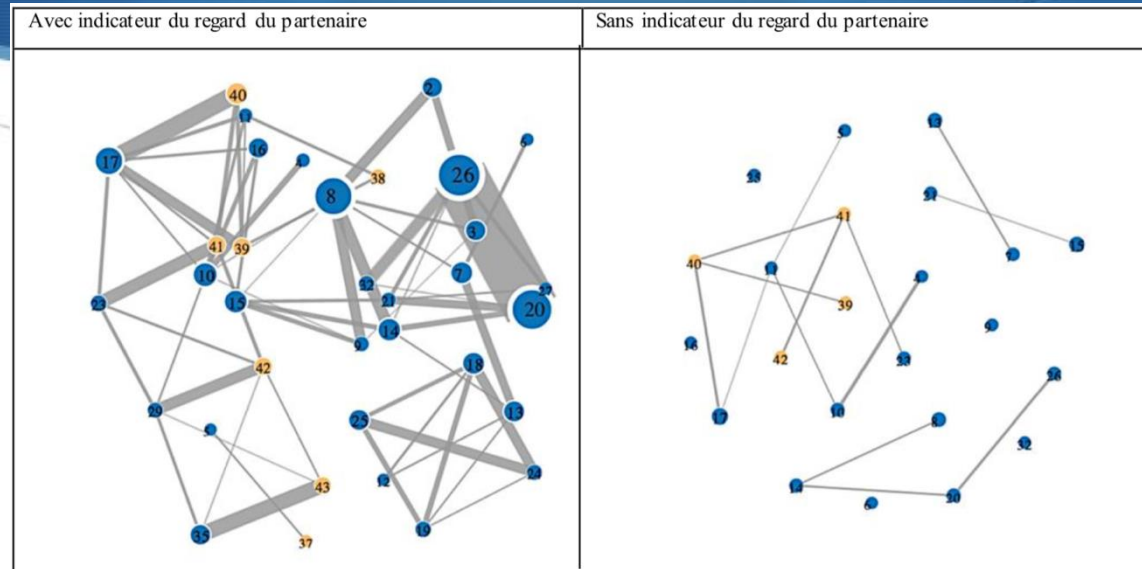
- La position du regard (noeud) et son mouvement (arc) sur un support pédagogique sont transformés en graphe non-dirigé





# Mesurer les interactions sociales

## *Schneider et al. 2013*



- ❖ Résultats :
- 💧 Démonstration de l'importance de considérer le regard comme indice d'attention soutenue mutuelle
  - 💧 En adaptant la théorie des graphes à leur problématique, identification de nouveaux indicateurs
  - Nombre de liens, taille des noeuds... permettent d'évaluer l'intensité d'une collaboration et la compréhension réciproque des apprenants



# Analyser le discours

- ◆ Apprentissage ne résulte pas des seules capacités cognitives ou du comportement d'un apprenant isolé
  - ◆ Langage : un des premiers vecteurs par lequel les apprenants construisent du sens
  - ◆ Usage influencé par les buts, les sentiments, les relations
    - Très variables selon les contextes
  - ◆ Fait partie de la situation d'apprentissage
- ◆ Déploiement des MOOCs
  - ◆ Forum : lieu de discussion asynchrone
    - Forum devient un objet de recherche



# Analyser le discours

## *De Liddo et al. 2011*

- ◆ Objectif : identification des types d'échanges
  - ◆ Challenge, exploration, évaluation, raisonnement
- ◆ Données :
  - ◆ Plateforme de délibération *Cohere*
    - ◆ Les sujets de conversation et le point de vue des apprenants
    - ◆ Les types de contribution et les accords/desaccords
    - ◆ L'organisation du réseau discursif et le rôle des apprenants dans ce réseau
- ◆ Technique :
  - ◆ Outils d'annotation : icône, lien annoté...

The screenshot shows a vertical thread of four messages on the Cohere platform. Each message is enclosed in a pink rounded rectangle. To the left of each message is a small circular icon: a lightbulb for the first, a red flag for the second, a green plus for the third, and a green plus for the fourth. To the right of the messages are green curved arrows indicating relationships: a red arrow labeled 'disagrees with' points from the second message to the first, and two green arrows labeled 'agrees with' point from the third and fourth messages to the second. The messages are as follows:

- Aurélie Miché** 22 October at 17:57 Reply • Report  
Woua, i love them all. Unfortunately, Vane i won't be able to be part of the exciting expedition for your birthday. I will be in Paris with a bunch of 30 excited students. Sorry and disappointed. Love. Lili
- Van Lopez** 26 October at 19:43 Reply • Report  
Hi guys, it seems also Dinar is traveling on the 5th .. Maca & Joan are also away, Lili too .. If you want we can postpone it for the weekend after .. although we have the KMi XMAS dinner on the 10 Friday, so it will have to be a short one Saturday-Sunday ... Your call! Besitos.
- Miriam Fernandez** 26 October at 21:05 Reply • Report  
hola chiquitos! I've got some friends visiting from saturday 11D to monday 13D so I will not be available that weekend ..
- Anna De Liddo** 26 October at 23:07  
Ciao Belli, just an info that there is no KMi christmas dinner on the 10th anymore! So in principle following w-e may work as well. BACIII
- Aurélie Miché** 27 October at 11:25 Reply • Report  
I guess we could as well and Sylvain might be there as well. His birthday on the 13th of Dec. Cao to all. Baci



# Donner à voir l'apprentissage

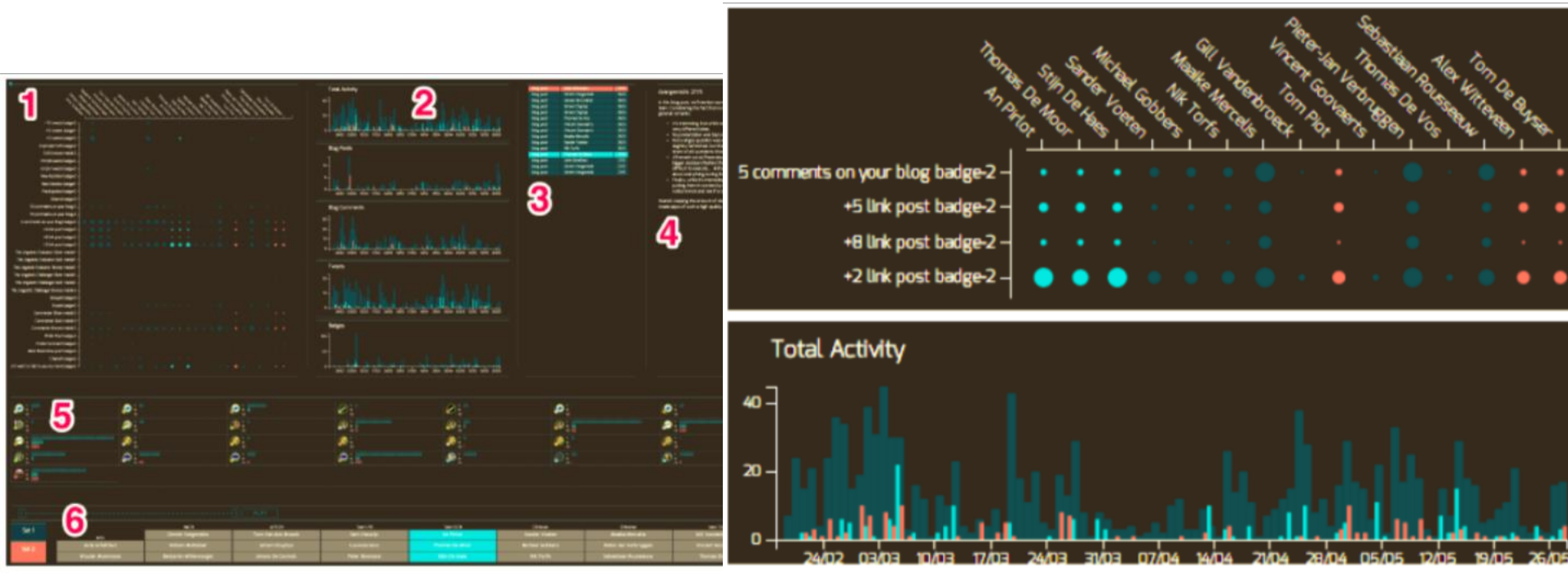
- ◆ Donner aux acteurs de la communauté éducative un pouvoir de décision
- ◆ Permettre aux apprenants de devenir acteur de leur apprentissage
  - ◆ Accéder à leurs propres données
- Tableaux de bord numériques (*Dashboards*)
  - ◆ Offrent aux humains une interprétation visuelle de larges ensembles de données
  - ◆ Pour découvrir, interroger, comprendre les modèles portés par les données et modifier leurs représentations



# Donner à voir l'apprentissage

## *Charleer et al. 2014*

- ◆ LARAe : tableau de bord pour les enseignants
- ◆ Visualisation des traces des apprenants et des enseignants, des badges, de l'activité des étudiants





# Donner à voir l'apprentissage

## *Broisin et al. 2017*

- Objectif :
  - permettre à l'étudiant d'analyser ses activités en temps réel mais aussi *a posteriori*
  - fournir un tutorat intelligent



- Données :
  - Plateforme Lab4CE : environnement web pour des Télé-TP
  - Toutes les interactions sont tracées : connexion, action soumise, correction...

