

**Vendredi 7 Décembre 2018****10h30****UT3 Paul Sabatier, IRIT, Auditorium J. Herbrand****Ophélie FRAISIER****Equipe IRIS - IRIT****Détection de points de vue sur les médias sociaux numériques***Jury :*

- Julien Velcin – *Professeur des universités au laboratoire ERIC (Université Lyon 2), Rapporteur*
- Gaël Dias – *Professeur des universités au GREYC (Université de Caen Normandie), Rapporteur*
- Sihem Amer-Yahia – *Directrice de recherche CNRS au LIG (Université Grenoble Alpes), Examinatrice*
- Guy Melançon – *Professeur des universités au LaBRI (Université de Bordeaux), Examineur*
- Julien Figeac – *Chargé de recherche CNRS au LISST (Université Toulouse 2 Jean Jaurès), Invité*
- Mohand Boughanem – *Professeur des universités à l'IRIT (Université Toulouse 3 Paul Sabatier), Directeur de thèse*
- Romaric Besançon – *Ingénieur de recherche à Nano-Innov (CEA LIST), Co-Directeur de thèse*
- Guillaume Cabanac – *Maître de conférences à l'IRIT (Université Toulouse 3 Paul Sabatier), Encadrant*
- Yoann Pitarch – *Maître de conférences à l'IRIT (Université Toulouse 3 Paul Sabatier), Encadrant*

**Mots-clés :** Fouille de points de vue, médias sociaux, expression politique, humanités numériques

**Résumé :** De nombreux domaines ont intérêt à étudier les points de vue exprimés en ligne, que ce soit à des fins de marketing, de cybersécurité ou de recherche avec l'essor des humanités numériques. Nous proposons dans ce manuscrit deux contributions au domaine de la fouille de points de vue, axées sur la difficulté à obtenir des données annotées de qualité sur les médias sociaux. Notre première contribution est un jeu de données volumineux et complexe de 22 853 profils Twitter actifs durant la campagne présidentielle française de 2017. C'est l'un des rares jeux de données considérant plus de deux points de vue et, à notre connaissance, le premier avec un grand nombre de profils et le premier proposant des communautés politiques recouvrantes. Ce jeu de données peut être utilisé tel quel pour étudier les mécanismes de campagne sur Twitter ou pour évaluer des modèles de détection de points de vue ou des outils d'analyse de réseaux. Nous proposons ensuite deux modèles génériques semi-supervisés de détection de points de vue, utilisant une poignée de profils-graines, pour lesquels nous connaissons le point de vue, afin de catégoriser le reste des profils en exploitant différentes proximités inter-profils. En effet, les modèles actuels sont généralement fondés sur les spécificités de certaines plateformes sociales, ce qui ne permet pas l'intégration de la multitude de signaux disponibles. En construisant des proximités à partir de différents types d'éléments disponibles sur les médias sociaux, nous pouvons détecter des profils suffisamment proches pour supposer qu'ils partagent une position similaire sur un sujet donné, quelle que soit la plateforme. Notre premier modèle est un modèle ensembliste séquentiel propageant les points de vue grâce à un graphe multicouche représentant les proximités entre les profils. En utilisant des jeux de données provenant de deux plateformes, nous montrons qu'en combinant plusieurs types de proximité, nous pouvons correctement étiqueter 98% des profils. Notre deuxième modèle nous permet d'observer l'évolution des points de vue des profils pendant un événement, avec seulement un profil-graine par point de vue. Ce modèle confirme





**IRIT**

*Institut de Recherche  
en Informatique de Toulouse*

qu'une grande majorité de profils ne changent pas de position sur les médias sociaux, ou n'expriment pas leur revirement.



05 61 55 65 10  
[info@irit.fr](mailto:info@irit.fr)  
[www.irit.fr](http://www.irit.fr)

