

ED STIC - Proposition de Sujets de Thèse pour la campagne d'Allocation de thèses 2015

Axe Sophi@Stic :

Titre du sujet :

Mention de thèse :

HDR Directeur de thèse inscrit à l'ED STIC :

Co-encadrant de thèse éventuel :

Nom :

Prénom :

Email :

Téléphone :

Email de contact pour ce sujet :

Laboratoire d'accueil :

Description du sujet :

What is the size of online social network in terms of its members? What is the average degree of the nodes? Is correlation between neighbours' degrees positive or negative? How frequently does a typical user visit the online social network and for how long he or she stays connected? Is there influence among the friends in the online social network? All such interesting questions can be answered by means of sampling techniques. The most basic technique is uniform sampling. However, it can be very costly in terms of API requests if the ID space of the users is sparsely populated. To remedy this problem one can use a variety of random walk methods, which visit the nodes by following friendship links. Still this is not without a cost. The random walk methods bring dependence and some bias (e.g., towards oversampling the nodes with large degrees). We aim to study analytically and empirically various network sampling processes and to

recommend best solutions. We and other authors [ART10,GKBM10] have preliminary results indicating that combining together random walk methods with uniform restart and thinning produce good results. We feel that the design of random walk based methods using advanced methods from Markov chain theory and adaptive control theory will lead to the creation of high performance sampling methods.

References:

[ART10] Avrachenkov, K., Ribeiro, B., and Towsley, D.

Improving random walk estimation accuracy with uniform restarts.

In Proceedings of WAW 2010, Algorithms and Models for the Web-Graph, pp. 98-109, Springer, 2010.

[GKBM10] Gjoka, M., Kurant, M., Butts, C. T., and Markopoulou, A.

Walking in facebook: A case study of unbiased sampling of OSNs.

In Proceedings of IEEE INFOCOM 2010, pp. 1-9, 2010.

Required competences:

We are looking for a candidate with sound knowledge of probability theory.

Good knowledge in one of the following topics is a plus: linear algebra, graph theory, control theory. We expect that a candidate can program well in matlab or python.

English version:

What is the size of online social network in terms of its members? What is the average degree of the nodes? Is correlation between neighbours' degrees positive or negative?

How frequently does a typical user visit the online social network and for how long he or she stays connected? Is there influence among the friends in the online social

network? All such interesting questions can be answered by means of sampling techniques. The most basic technique is uniform sampling. However, it can be very costly in

terms of API requests if the ID space of the users is sparsely populated. To remedy this problem one can use a variety of random walk methods, which visit the nodes by

following friendship links. Still this is not without a cost. The random walk methods bring dependence and some bias (e.g., towards oversampling the nodes with large degrees).

We aim to study analytically and empirically various network sampling processes and to recommend best solutions. We and other authors [ART10,GKBM10] have preliminary

results indicating that combining together random walk methods with uniform restart and thinning produce good results. We feel that the design of random walk based methods

using advanced methods from Markov chain theory and adaptive control theory will lead to the creation of high performance sampling methods.

References:

[ART10] Avrachenkov, K., Ribeiro, B., and Towsley, D.

Improving random walk estimation accuracy with uniform restarts.

In Proceedings of WAW 2010, Algorithms and Models for the Web-Graph, pp. 98-109, Springer, 2010.

[GKBM10] Gjoka, M., Kurant, M., Butts, C. T., and Markopoulou, A.
Walking in facebook: A case study of unbiased sampling of OSNs.
In Proceedings of IEEE INFOCOM 2010, pp. 1-9, 2010.

Required competences:

We are looking for a candidate with sound knowledge of probability theory.

Good knowledge in one of the following topics is a plus: linear algebra, graph theory, control theory. We expect that a candidate can program well in matlab or python.