- Technique :
  - Visualisation d'indicateurs, patterns de comportement

```
venant@sieraremi2:~$ chmod -A 700 toto
chmod: -A not implemented
venant@sieraremi2:~$ chmod -a 700 toto
chmod: Unable to translate '700' to a UID
venant@sieraremi2:~$ chmod -b 700 toto
chmod: illegal option -- b
usage: chmod [-fhv] [-R [-H | -L | -P]] [-a | +a | =a [i] [# [ n]]] modeentry file ...
        chmod [-fhv] [-R [-H | -L | -P]] [-E | -C | -N | -i | -I] file ...
venant@sieraremi2:~$ man chmod
venant@sieraremi2:~$ chmod -B 700 toto
chmod: illegal option -- B
usage: chmod [-fhv] [-R [-H | -L | -P]] [-a | +a | =a [i] [# [ n]]] modeentry file ...
        chmod [-fhv] [-R [-H | -L | -P]] [-E | -C | -N | -i | -I] file ...
venant@sieraremi2:~$ man chmod
venant@sieraremi2:~$ chmod 700 -B toto
chmod: illegal option -- B
venant@sieraremi2:~$
```

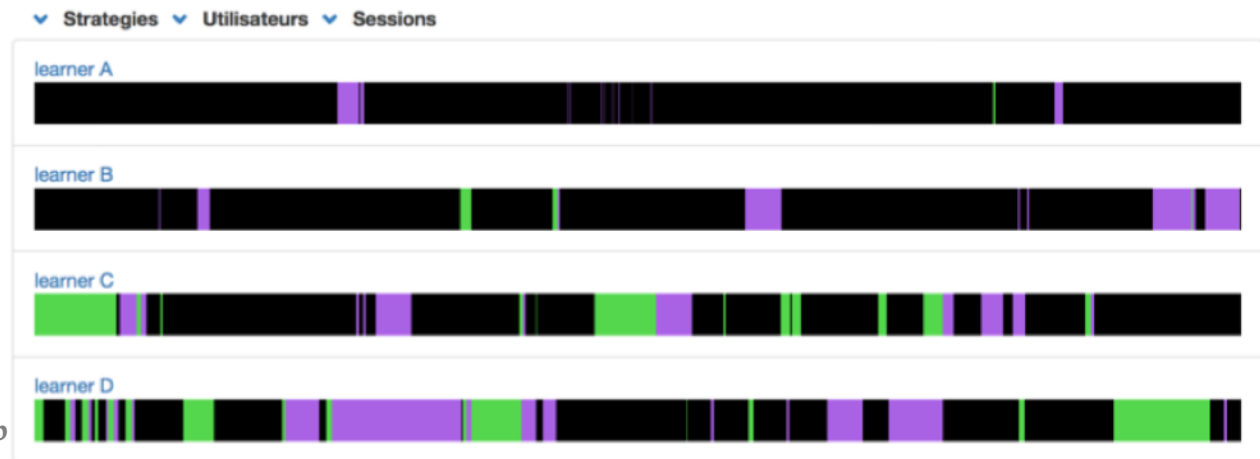
Pas de panique, il semble que [broisin](#) pourrait t'aider, n'hésite pas à lui demander !



# Donner à voir l'apprentissage

## *Broisin et al. 2017*

- Objectif :
  - outil d'analyse du comportement des apprenants pour les **enseignants / chercheurs**
  - utilisé aussi par les apprenants pour l'auto-régulation
- Données :
  - Les mêmes
- Technique :
  - Découverte de patterns

