ED STIC - Proposition de Sujets de Thèse

pour la campagne d'Allocation de thèses 2017

Axe Sophi@Stic :

SystèmesRéseaux|

Titre du sujet :

Exploring trust on Twitter

Mention de thèse :

Informatique

HDR Directeur de thèse inscrit à l'ED STIC :

Arnaud Legout

Co-encadrant de thèse éventuel :

Nom :

Giroire

Prénom :

Frédéric

Email :

frederic.giroire@cnrs.fr

Téléphone :

Email de contact pour ce sujet :

arnaud.legout@inria.fr

Laboratoire d'accueil :

INRIA

Description du sujet :

Twitter is the most popular micro-blogging service in the world. It was created to enable people to find
out what is currently happening with people and organizations they are interested in. The relation
between users on Twitter is different from classical social networks like Facebook. Instead of
bidirectional friendship link that are initiated by one user and accepted by another, Twitter uses the
concept of following. Users can follow other users they are interested in, which means they subscribe
for all the messages they sent. So, the links on Twitter are unidirectional, if someone follows you, you
don't need to follow back. Twitter is a very interesting object of study because the unidirectional

model
of relationship is the closest to real-life communications, thus a huge societal impact.

Whereas Twitter is de facto standard to find trends in the Internet, it has been criticized for the difficulty to filter out spams and attempts to manipulate trends. Even Twitter is still not able to efficiently filter out the noise.

The goal of this Ph.D. is to tackle the difficult and impactful problem of filtering out spams and attempts
to manipulate trends in a way that is robust to escape techniques. The current techniques are based
on the content of the tweets or on the characteristics of the profiles sending the tweets. However, such semantic techniques can be easily worked around. Instead, we propose to develop new techniques based on the structural properties of the social graph, which is way harder to work around.

For instance, we propose to study how to propagate a trust relationship by exploiting a transitive relationship called chain of trust. We also propose to work on structural properties of the graph of hashtags to find anomalies representing attempts to manipulate.

The student will have the opportunity to work on a unique dataset we collected in 2012 [GRL_SIG_14].
This dataset represents the entire Twitter social graph with more than 500 million accounts and 24
billion links. He will have to overcome several challenges such as the collection and construction of
hashtag graphs based on real data, or the design of fast graph matching algorithms to perform on-thefly
classification.

[GRL_16] Maksym Gabielkov, Arthi Ramachandran, Arnaud Legout, Augustin Chaintreau. Social Clicks: What and Who Gets Read on Twitter? In Proc. of ACM SIGMETRICS 2016, June 14--18, 2016,
Antibes Juan-les-Pins, France.

[GRL_SIG_14] Maksym Gabielkov, Ashwin Rao, and Arnaud Legout. Studying Social Networks at Scale: Macroscopic Anatomy of the Twitter Social Graph. In Proc. of ACM SIGMETRICS 2014, June 16--
20, 2014, Austin, Texas, USA.

[DGPP_17] Guillaume Ducoffe, Frédéric Giroire, Stéphane Pérennes, Stefano Ponziani. Revisiting Preferential Attachment with applications to Twitter, Inria Research Report, 2017.

[FGJ+_14] Fedor Fomin, Frédéric Giroire, Alain Jean-Marie, Dorian Mazauric, and Nicolas Nisse. To Satisfy Impatient Web surfers is Hard. Theoretical Computer Science, 526:1–17, 2014.

URL : http://www-sop.inria.fr/members/Arnaud.Legout/Projects/sotweet.html

English version: