Theoretical Computer Science 584 (2015) 3–18

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

Self-organizing flows in social networks

Nidhi Hegde [a], Laurent Massoulié [b,1], Laurent Viennot [c,*,2]

[a] Bell Labs France Alcatel-Lucent, Route de Villejust, 91620 Nozay, France
[b] Microsoft Research – Inria Joint Centre, campus de l'école polytechnique, 1 rue Honoré d'Estienne d'Orves, 91120 Palaiseau, France
[c] Inria – Paris Diderot University, LIAFA Case 7014, 75205 Paris, France

ARTICLE INFO

Article history:
Received 25 November 2013
Received in revised form 9 February 2015
Accepted 10 February 2015
Available online 16 February 2015

Keywords:
Network formation
Self-organization
Budget of attention
Price of anarchy
Social filtering

ABSTRACT

Social networks offer users new means of accessing information, essentially relying on "social filtering", i.e. propagation and filtering of information by social contacts. The sheer amount of data flowing in these networks, combined with the limited budget of attention of each user, makes it difficult to ensure that social filtering brings relevant content to interested users. Our motivation in this paper is to measure to what extent self-organization of a social network results in efficient social filtering.

To this end we introduce *flow games*, a simple abstraction that models network formation under selfish dynamics, featuring user-specific interests and budget of attention. In the context of homogeneous user interests, we show that selfish dynamics converge to a stable network structure (namely a pure Nash equilibrium) with close-to-optimal information dissemination. We show that, in contrast, for the more realistic case of heterogeneous interests, selfish dynamics may lead to information dissemination that can be arbitrarily inefficient, as captured by an unbounded "price of anarchy".

Nevertheless the situation differs when user interests exhibit a particular structure, captured by a metric space with low doubling dimension. In that case, natural autonomous dynamics converge to a stable configuration. Moreover, users obtain all the information of interest to them in the corresponding dissemination, provided their budget of attention is logarithmic in the size of their interest set.

1. Introduction

Information access has been revolutionized by the advent of social networks such as Facebook, Google+ and Twitter. These platforms have brought about the new paradigm of "social filtering", whereby one accesses information by "following" social contacts.

This is especially true for twitter-like microblogging social networks. In such networks the functions of filtering, editing and disseminating news are totally distributed, in contrast to traditional news channels. The efficiency of social filtering is critically affected by the network topology, as captured by the contact-follower relationships. Today's networks provide recommendations to users for potentially useful contacts to follow, but don't interfere any further with topology formation. In this sense, these networks self-organize, under the selfish decisions of individual users.

* Corresponding author.
E-mail addresses: nidhi.hegde@alcatel-lucent.com (N. Hegde), laurent.massoulie@inria.fr (L. Massoulié), laurent.viennot@inria.fr (L. Viennot).
[1] Part of this work was done while at Technicolor.
[2] Supported by the Inria project-team "Gang" in the "LIAFA" laboratory and by the "LINCS" laboratory.

http://dx.doi.org/10.1016/j.tcs.2015.02.018
0304-3975/

This begs the following question: when does such autonomous and selfish self-organizing topology lead to efficient information dissemination? The answer will in turn indicate under what circumstances self-organization is insufficient, and thus when additional mechanisms, such as incentive schemes, should be introduced.

Two parameters play a key role in this problem. On the one hand each user aims to maximize the coverage of the topics of his interest. On the other hand, a user pays with his attention: filtering interesting information from spam (i.e. information that does not fall in his topics of interest) incurs a cost. Users must therefore trade-off topic coverage against attention cost. As pointed out by Simon [1], as information becomes abundant, another resource becomes scarce: attention.

Furthermore, there is an interplay between participants in a social network where filtering by one user may benefit another, inducing complex dependencies in decisions on creating connections. To model this, we introduce a network formation game called *flow game* where some users produce news about specific topics and each user is interested in receiving all news about a set of topics specific to him. Each user is a selfish agent that can choose his incoming connections within a certain budget of attention in order to maximize the coverage of his set of topics of interest.

This model is of interest on its own, as it enriches the class of existing network formation games with a focus on flow dissemination under bounded connections. This model could also be of interest in the context of peer-to-peer streaming and file sharing or publish/subscribe applications.

1.1. Our results

An important feature in our model is a user's budget of attention for the consumption of content. In previous work [2] the budget of attention was modeled as a limit on the rate with which a user consults a friend, with a different objective of minimizing delay in receiving all content. In the present work we are interested in a more fundamental question, of how efficient social networks are formed in the first place. We consider the model where users are interested in specific subsets of topics and their objective is to maximize the number of flows received corresponding to these topics. As such, we model the budget of attention as a constraint on the number of connections a user may create (rather than a rate of consultation). Our aim is to build a simple model capturing the complexity of the problem. This way of capturing the budget of attention amounts to assuming that each connection consumes the same amount of attention. We discuss in Section 6 how we can tweak our model to more finely model attention consumption.

We capture users' interests in topics through user-specific values for each topic and define the *utility* a user receives to be the sum of values of all received topics. Each user's objective in a *flow game* is then to choose connections so as to maximize his utility. We additionally assume that a user may produce news about one topic at most even if he redistributes other topics. This is coherent with an empirical study of twitter traces [3] where it is shown that ordinary users (as opposed to celebrities or newspapers) can gain influence by concentrating on a single topic.

Our main results relate to the stability and efficiency of the formation of information flows. We derive conditions where selfish dynamics converge to a pure Nash equilibrium. We then give approximation ratios bounding the quality of an equilibrium compared to an optimal solution. This is traditionally measured through the price of anarchy, the ratio of the global welfare (measured as the sum of user utilities) at an optimal solution compared to that at the worst equilibrium.

More precisely, we first consider homogeneous games where all users are interested in the same set of topics. We can then prove that selfish dynamics always converge to an equilibrium. Selfish dynamics comprise of any sequence of moves, where in each move a user is given the opportunity to selfishly rewire his connections to increase his utility. We show moreover that convergence occurs within a polynomial number of rounds where a round is a sequence of selfish moves including at least one move per user. We additionally show that the price of anarchy is bounded and approaches 1 as the budget of attention of users increases.

In the more general case where users interests are heterogeneous, selfish dynamics may not converge and price of anarchy may be unbounded. However, we observe that fast convergence towards efficient configurations can occur when users' interests are captured by a metric space with sufficient structure. Here, the interests of a user are modeled a point in this space such that nearby topics are of interest to the user. Sufficient structure typically arises when the metric space is a Euclidean space with low dimension. Our results are tailored to the more general case of metrics with low doubling dimension. Low dimension assumptions are classically used in information retrieval when data can be viewed as a matrix which is approximated with a low rank matrix. For example, a ranking technique for the web is proposed in [4] using a 16-dimensional space for representing topics of web pages. Closer to the context of our study, modeling people's opinions as points in a low Euclidean space is a classical approach in social sciences. Political spectrum for example is often modeled as a one dimensional space along a Left–Right axis, consist in introducing more dimensions. This concept can be formalized with single-peaked preference curves [5]. An online system for exchanging political views could be a concrete example where the technical conditions of our model are met. We believe that the same applies for the various domains of interest of a user, implying that our model remains valid more broadly if we can attribute several points of interest (one per domain) to each user. An extension of our model in that direction is proposed.

1.2. Related work

Information spread in networks has been studied extensively. Much of the past work study the properties of information diffusion on given networks with given sharing protocols. Our goal in this work is to study how networks form when users