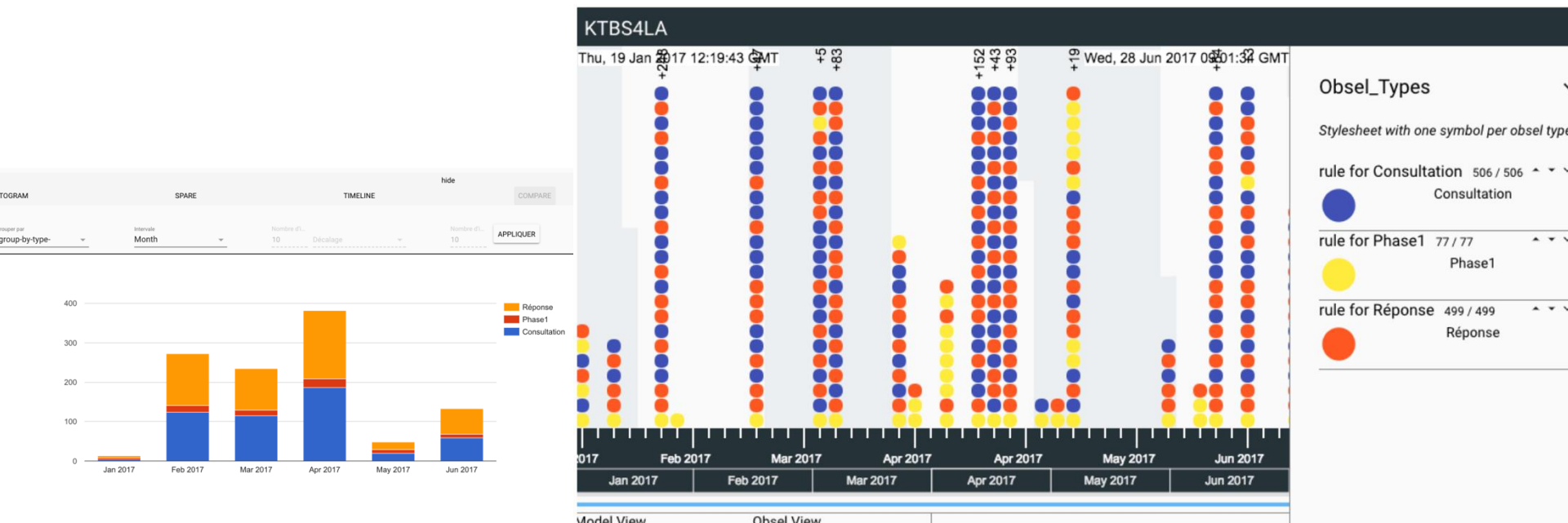




# Donner à voir l'apprentissage

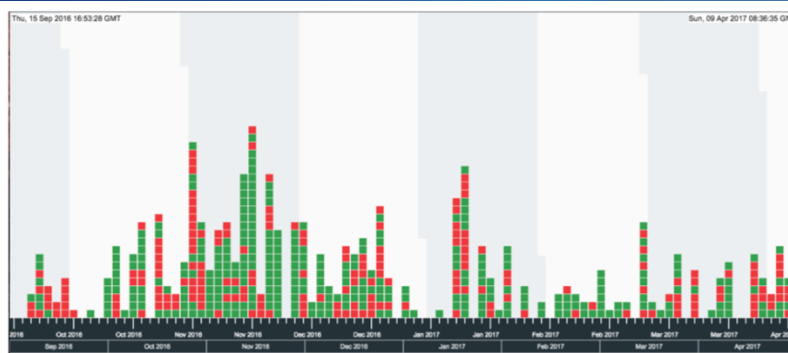
## *Casado et al. 2017*

- Plateforme d'analyse de traces kTBS4LA (kernel for Trace-Based Systems for Learning Analytics)
- Permet une manipulation graphique des données collectées



