

Julien Ah-Pine

SIGMA Clermont
20 avenue Blaise Pascal
TSA 62006 63178 Aubière Cedex, France
Mail: julien.ahpine@sigma-clermont.fr

Date of birth: 16/02/1981
Place of birth: Reunion Island
Citizenship: French
Marital Status : Married, 2 children

EDUCATION

- 2003 - 2007 **DOCTORATE IN MATHEMATICS** FROM UNIVERSITY OF PARIS 6, FRANCE, DEFENDED IN 10/2007
Thesis title: Algebraic and combinatoric aspects of Relational Analysis. Applications in clustering, social choice theory and braid theory
- Research work on **relational analysis** (association measures and aggregation procedures for binary relations such as equivalence relations or order relations by using statistical approaches and linear programming), **data-mining, text mining, clustering, social choice theory, rank aggregation, braid and knot theory.**
- 2002 - 2003 **MASTER DEGREE IN STATISTICS** FROM UNIVERSITY OF PARIS 6, FRANCE
- Courses on **parametric and non-parametric statistics, stochastic processes, general linear model, data analysis, spatial statistics, image processing and remote sensing.**
- 2000 - 2003 **MAGISTÈRE DEGREE** (4 YEAR UNIVERSITY DEGREE) IN **ECONOMICS AND STATISTICS** FROM UNIVERSITIES OF TOULOUSE 1 AND 3, FRANCE
- Courses on **econometrics, operations research, microeconomics, macroeconomics, development economics.**
- 2001 - 2002 **BACHELOR DEGREE IN ECONOMETRICS** FROM UNIVERSITY OF TOULOUSE 1, FRANCE

PROFESSIONAL EXPERIENCES

- 2023 - **ASSOCIATE PROFESSOR IN DATA SCIENCE** AT
SIGMA CLERMONT (CLERMONT AUVERGNE INP) AND
LIMOS (UMR 6158 - CNRS/UCA/EMSE -)
- Research work on **machine learning, functional data analysis, time series forecasting, anomaly detection, applications of machine learning to sustainability and disaster sciences, poverty and vulnerability indices.**
 - Associate member of CERDI (UMR 6587 - CNRS/UCA/IRD -) (Research Centre on International Development).
 - Involved in the set up of **1 research project** :
 - ★ Principal Investigator of the DLISCES project (starting in spring 2024) on **Deep Learning, Satellite Images and Maps of Socio-Economic Indicators**, funded by *I-Site CAP 2025/IRC2D*. This multidisciplinary project aims to develop machine learning models in order to infer from diverse and heterogeneous spatial data (satellite images, georeferenced survey data, topographic maps, ...), maps of vulnerability indicators in the context of natural disasters.

2010 - 2023

ASSOCIATE PROFESSOR IN DATA SCIENCE AT UNIVERSITY OF LYON 2 AND ERIC LAB.

- Research work on **machine learning, text mining, natural language processing, preferences elicitation, information fusion, optimization.**
- Publication of **24 articles in peer-reviewed conferences or journals** (7 of ranks **A** or **A***).
- Main contributions :
 - ★ a new framework for **scalable hierarchical clustering**,
 - ★ a new model and inference approach for approximately solving the **graph-based clustering** problem,
 - ★ several contributions in **text mining** (hybrid methods for word sense disambiguation and named entity recognition, sentiment analysis, cluster hypothesis in information retrieval),
 - ★ a new **aggregation function** that allows interaction between criteria using triangular norms,
 - ★ new optimization models for **preference elicitation** with the bipolar **Choquet integral**,
 - ★ a unifying model for graph-based **information fusion in multimedia text-image information retrieval.**
- Involved in the set up of **2 research projects** :
 - ★ Leader of the scalable information retrieval task of the Request project, funded by “*Projets d’Investissement d’Avenir*” of the french government (2014-2017). Collaboration on the **cluster hypothesis in information retrieval**, the development of **scalable framework for hierarchical text clustering and co-clustering**. This project gathered around 15 partners.
 - ★ Principal investigator of the DocNet project on the **social and expertise networks** of scholars in humanities in the Lyon and Saint-Etienne area, funded by the University of Lyon 2 (2011-2012). This project gathered 3 partners.
- Participation in the Imagiweb project on investigating the image of entities in social media, funded by the “*Agence Nationale de la Recherche*” of the french government (2013-2016). Collaboration on **text mining** and **social media analysis** tasks. This project gathered 6 partners.

