Proposition de recherche doctorale

Détection d'événements et/ou d'anomalies dans les dynamiques de graphes

Mots clés :

- **Directeur de thèse** : Matthieu Latapy
- **Co-encadrant(s)** :
- **Unité de recherche** : Laboratoire d'informatique de Paris 6
- **Ecole doctorale** : École Doctorale Informatique, Télécommunications, Électronique de Paris
- **Domaine scientifique principal**: Divers

Résumé du projet de recherche (Langue 1)

Dans de nombreux contextes (trafic réseau, transactions financières, communications entre individus, etc), on est confronté à des données sous forme de flux de liens : la donnée est essentiellement composée d'une série de couples {(A,B)} avec un horodatage {t} indiquant que {A} a interagi avec {B} à l'instant {t} (par exemple, la machine {A} a envoyé un paquet à la machine {B}, ou le compte bancaire {A} a transféré de l'argent au compte {B}, la personne {A} a envoyé un message à la personne {B}, etc). Dans tous ces contextes, l'analyse de la dynamique des interactions, et notamment la détection d'anomalies et/ou d'événements dans cette dynamique, est un enjeu crucial tant pour les applications que d'un point de vue fondamental. On peut penser par exemple à la détection d'attaques sur un réseau, de fraudes bancaires ou de corruption, ou encore de changements significatifs dans un réseau social. Il est très naturel de voir ce type de données comme des graphes dynamiques (avec les notations ci-dessus, {A} et {B} sont des sommets, et {(A,B)} une arête) et de formuler les problématiques de détection d'événements et/ou d'anomalies en ces termes. Cette approche permet notamment d'espérer exploiter la très riche structure des interactions observées, ce qui laisse présager un renouveau profond et des progrès considérables pour le thème. Cette direction a toutefois été très peu abordée pour l'instant, et ce projet propose de l'explorer. Un exemple représentatif est la détection d'anomalies (attaques notamment) sur des réseaux informatiques. On peut typiquement effectuer une capture de trafic sur un routeur d'entrée ou un firewall, ce qui permet d'observer les paquets entrants et sortants du sous-réseau surveillé. Les IDS ({Intrusion Detection Systems}) actuels se basent essentiellement sur des analyse des volumes échangés et des protocoles utilisés. Mais l'information disponible est bien plus riche, puisqu'à chaque paquet correspond un expéditeur et un destinataire. Voir le paquet comme un {lien} entre l'expéditeur et le destinataire, dans un graphe dynamique constitué par tout le trafic observé, permet d'analyser cette information avec une approche {graphes} telle que celle que nous proposons ici. Une attaque serait alors visible par une évolution anormale de la dynamique de ce graphe, qui reste à caractériser. De la même manière, des logs d'accès à des services (un site web par exemple) peuvent être modélisés par des liens entre les utilisateurs (ou leur adresse IP) et les services utilisés (les pages web consultées par exemple) ; une attaque est alors visible dans la dynamique du graphe obtenu. Les achats en ligne peuvent être vus comme des liens entre clients et produits, et la fraude comme une dynamique anormale dans le graphe induit. Les exemples sont multiples, et les applications très nombreuses. L'équipe d'accueil dispose de diverses données (notamment des échanges en ligne, des logs d'activité réseau, ou des captures de trafic) qu'il serait intéressant d'aborder sous cet angle. Elle a de plus défini de premières propriétés pour l'analyse de graphes dynamiques de ce type et la détection d'événements dans ces dynamiques. L'essentiel reste toutefois à faire, et nous souhaitons développer cette thématique.

Résumé du projet de recherche (Langue 2)

Il s'agira tout d'abord de considérer plusieurs cas pratiques, représentatifs de la gamme des objets modélisables de cette façon, et pour lesquels l'équipe d'accueil a les données et l'expertise nécessaires. Il faudra alors produire un outil logiciel générique et performant pour analyser ces dynamiques, et l'appliquer à des cas variés. On identifiera et on formalisera les questions en commun soulevées par les différents cas pratiques, comme la convergence de propriétés observées ou la détection d'événements et d'anomalies. On s'attachera notamment à la définition de notions spécifiques aux graphes dynamiques de ce type, qui seront un pendant aux notions usuelles sur les graphes classiques. Une direction particulièrement prometteuse est la prise en compte de l'ancienneté des liens et l'identification de plusieurs échelles de temps pertinentes pour l'observation. Dans cette optique, des approches combinant graphes, traitement du signal, statistiques, fouille de données et/ou algorithmique semblent les plus pertinentes.

Informations complémentaires (Langue 1)

Le projet est naturellement inséré dans la recherche internationale et les nombreuses collaborations de l'équipe. Des séjours à l'étranger seront encouragés.