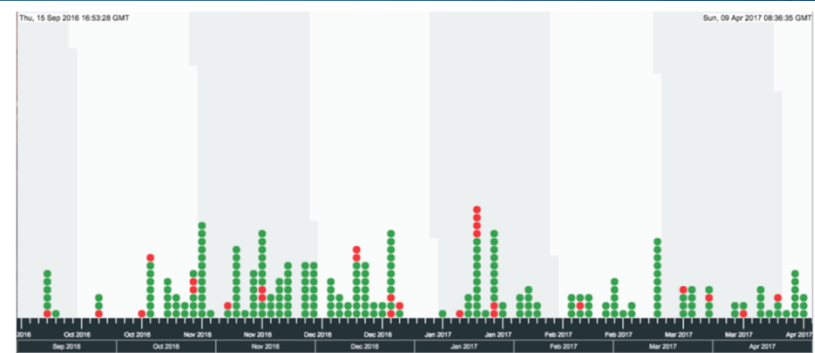


# Donner à voir l'apprentissage



## Garçons

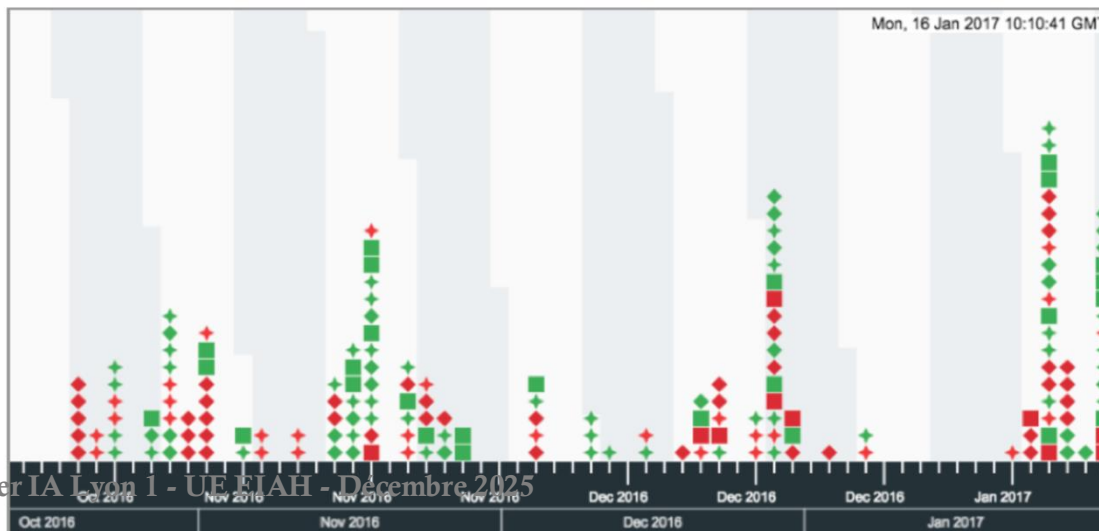
- Pouvoirs collaboratifs (304/493) - 61.7%
- Pouvoirs individuels (189/493) - 38.3%



## Filles

- Pouvoirs collaboratifs (228/254) - 89.8%
- Pouvoirs individuels (26/254) - 10.2%

Timeline avec la visualisation des obsels de genre et de type de pouvoir. Reproduit à partir de « Analyse des traces numériques d'interaction de Classcraft avec kTBS4LA », par Bonvin, G. et Sanchez, E., 2017



## Équipe 1

- P. collab. (28/38) - 73.7%
- P. indiv. (10/38) - 26.3%

## Équipe 2

- ◆ P. collab. (27/71) - 38.0%
- ◆ P. indiv. (44/71) - 62.0%

## Équipe 3

- ◆ P. collab. (44/75) - 58.7%
- ◆ P. indiv. (31/75) - 41.3%



# Donner à voir l'apprentissage *ASKER*

Vue dans le temps des notes obtenues par l'étudiant

Bases de l'algorithmique



Bases du C



Fonctions et procédures



Date : 28/09/2024  
Heure : 16:21:12  
Note : 83.33  
Modèle : La boucle FOR pour afficher

Passage de paramètres



Les tableaux



Les chaînes de caractères



# Donner à voir l'apprentissage *ASKER*

Répartition globale des notes de l'élève dans le dossier LIFAPI





# Donner à voir l'apprentissage

## *Badier et al. 2023*

### Objectif :

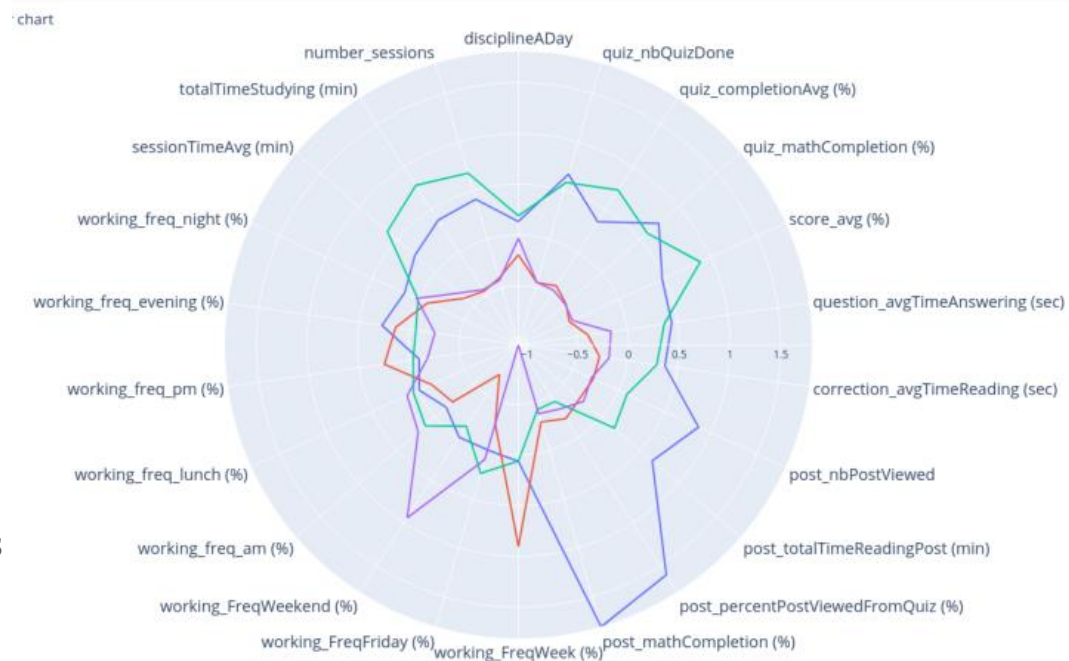
- Identifier un nombre restreint de profils comportementaux similaires sur une application de révision parascolaire