2007 - 2010

RESEARCH SCIENTIST AT XEROX RESEARCH CENTRE EUROPE

- Research work on **multi-modal information retrieval, meta-search** problems, **text mining** and **machine learning**.
- Main contributions :
 - ★ a new **aggregation function** for **meta-search** (patented),
 - ★ a new hybrid system, **clique-based clustering**, for **named entities disambiguation** (patented).
- Participation in **ImageCLEF 2008 and 2009**, international evaluation campaigns in information retrieval. Our runs at XRCE in multi-modal (image and text) information retrieval (**Photo retrieval tasks**) were ranked firsts among around ten participants.
- Involved in the research activities of the Infom@gic project, funded by the “*Pôle de compétitivité*” *Cap Digital*. Collaboration on the design of methods and tools for **multi-modal information seeking** and their applications to the **Wikipedia corpus**. This project gathered around 20 partners.

- Involved in the set up and the research activities of the SYNC3 project, funded by the *European Union FP7-ICT*. Collaboration on **news event analysis, clustering and opinion extraction from blogs**. This project gathered around 10 partners from 4 different countries.

2003 - 2007

RESEARCH ASSISTANT AT THALES, CENTER FOR INFORMATION ANALYSIS

- Research work on **relational analysis, clustering algorithms, text mining, social choice theory, braid theory**.
- Some main contributions :
 - ★ a new **scalable clustering algorithm (and pre-defined number of clusters not required)** for text mining (patented).
 - ★ several **theoretical results on the relational analysis** framework.
- Involved in the set up of the Infom@gic project on multi-modal information retrieval and analysis, funded by the “ *Pôle de compétitivité*” *Cap Digital*.

RESEARCH INTERESTS

- Machine learning including unsupervised, supervised, deep, transfer learning and optimization.
- Complex data analysis including, images, texts, networks and functional data.
- Information fusion including multimedia information retrieval and seeking, multiview data analysis.
- Decision-making including aggregation functions, graph-based methods, social choice theory.
- Applications in sustainability and disaster sciences.

LIST OF PUBLICATIONS

Peer-reviewed

Journal Articles

- **J. Ah-Pine**. 2022. Learning Doubly Stochastic and Nearly Idempotent Affinity Matrix for Graph-Based Clustering. *European Journal of Operational Research*, 299(3), 1069-1078. [Lien vers le journal, <https://www.sciencedirect.com/science/article/abs/pii/S0377221721010900>].
- **J. Ah-Pine**. 2018. An Efficient and Effective Generic Agglomerative Hierarchical Clustering Approach. *Journal of Machine Learning Research*. 19. [Lien vers le journal, <http://www.jmlr.org/papers/v19/18-117.html>].
- E.P. Soriano-Morales, **J. Ah-Pine**, S. Loudcher. 2016. Using a Heterogeneous Linguistic Network for Word Sense Induction and Disambiguation. *Computación y Sistemas*. 20(3): 315-325. [Lien vers le journal, <http://www.cys.cic.ipn.mx/ojs/index.php/CyS/article/view/2466>].
- **J. Ah-Pine**. 2016. On aggregation functions based on linguistically quantified propositions and finitely additive set functions. *Fuzzy Sets and Systems*. 287:1-21. [Lien vers le journal, <http://www.sciencedirect.com/science/article/pii/S0165011415002870>].
- A. Rolland, **J. Ah-Pine**, B. Mayag. 2015. Elicitation of 2-additive bi-capacity parameters. *EURO Journal on Decision Processes*. 3(1). [Lien vers le journal, <http://link.springer.com/article/10.1007/s40070-015-0043-3>].
- **J. Ah-Pine**, Gabriela Csurka, Stéphane Clinchant. 2015. Unsupervised Visual and Textual Information Fusion in CBMIR Using Graph-Based Methods. *ACM Transactions on Information Systems*. 33(2). [Lien vers le journal, <http://dl.acm.org/citation.cfm?id=2699668>].

