

Colocation Cascade: Bringing Social Cascade to the Physical Space

Aysha Alsafty

Fayoum University, Egypt
Aysha.Alsafty at cantab dot net

David J. Greaves

The Computer Laboratory, University of Cambridge
David.Greaves at cl dot cam dot ac dot uk

Abstract

The need to exploit the physical vicinity of mobile users in social platforms has never been more evident as increasingly more users access social networks from mobile devices. The effect of online social connections has already been shown as instrumental in spreading content and influence via social cascade. In this paper, we examine the phenomenon of social cascade when confined to the user's physical vicinity. We examine the intersection between patterns of both social cascade and physical collocation. We combine traces of user mobility with models of online social cascade and show that content propagated via social cascade on online social networks are likely to be already cached on the devices of colocated peers.

Introduction

Increasingly, users access online social networks (OSNs) via mobile devices; whereas of the second quarter of 2014, 62% of Facebook advertisement revenue is generated from mobile traffic. In the arena of mobile social networks, several studies have shown self-reported social links, such as the ones manifested in online social networks, to be a predictor for the frequency of physical collocation between socially-linked users (Hui et al. 2011; Bigwood et al. 2011). The increased probability of physical encounters among socially-connected users has been further exploited for routing content in delay-tolerant networks (Costa et al. 2008; Yoneki et al. 2008; Jahanbakhsh et al. 2010). Moreover, as social relations were shown to affect mobility and collocation patterns, we aim to explore how social cascade patterns in the OSN relate to collocation patterns in the mobile social network (MSN). Correlation of OSN content access and physical collocation can be exploited in multiple opportunities and applications, such as realizing large scale deployment of distributed web of trust systems and friend recommendation. A suite of applications can be built to propagate “nearby” popular content to accompany and drive location-based advertisement and services. As short

range, wide bandwidth connectivity become more ubiquitous, the opportunity to localize the retrieval of bandwidth-expensive content to colocated peer devices is yet another opportunity.

Using Social Cascade for Localized Access

The social cascade phenomenon captures the impact of social connections in increasing the user's bias to access the same content as her peers. The user becomes interested in accessing content after and in part due to her becoming aware that one or more social connections have accessed such content. In this paper we attempt to quantify the feasibility of the social cascade phenomenon within a collocation network. This is a preliminary study, where we quantify the coincidental probability of social cascade intersecting with collocation patterns. If such correlation exists, we envision multiple applications and incentives to further encourage accessing social content from colocated peers.

We define *colocation cascade* as the probability of successfully obtaining a copy of OSN cascade content from a socially-connected and physically-colocated peer versus retrieving it from the OSN directly. Colocation cascade thus captures the correlation between the progression of online cascade users' physical encounters. We quantify colocation cascade as a probability. If a data item is propagated via online social cascade from node A to node B at time t , and if B is colocated with A within time window w from t , this occurrence is recognized as an instance of colocation cascade. If the data item can be retrieved via colocation cascade, this is considered as a hit in the putative distributed social cache that could be emergent from the devices of colocated peers, otherwise it is a miss. We then calculate the probability of hits versus misses for the full cascade study.