### Méthode :

- Traces de 1000 utilisateurs en classe de 3e entre octobre et décembre 2021
- Définitions d'indicateurs comportementaux
- Analyse par clustering (méthode k-means)

### Résultats :

- Des données très bruitées et très diverses
- Majorité d'utilisation de faible intensité
- Profils identifiés non directement convertibles en stratégies de recommandation





# Plan du cours

- ◆ Objectif de l'analyse de traces en EIAH
- ◆ Ce que l'on peut trouver dans les traces d'apprentissage
- ◆ Deux communautés de recherche : EDM et LAK
- ◆ Exemples de travaux de recherche
- ◆ Ce qu'il reste à faire...
- ◆ Pour aller plus loin : références, conférences et revues



# Perspectives actuelles des 2 communautés

- ◆ Exploiter toutes les traces pour établir les contextes d'observation
  - ◆ Educatives : lire une vidéo, répondre à un quizz
  - ◆ Non-éducatives : issus des capteurs des objets connectés
- ◆ Proposer des modèles et méthodes d'analyse tenant compte de la spécificité du domaine éducatif
  - ◆ Le statut de l'erreur de l'apprenant, son caractère non monotone
  - ◆ L'analyse des interactions sociales vis-à-vis de l'apprentissage
  - ◆ Les différents niveaux de granularité d'analyse (apprenant, classe, institution, ...)
  - ◆ La variété des situations d'apprentissage et des connaissances en jeu



# Perspectives actuelles des 2 communautés

- ◆ Partager les indicateurs, les processus d'analyses, les connaissances et modèles déduits
  - Projet HUBBLE : <http://hubblelearn.imag.fr>
- ◆ Le passage de l'expérimentation à l'industrialisation
  - ◆ Arrivée impromptue des géants de l'économie numérique
    - Scepticisme du public
    - Garde-fous plus que jamais nécessaires car public, pour partie, mineur et données cognitives sensibles
    - Proposer des dispositifs responsables, transparents, confidentiels et accessibles...



# Plan du cours

- ◆ Objectif de l'analyse de traces en EIAH
- ◆ Ce que l'on peut trouver dans les traces d'apprentissage
- ◆ Deux communautés de recherche : EDM et LAK
- ◆ Exemples de travaux de recherche
- ◆ Ce qu'il reste à faire...
- ◆ Pour aller plus loin : références, conférences et revues



# Les conférences et revues

## ◆ Conférence

- ◆ International Educational Data Mining Conferences
- ◆ Conference on Learning Analytics and Knowledge (LAK)

## ◆ Revue

- ◆ Journal of Educational Data Mining : <https://jedm.educationaldatamining.org/>
- ◆ Journal of Learning Analytics : <http://learning-analytics.info/>

## ◆ Société savante

- ◆ IEDMS : <http://educationaldatamining.org/>
- ◆ SoLAR : <https://solaresearch.org/>