- **J. Ah-Pine**. 2011. On data fusion in information retrieval using different aggregation operators. *Web Intelligence and Agent Systems*. 9(1):43-55. [Lien vers le journal, <http://www.iospress.nl/loadtop/load.php?isbn=15701263>].
- **J. Ah-Pine**, F. Marcotorchino. 2010. Unifying some association criteria between partitions using relational matrices. *Communications in Statistics - Theory and Methods*. 39(3):531-542. [Lien vers le journal, <http://www.informaworld.com/smpp/content~content=a918382970~db=all~jumptype=rss>].
- **J. Ah-Pine**, M. Bressan, S. Clinchant, G. Csurka, Y. Hoppenot, J.M. Renders. 2009. Crossing textual and visual content in different application scenarios. *Multimedia Tools and Applications Journal*. 42(1):31-56. [Lien vers le journal, <http://www.springerlink.com/content/5777145v45407864/>].

Book

Chapters

- **J. Ah-Pine**, S. Clinchant, G. Csurka, F. Perronnin, J.M. Renders. 2010. Leveraging text, image and cross-media similarities for diversity-focused multimedia retrieval. *ImageCLEF*. Springer. The Information Retrieval series. Vol 32.
- **J. Ah-Pine**, J.F. Marcotorchino. 2010. Overview of the Relational Analysis Approach and its applications in Data-Mining and Multicriteria Decision Making. *Web intelligence and Intelligent agents*. I-Tech Education and Publishing.

Peer-reviewed

Conferences

- **J. Ah-Pine**. 2023. Sur l'apprentissage d'une matrice d'affinité bistochnastique en clustering. *Actes des Rencontres de la Société Francophone de Classification (SFC 2023)*. [Lien vers la conférence, <https://pfia23.icube.unistra.fr/conferences/sfc/index.html>].
- **J. Ah-Pine** and N. Lebreton. Fusion tardive an analyse de données fonctionnelles élastique. In *53ème Journées de Statistiques, (JDS 2022)*. [Lien vers les proceedings, <https://jds22.sciencesconf.org/data/pages/LivretJds22.pdf>].
- **J. Ah-Pine**. Sur l'apprentissage d'une matrice d'affinité bistochnastique en clustering. In *52ème Journées de Statistiques, (JDS 2021)*. [Lien vers les proceedings, https://jds2021.sciencesconf.org/data/pages/book_jds2021_fr_compressed.pdf].
- **J. Ah-Pine** and A-F. Yao. 2020. Une approche par noyaux multiples pour l'apprentissage non-supervisé de représentation de données fonctionnelles dans des espaces de Sobolev. In *51ème Journées de Statistiques, (JDS 2020)*. [Lien vers les proceedings, https://jds2020.sciencesconf.org/data/pages/book_jds2020_fr_compressed.pdf].
- E.P. Soriano Morales, **J. Ah-Pine**, S. Loudcher. 2017. Fusion Techniques for Named Entity Recognition and Word Sense Induction and Disambiguation. *Proceedings of the 20th International Conference on Discovery Science (DS 2017)*. [Lien vers la conférence, <http://www.iip.ist.i.kyoto-u.ac.jp/ds2017/index.html#1>].
- X. Wang, **J. Ah-Pine**, Jérôme Darmont. 2017. SHCoClust, a scalable similarity-based hierarchical co-clustering method and its application to textual collections. *Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2017)*. [Lien vers la conférence, <http://www.fuzzieee2017.org/>].
- **J. Ah-Pine**. 2017. Sur la normalisation de le matrice Laplacienne en partitionnement spectral. *Actes des XXIVèmes Rencontres de la Société Francophone de Classification (SFC 2017)*. [Lien vers la conférence, <http://polytech-sfc2017.univ-lyon1.fr/>].