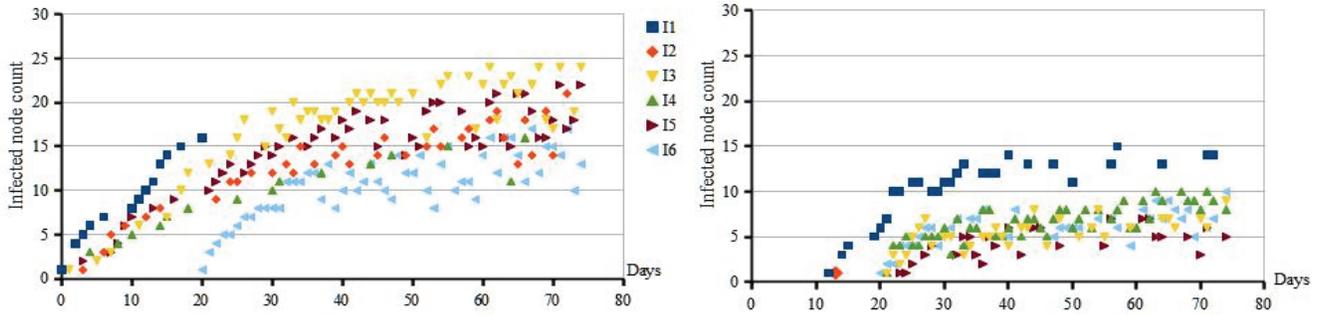


Figure 1: Infection propagation in the (a) long chain model and (b) short chain model

Modeling Social Cascade

To begin to understand colocation cascade, we need to first model online social cascade. There is not yet a consensus on such a model. In light of previous studies of cascade dynamics, we recognize two possible models; long chain model (LCM) and short chain model (SCM). We construct the long chain model based on the empirical analysis of Cha et al. (Cha et al. 2008) for the propagation of popular “favourited” photos on Flickr. The short cascade model is based on Sun et al. (Sun et al. 2009) that analyzed the act of “fanning” pages via Facebook news feeds. As early analysis of the social cascade phenomenon built on epidemiology literature, the term *infection* is used to refer to the spread of the interest in a data item from a user to a social peer.

Characteristic	Long chain model	Short chain model
Starter nodes	Few starter nodes	14-47% of all nodes
Infection introduction rate	Biased to early in cascade lifetime	Spread throughout cascade life time
Infection propagation rate	Controlled by exposure density	unlimited
Chain length	unlimited	Limited according to a right-skewed distribution

Table 1: Characteristics of long vs. short cascade models

The long chain model further builds on epidemiology literature in assuming an infection is introduced to a social network through one node. In contrast, Sun et al. demonstrated that cascade behavior is started by many starter nodes. Each starter node then triggers multiple cascade chains, most of which are of short length. Many short chains eventually conglomerate into one big cluster as these chains overlap at multiple nodes. The model further imposes a right-skewed distribution on chain length and a

cascade chain is obsoleted after reaching this length and it is not propagated any further. In contrast, the long chain model imposes no limit on the length of a cascade chain and assumes an infection is sustained for the length of the experiment.

While both cascade models show a correlation between exposure to the cascade item and the probability of adopting the infection object, they differ in their findings on the number of starter nodes and the chain length or depth. The long chain model assumes a minimal number of starter nodes, while the short chain model found that 14-47% of nodes are involved as cascade starters.

The main difference between the two models, and which motivates naming the models is the cascade depth. LCM adopts that cascade chains grow indefinitely; in contrast, SCM found that chain length followed a right-skewed distribution, where most chains were fewer than 3-nodes long. More recently, Dow et al. contributed to the discussion on whether a deep or a shallow cascade model is more accurate (Dow et al, 2013). They found out that by correcting for exposure sequence, a deep model was a more accurate presentation of Facebook large cascades. We here introduce both models still as the dynamics of multiple online social communities vary enough to potentially allow for both models to exist in different communities.

Long Chain Model (LCM)

The long chain model (LCM) distinguishes two types of propagation steps; initial steps that introduce a popular photo to a social cluster via starter nodes versus following steps, where the photo propagates via social links in the network.

LCM introduces most infections early on in the lifetime of the cascade, where 50% of the infections occur within the first 3 days of object introduction. To simulate LCM, we control the probability of following infections so that infections are introduced into the network according to the number of exposures a node sees from friends. Our model dictates that 45% of infected users were exposed to the photo via three or more infectors, 35% of users had a sin-

gle infector and 20% of users had two infectors. The length of the chain is unlimited and infections persist for the period of the experiment.