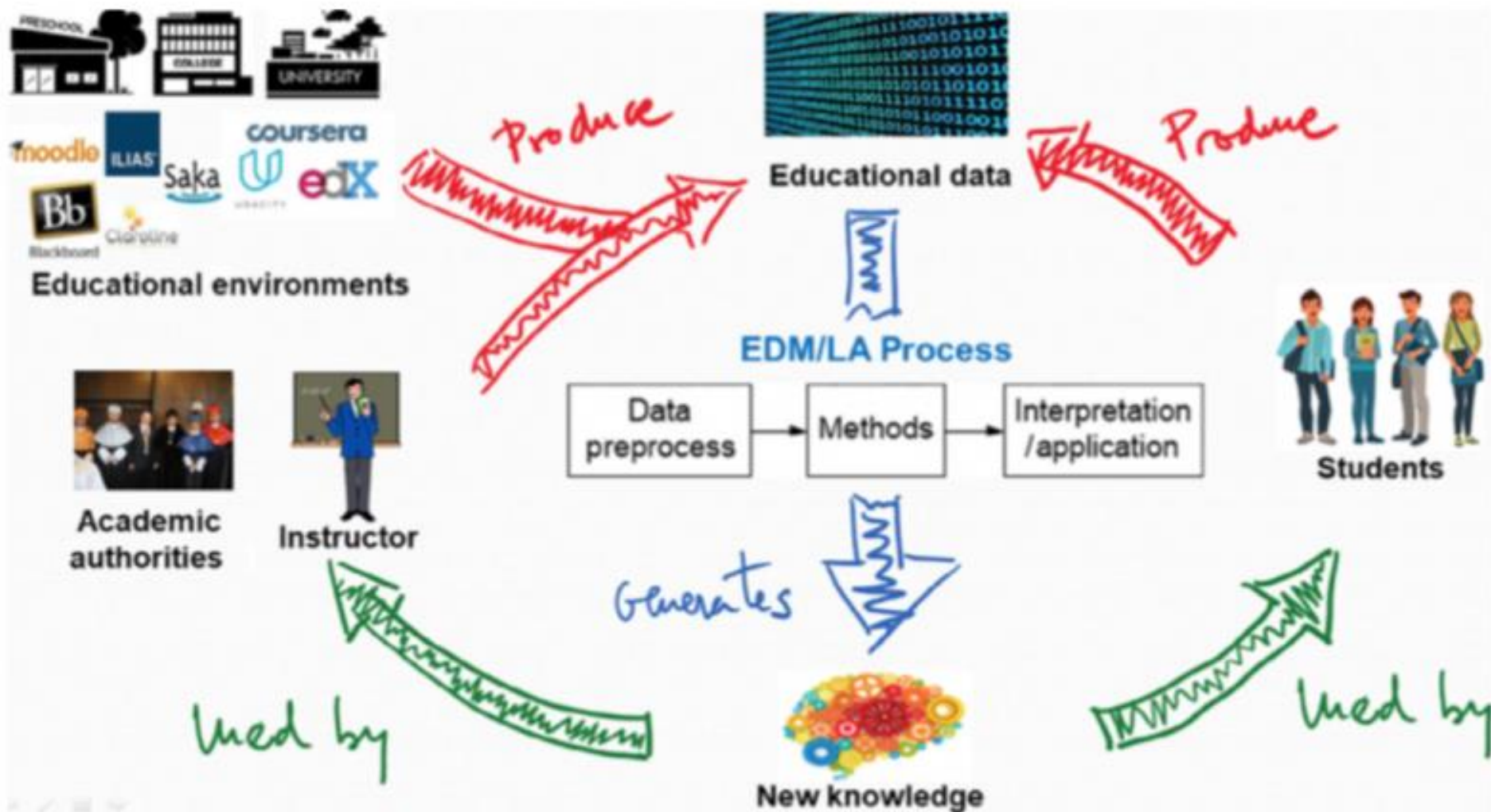


# Groupes de travail

- ◆ Numéri'Lab
  - ◆ Mission d'incubation de la direction du numérique pour l'éducation
- ◆ GTnum2 : Analyse des traces d'apprentissage « learning analytics »
  - ◆ <http://techne.labo.univ-poitiers.fr/gtnum2/>
  - ◆ Axe 1 : Outils et méthodes pour la production et analyse de traces (LIUM & LIRIS)
  - ◆ Axe 2 : Usages des learning analytics (TECHNE)
  - ◆ Axe 3 : Entreprises et organismes pour la production et l'analyse (CNED )
  - ◆ Axe 4 : Droit, Ethique et déontologie (Cecoji)
  - ◆ Axe 5 : Learning analytics vs Intelligence artificielle (LIP & LIRIS)
  - ◆ Axe 6 : Terminologie (TECHNE)



Processus cyclique de découverte des connaissances en Educational Data Mining et Learning Analytics. Reproduit à partir de « Educational data mining and learning analytics: An updated survey », par Romero, C. et Ventura, S., 2020, Wiley Interdisciplinary Reviews-Data Mining and Knowledge Discovery, 10(3), p. 8.