- X. Wang, **J. Ah-Pine**, Jérôme Darmont. 2017. A New Test of Cluster Hypothesis Using a Scalable Similarity-Based Agglomerative Hierarchical Clustering Framework. *Proceedings of the 14th “CONFérence en Recherche d’Informations et Applications - Rencontres des Jeunes Chercheurs” (CORIA-RJCRI 2017)*. [Lien vers la conférence, <http://www.lsis.org/coria2017/>].
- **J. Ah-Pine**, X. Wang. 2016. Classification ascendante hiérarchique à noyaux et une application aux données textuelles. *Actes des 17ème Journées Francophones Extraction et Gestion des Connaissances (EGC 2017)*. [Lien vers l’article, <https://editions-rnti.fr/?inprocid=1002306>].
- **J. Ah-Pine**, X. Wang. 2016. Similarity Based Hierarchical Clustering with an Application to Text Collections. *Proceedings of the 15th International Symposium on Intelligent Data Analysis (IDA 2016)*. [Lien vers la conférence, <https://ida2016.blogs.dsv.su.se/>].
- **J. Ah-Pine**, E.P. Soriano Morales. 2016. A Study of Synthetic Over-sampling for Twitter Imbalanced Sentiment Analysis. *Proceedings of the Workshop on Interactions between Data Mining and Natural Language Processing 2016 (DMNLP@ECML-PKDD 2016)*. [Lien vers la conférence, <https://dmnlp.irisa.fr/dmnlp16/>].
- E.P. Soriano Morales, **J. Ah-Pine**, S. Loudcher. 2016. Hypergraph Modelization of a Syntactically Annotated Wikipedia Dump. *Proceedings of the 10th Edition of Language Resources and Evaluation Conference (LREC 2016)*. [Lien vers la conférence, <http://lrec2016.lrec-conf.org/>].
- **J. Ah-Pine**, X. Wang. 2015. Classification ascendante hiérarchique à noyaux et pistes pour un meilleur passage à l’échelle. *Communication aux Journées de Statistiques (JDS 2015)*. [Lien vers la conférence, http://papersjds15.sfds.asso.fr/submission_116.pdf].
- **J. Ah-Pine**, Brice Mayag and Antoine Rolland. 2013. Identification of a 2-additive bi-capacity by using mathematical programming. *Proceedings of the 3rd International Conference on Algorithmic Decision Theory (ADT 2013)*. [Lien vers la conférence, <http://www.adt2013.org/>].
- **J. Ah-Pine**. 2013. Graph Clustering by Maximizing Statistical Association Measures. *Proceedings of the 12th International Symposium on Intelligent Data Analysis (IDA 2013)* [Taux d’acceptation < 23%]. [Lien vers la conférence, <http://www.ida2013.org/>].
- **J. Ah-Pine**. 2013. A general framework for comparing heterogeneous binary relations. *Proceedings of the 1st International Conference on Geometric Sciences of Information (GSI 2013)*. [Lien vers la conférence, <http://www.gsi2013.org/>].
- B. Mayag, A. Rolland, **J. Ah-Pine**. 2012. Elicitation of a 2-additive bi-capacity through cardinal information on trinary actions. *Proceedings of the 14th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU 2012)*. [Lien vers la conférence, <http://www.ipmu2012.unict.it/>].
- S. Clinchant, **J. Ah-Pine**, G. Csurka. 2011. Semantic Combination of Textual and Visual Information in Multimedia Retrieval. *Proceedings of the 1st ACM International Conference on Multimedia Retrieval (ICMR 2011)* [Taux d’acceptation < 35%]. [Lien vers la conférence, <http://www.icmr2011.org/>].
- **J. Ah-Pine**. 2010. Normalized kernels as similarity indices. *Proceedings of the 14th Pacific Asia conference on Knowledge Discovery and Data Mining (PAKDD 2010)* [Taux d’acceptation < 11%]. [Lien vers la conférence, <http://www.iiit.ac.in/conferences/pakdd2010/>].

