

Aide à l'Analyse Visuelle de Réseaux Sociaux pour la Détection de Comportements Suspects

Amyr Bannamane*, Hakim Hacid*, Arnaud Ansiaux*, Alain Cagnati**

*Alcatel-Lucent Bell Labs France
Centre de Villarceaux Route de Villejust, 91620 Nozay, France
{prénom.nom}@alcatel-lucent.com,
**Ministère de l'Intérieur, ST(SI)²
alain.cagnati@interieur.gouv.fr

Résumé. Cet article traite de l'analyse visuelle de réseaux sociaux pour la détection de comportements suspects à partir de données de communications fournies à des enquêteurs suivant deux procédures : l'interception légale et la rétention de données. Nous proposons les contributions suivantes : (i) un modèle de données et un ensemble d'opérateurs pour interroger ces données dans le but d'extraire des comportements suspects et (ii) une représentation visuelle conviviale pour une navigation simplifiée dans les données de communication accompagnée avec une implémentation.

1 Introduction

Au delà de l'intérêt qu'apportent les réseaux sociaux en ligne et les moyens de communication à l'utilisateur et aux fournisseurs de services, ces moyens ne sont malheureusement pas toujours utilisés de manière licite et peuvent être utilisés, par exemple, dans la synchronisation d'opérations illégales. Une opération illégale peut aller d'une utilisation illégale simple, p.ex. fraude, des moyens de communication, à des formes plus complexes comme le terrorisme. Pour identifier des comportements suspects, les enquêtes sont généralement menées par des enquêteurs qui ne sont pas experts en informatique. En effet, les enquêteurs sont des «Hommes» de terrain et utilisent les outils informatiques en tant que source d'information additionnelle pour vérifier des hypothèses. Dans ce contexte, les tableurs comme Excel sont utilisés pour gérer des listes, faire des tris et des calculs dans l'espoir de construire des relations significatives entre les données, extrayant ainsi des connaissances utiles et des faits. Les tableurs sont très puissants pour analyser des données tabulaires (Lakshmanan et al. (1998)), mais ne sont pas adaptés pour la mise en évidence de patterns dans les données semi-structurées ou non structurées comportant de nombreuses relations, références croisées et métadonnées. À moins que l'utilisateur ne sache ce qu'il est en train de chercher, il est très difficile d'extraire directement de nouvelles connaissances depuis les tables.

Le problème majeur auquel nos utilisateurs sont confrontés est la capacité à extraire des informations utiles depuis des jeux de données de grande taille, hétérogènes, et éparpillés. Ce