# Les références citées

- ◆ Baker, R., de Carvalho, A. M. J. A., Raspat, J., Aleven, V., Corbett, A. T. et Koedinger, K. R. (2009). Educational Software Features that Encourage and Discourage "Gaming the System". Dans Proceedings of the 14th International Conference on Artificial Intelligence in Education (p. 475-482), IOS Press.
- ◆ Beal, C.R., Qu, L. et Lee, H. (2006). Classifying learner engagement through integration of multiple data sources. Dans AAAI'06. Proceedings of the 21st National Conference on Artificial Intelligence (p. 151-156), New York NY : ACM.
- ◆ Bowers, A. J. (2010). Analyzing the Longitudinal K-12 Grading Histories of Entire Cohorts of Students: Grades, Data Driven Decision Making, Dropping Out and Hierarchical Cluster Analysis. Practical Assessment, Research & Evaluation, 15(7), 1-18.
- ◆ Broisin, J., Venant, R., Vidal, P. (2017). Lab4CE: a Remote Laboratory for Computer Education. International Journal of Artificial Intelligence in Education, 27(1), pp. 154-180.
- ◆ Cen, H., Koedinger, K. et Junker, B. (2006). Learning Factors Analysis - A general method for cognitive model evaluation and improvement. Dans Proceedings of the 8th International Conference on Intelligent Tutoring Systems (p. 164-175), Springer-Verlag Berlin Heidelberg.
- ◆ Charleer, S., Klerkx, J. et Duval, E. (2014). Learning dashboards. JLA, 1(3), 199-202.
- ◆ De Liddo, A., Buckingham Shum, S., Quinto, I., Bachler, M. et Cannavacciuolo, L. (2011). Discourse-centric learning analytics. Dans LAK 2011: 1st International Conference on Learning Analytics & Knowledge, 27 Feb - 01 Mar 2011, Banff, Alberta (p. 23-33), New-York, NY : ACM.
- ◆ Hershkovitz, A., Nachmias, R. (2008). Developing a log-based motivation measuring tool. Dans Proceedings of the First International Conference on Educational Data Mining (p. 226-233).
- ◆ Macfadyen, L. P. et Dawson, S. (2010). Mining LMS data to develop an "early warning system" for educators: A proof of concept. Computers & Education, 54(2), 588-599.
- ◆ Martin, T., Aghababayan, A., Pfaffman, J., Olsen, J., Baker, S., Janisiewicz, P., ... Petrick Smith, C. (2013). Nanogenetic Learning Analytics: Illuminating Student Learning Pathways in an Online Fraction Game. LAK'13, Proceedings of the Third International Conference on Learning Analytics and Knowledge (p. 165-169), New York, NY : ACM.
- ◆ Schneider, B., Abu-El-Haija, S., Reesman, J., et Pea, R. (2013). Toward collaboration sensing: applying network analysis techniques to collaborative eye-tracking data. Dans Proceedings of the Third International Conference on Learning Analytics and Knowledge (p. 107-111), New York, NY : ACM.
- ◆ Siemens, G., Gasevic, D., Haythornthwaite C., Dawson S., Buckingham Shum S., Ferguson R., ... Baker R. (2011). Open Learning Analytics: an integrated & modularized platform Proposal to design, implement and evaluate an open platform to integrate heterogeneous learning analytics techniques. Récupéré de <http://solaresearch.org/OpenLearningAnalytics.pdf>
- ◆ Zimmermann, J., Brodersen K. H., Heinemann, H. R. et Buhmann, J. M. (2015). A Model-Based Approach to Predicting Graduate-Level Performance Using Indicators of Undergraduate-Level Performance. JEDM, 7(3), 151-176.



# Quelques références pour creuser

- ♦ Baker, R. et Siemens, G. (2014). Educational Data Mining and Learning Analytics. Dans K. Sawyer (dir.), Cambridge Handbook of the Learning Sciences: 2nd Edition (p. 253-274), New York , NY : Cambridge University Press .
- ♦ M. A. Chatti, A. L. Dyckhoff, U. Schroeder, et H. Thüs, « A Reference Model for Learning Analytics », Int. J. Technology Enhanced Learning, vol. 4, no 5/6, p. 318-331, 2012.
- ♦ Rapport utilisé pour construire ce cours :
  - ♦ L'analytique des apprentissages numériques.  
Hugues LABARTHE, Vanda LUENGO. Rapport LIP6. Décembre 2016