- **J. Ah-Pine**. 2010. Une famille d'indices de similarité généralisant la mesure cosinus. *Actes des XVIIèmes Rencontres de la Société Francophone de Classification (SFC 2010)*. [Lien vers la conférence, <http://sfc2010.univ-reunion.fr/sfc2010>].
- **J. Ah-Pine**. 2010. On aggregating binary relations using 0-1 integer linear programming. *Proceedings of the 11th International Symposium on Artificial Intelligence and Mathematics (ISAIM 2010)*. [Consultable à l'adresse suivante, <http://gauss.ececs.uc.edu/Workshops/isaim2010/papers/ah-pine.pdf>].
- **J. Ah-Pine**, S. Clinchant, G. Csurka. 2010. Comparison of several combinations of multimodal and diversity seeking methods for multimedia retrieval. *Multilingual Information Access Evaluation Vol. II Multimedia Experiments, 10th Workshop of the Cross-Language Evaluation Forum (CLEF 2009), Lecture Notes in Computer Science (LNCS 6242)*. [Lien vers le journal, <http://www.springerlink.com/content/u517184286474j81/>]
- **J. Ah-Pine**, J.M. Renders, M.L. Viaud. 2009. A Continuum between Serendipitous Browsing and Query-based Search for Multimedia Information Access. *Proceedings of the Adaptive Multimedia Retrieval (AMR 2009) [Workshop]*. [Lien vers la conférence, <http://cabrillo.lsi.uned.es/nlp/amr2009/program>]
- **J. Ah-Pine**. 2009. Cluster analysis based on the central tendency deviation principle. *Proceedings of Advanced Data Mining and Applications (ADMA 2009), Lecture Notes in Artificial Intelligence (LNAI 5678)* [Taux d'acceptation < 12%]. [Lien vers le journal, <http://www.springerlink.com/content/08357122g0831064/>].
- **J. Ah-Pine**, G. Jacquet. 2009. Clique based clustering for improving Named Entity Recognition systems. *Proceedings of the European chapter of the Association of Computational Linguistics (EACL 2009)* [Taux d'acceptation < 28%]. [Consultable à l'adresse suivante, www.aclweb.org/anthology/E09-1007].
- **J. Ah-Pine**, G. Csurka, J.M. Renders. 2009. Evaluation of diversity-focused strategies for Multimedia Retrieval. *Evaluating Systems for Multilingual and Multimodal Information Access, 9th Workshop of the Cross-Language Evaluation Forum (CLEF 2008), Lecture Notes in Computer Science (LNCS 5706)*. [Lien vers le journal, <http://www.springerlink.com/content/h3123p4555v79112/>].
- **J. Ah-Pine**. 2008. Data fusion in information retrieval using consensus aggregation operators. *Proceedings of Web Intelligence (WI 2008) held in conjunction with Intelligent Agent Technologies (IAT 2008) (IEEE/WIC/ACM)* [Taux d'acceptation < 20%]. [Lien vers la conférence, <http://datamining.it.uts.edu.au/conferences/wi08/>].
- **J. Ah-Pine**, F. Marcotorchino. 2007. Statistical, geometrical and logical independences between categorical variables. *Proceedings of Applied Stochastic Models and Data Analysis (ASMDA 2007)*. [Consultable à l'adresse suivante, http://www.asmda.com//CDasmda2007a/conf_papers.html].
- J. Lemoine, H. Benhadda, **J. Ah-Pine**. 2006. Classification non supervisée de documents hétérogènes: Application au corpus 20 Newsgroups. *Proceedings of Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU 2006)*. [Lien vers la conférence, <http://ipmu2006.lip6.fr/>].
- **J. Ah-Pine**, J. Lemoine, H. Benhadda. 2005. Un nouvel outil de classification non supervisée de documents pour la découverte de connaissance et la détection de signaux faibles: Rares Text. *Actes du colloques Ile Rousse 2005*. [Consultable à l'adresse suivante, http://isdms.univ-tln.fr/PDF/isdms22/isdms22_ahpine.pdf].

