Pervasive and Mobile Computing 9 (2013) 324–335

Contents lists available at SciVerse ScienceDirect

Pervasive and Mobile Computing

journal homepage: www.elsevier.com/locate/pmc

Fast track article

GoDisco++: A gossip algorithm for information dissemination in multi-dimensional community networks

Rajesh Sharma *, Anwitaman Datta

School of Computer Engineering, Nanyang Technological University, Singapore

ARTICLE INFO

Article history:
Available online 4 October 2012

Keywords:
Gossip algorithm
Community networks
Selective dissemination
Egocentric social network

ABSTRACT

In this paper we propose GoDisco++, a gossip based approach for information dissemination in online social community networks. GoDiscoo++ uses local information available to nodes—that is information associated with a node and its neighbors. The algorithm exploits multiple relations which may exist between nodes, and applies social principles and behavior inspired decentralized mechanisms for *targeted dissemination*. The dissemination process works with the dual aims of (i) maximizing the spread among relevant nodes (*high recall*) and (ii) minimizing spamming among non-relevant nodes (*high precision*). Such a designed dissemination scheme can have interesting applications like probabilistic publish/subscribe, decentralized recommendation and contextual advertisement systems, to name a few. We validate the proposed approach with simulation experiments performed using real and synthetic datasets.

1. Introduction

Interest communities are present in many networked and distributed environments. In online social networks there is often an explicit concept of communities or groups based on declared interests of individuals, as well as implicit communities that can be inferred from the link structures. Likewise, in other networks, such as those derived from communication patterns, or in traditional file sharing peer-to-peer systems where people with similar interests self-organize into a semantically clustered overlay [1,2], implicit communities may be identified. Recently, various efforts to build peer-to-peer online social networks (P2P OSNs) and applications have also been underway [3–6]. Such systems are being investigated to achieve privacy and autonomy from big brotherly service providers [7]. Likewise, in virtual worlds such as massively multiplayer online games (MMOGs), virtual communities are formed.

1.1. High level objective

It is often necessary to have a mechanism to disseminate information in such virtual communities. Since the users who form a community (be it explicit or implicit) are driven by shared interests, a dissemination mechanism should take into consideration that ideally only users who might be interested in a particular message should receive it. Essentially, such a selective dissemination mechanism needs to

- maximize spreading (high recall): The message should reach out to as many as possible relevant users who would be interested in the information.
- minimize spamming (high precision): The message should reach out to as few as possible non-relevant nodes.

* Corresponding author.
E-mail addresses: raje0014@e.ntu.edu.sg, raje0014@ntu.edu.sg (R. Sharma), anwitaman@ntu.edu.sg (A. Datta).

1574-1192/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.pmcj.2012.09.005

As compared to centralized settings, this is non-trivial to achieve in decentralized settings due to lack of global knowledge or coordination, as well as the system's dynamics, such as churn or change of user interests. Thus it is desirable to design mechanisms that utilize only information that is locally available, and hence, easy to update as well.

1.2. Motivation

Circulation of call for papers (CFP) among academic peers is a good analogy to understand the motivation behind our work. For carrying out such circulation, one may use mailing lists, which is analogous to a very simplistic 'topic based' publish/subscribe (pub/sub) system. If the mailing lists' scopes are too narrow, a user would however have to subscribe to many such mailing lists. A user might miss out some information if he fails to subscribe to any relevant mailing list. On the other hand, if there are few mailing lists, each covering a vast amount of topics, a user might receive unwanted information most of the times.

We often observe colleagues and collaborators also adopt ad-hoc, word-of-mouth ways to propagate CFPs by personal email. However the decisions as to whom to and whom not to forward a CFP is taken by individuals based on their perception of what their buddies' interests may be. Likewise, as a receiver, each individual may build local (implicit) trust metrics to decide which friends typically forward useful or useless information, providing a personal, subjective context to determine the quality of information. However, this unstructured word of mouth (gossip/epidemic) approach does not guarantee full coverage, and can also generate many duplicates. At the same time, such redundancy can however also make the dissemination process robust against isolated communication failures. Such resilience of gossip/epidemic algorithms is well studied and exploited [8–11] in general.

Despite the well recognized role of word of mouth mechanisms in selective information dissemination in real life as well as over the internet (such as by emails or on the blogosphere and social media), to the best of our knowledge, there has not been any algorithmic (designed) information dissemination mechanism leveraging on the community structures and semantics available in social information systems. This motivated us to propose GoDisco [12] in the past which can be applied in systems such as P2P OSNs, [3] virtual communities [5] and MMOGs [13] where the mechanism can be naturally explored. In this paper, we refine the previous ideas to propose GoDisco++.

1.3. Difference between GoDisco and GoDisco++

In GoDisco [12] we proposed and investigated a gossip based, social principles and behavior inspired decentralized mechanism to disseminate information in a distributed setting, using exclusively social links and exploiting semantic context to keep the dissemination process selective to relevant nodes by exploring social aspects such as degree and activeness of the users along with behaviorial aspects such as history as a forwarder. This work is primarily a refinement of our previous work [12]. The novelty of this extension is the exploration of a multi-dimensional social network, whose semantics can be exploited to achieve better dissemination characteristics as compared to what is achieved using a single dimensional social network.

1.4. Contribution

Our main contribution in this paper (some shared with GoDisco [12]) is the information dissemination algorithm (we call it GoDisco++) which does not require global knowledge of the network. The mechanism explores social principles such as social triad, activeness, degree and is behavior inspired, for example — exploiting forwarding behavior of the neighbors.

The designed mechanism is expected to be useful for distributed social information systems [6,3,14,15] as well as in other settings such as word of mouth marketing strategies in job market. We evaluate our mechanisms using both synthetic and real datasets of social communities, particularly looking at metrics such as recall, precision, etc. Experiments show that the proposed mechanism is moderately effective and efficient.

The rest of the paper is organized as follows. In Section 2, we present related works. We present our approach including various concepts and algorithm in Section 3. Section 4 evaluates the approach using synthetic and real networks. We conclude in Section 5 by indicating potential future directions.

2. Related works

Targeted dissemination of information can be carried out efficiently using a structured approach as is the case with publish/subscribe systems [16] that rely on infrastructure like overlay based application layer multicasting [17] or gossiping techniques [10] to spread the information within a well-defined group. Well defined groups are not always practical, and alternative approaches to propagate information in unstructured environments are appealing. Broadcasting information to everyone ensures that relevant people get it (high recall), but is undesirable, since it spams many others who are uninterested (poor precision). In mobile ad-hoc and delay tolerant networking scenarios [18], selective gossiping techniques relying on user profile and context to determine whether to propagate the information have been used. This is analogous to how multiple diseases can spread in a viral manner — affecting only "susceptible" subset of the population. Autonomous