Conferences

- **J. Ah-Pine**, A-F. Yao. 2021. Multiple kernel SVM for classifying functional data in Sobolev spaces (*Journées MAS 2020*) [Lien vers la conférence, <https://mas2020.sciencesconf.org/>].
- **J. Ah-Pine**, A-F. Yao. 2019. A study of the manifold hypothesis for functional data by using spectral clustering. *12th International Conference of the ERCIM WG on Computational and Methodological Statistics (CMStatistics 2019)* [Lien vers la conférence, <http://www.cmstatistics.org/CMStatistics2019/>].
- **J. Ah-Pine**, S. Clinchant, G. Csurka, Y. Liu. 2009. XRCE's participation in ImageCLEF 2009. *Proceedings of CLEF 2009, workshop to be held during European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2009)*. [Consultable à l'adresse suivante, www.clef-campaign.org/2009/working_notes/ah-pine-paper-new.pdf].
- **J. Ah-Pine**, C. Cifarelli, S. Clinchant, G. Csurka, J.M. Renders. 2008. XRCE's participation to ImageCLEF 2008. *Proceedings of CLEF 2008, workshop to be held during European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2008)* [Workshop]. [Consultable à l'adresse suivante, http://www.clef-campaign.org/2008/working_notes/ah-pine-paperCLEF2008.pdf].

Scientific Magazines

- S. Lorin, **J. Ah-Pine**, S. Caban, H. Benhadda. 2007. L'analyse avancée de l'information par fusion de données textuelles et numériques. *Revue de l'Electricité et de l'Electronique (REE)*, 9, 2007. [Lien vers le journal, <http://cat.inist.fr/?aModele=afficheN&cpsidt=19138736>].

Master Thesis

- **J. Ah-Pine**. 2003. Etude des incertitudes lors de calculs de risques d'inondations: modélisation des erreurs dans les Modèles Numériques de Terrain (MNT).

PATENTS

Acceptés

- **J. Ah-Pine**, H. Benhadda, J. Lemoine. Procédé de classification non supervisée linéaire et stable sur l'ordre des objets. Référencé par FR2895110 (A1) par l'INPI. Référencé WO/2007/068741 par WIPO (procédure internationale)
- **J. Ah-Pine**, S. Clinchant, G. Csurka. System for Multimedia Information Retrieval (octroyé en 05/2013)
- **J. Ah-Pine**, J.M. Renders, M.L. Viaud. System, Method and Apparatus for Information Seeking in a multimedia Text/Image collection (octroyé en 05/2012)
- **J. Ah-Pine**. Data fusion using consensus aggregation functions (octroyé en 04/2012)
- **J. Ah-Pine**, G. Jacquet. Clique based clustering for named entity recognition system (octroyé en 09/2012).

EXPERTISE

Reviewer

- International journals: Journal of Machine Learning Research, Advances in Data Analysis and Classification, IEEE Transactions on Signal Processing, IEEE Transactions on Multimedia, Pattern Recognition Letters.

- International conferences: AAAI, EMNLP, COLING
- Thesis committee
- Examineur de la soutenance de thèse de doctorat de Pierre Bertrand sur le sujet “Conditions de Monge, Transport Optimal et Pont Relationnel - Propriétés, applications et extension du couplage d’indétermination” en 2021.
- Expert R&D projects
- Research projects: French National Agency for Research (ANR -“Agence Nationale de la Recherche”-), ERDF (European Regional Development Fund), French National Association for Research and Technology (ANRT -“Association nationale de la recherche et de la technologie”-).

SKILLS

Programming

Language

- R + Data science libraries
- Python + Data science libraries + Deep learning libraries
- Matlab, Octave
- AMPL

Système

d’exploitation

- Unix/Linux
- Windows

Applications

- $\text{T}_{\text{E}}\text{X}$, $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$, $\text{BIB}\text{T}_{\text{E}}\text{X}$
- Microsoft Office, Open Office

Langues

- English: proficient
- Mandarin: beginner