Short Chain Model (SCM)

In the short chain cascade model (SCM), infections are introduced to the population throughout the cascade lifetime. The length of a chain is controlled to reflect a distribution whereby 25% of chains are of length 1, 25% of chains are of length 2, 25% of chains are of length 3, 20% of chains are of length 11 or fewer and 98% are of length 18 or fewer. Our model adheres to these distributions and ends the propagation of an infection when the maximum length is reached. This is in addition to 14-47% of nodes being cascade starters.

Experiments

Evaluating colocation cascade requires three-fold knowledge: the user’s social graph, traces of online social cascade and colocation traces. For this initial study of the concept, we rely on simulation based on synthesizing two datasets. Social cascade traces are constructed via the long and short chain models discussed in the previous sections. The two cascade models are applied to the real-life social graph of a group of users and for a period of 75 days (to match the time length of the colocation traces). The propagation of cascade infections on the social graph is recorded as cascade traces. These traces are then replayed against the real-life colocation traces of the same group of users.

Dataset

For the social graph and colocation traces, we use the SASSY dataset, collected by St. Andrews university and available via the CRAWDAD archive for crawled datasets (Bigwood et al. 2011). This dataset was collected over 75 days, where 25 users carried T-mote sensors in a university environment. The dataset captures 112264 encounters between users with a mean of around 1496 encounters a day and a median of 252 encounters a day.

Simulation Setup

Using a discrete event simulator we developed in Python, we capture the propagation of the cascade through the social graph during 75 days. We construct the social cascade traces by applying the Long and short chain models to the social graph of the users in the SASSY dataset. A number of independent infections is generated according to LCM and SCM. For the results shown here, we generate 7 independent infections and run the simulation 10,000 times, to account for the random factors in LCM and SCM. One example infection trace is shown in figure 1. The se-

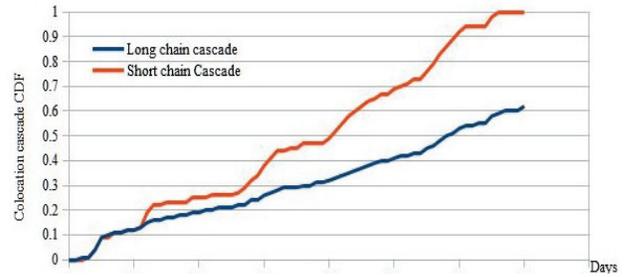


Figure 2: Colocation cascade CDF, long versus short chain models

cond phase of the simulation is then to replay these cascade trace while simulating the physical encounters of the same group of users.

Measuring Colocation Cascade

We calculate the probability of colocation cascade by measuring the number of times an online cascade between two users can be satisfied via a colocation with the same user. This occurrence counts as a hit and is counted against misses as the probability of colocation cascade. We use a time window of 8 hours; that is, a cascade interaction is considered a hit or what we call a collocated cascade if the two users involved have a physical encounter within a period of 8 hours from the online cascaded time. The time window choice is due to the natural breakdown of daily activities, This assumes that a user affected by cascade of online content will still be interested in accessing the content up to 8 hours from exposure time.

The long chain model achieves a cumulative distribution of 0.69 colocation cascade hits, while the short chain model sees a higher rate reaching a CDF of 1 before the end of the simulation as shown in figure 2. This is perhaps due to the small size of population in the dataset leading to early saturation in the case of the long chain model. Meanwhile, as the short chain model distributes starter infections throughout the lifetime of the cascade, new starter infections can help the cascade chain escape a saturated clique of users to further spread the content throughout the full population.

Discussion

Our simulation finds that colocation cascade occurs as a natural side-effect of online social networks going mobile. Contributing factors are the social cascade phenomenon, our tendency to befriend people we frequently encounter and to be around some of our online social friends. A suite of applications that fully expose and exploit this phenomenon is yet to be developed. While some mobile applica-

tions already transparently utilize user location to optimize application behavior, these have been limited to location but not colocation data— perhaps due to privacy challenges. Our simulation accounts only for coincidental colocation cascade; that is naturally occurring colocation cascade— in absence of any incentive or application-driven interactions. We thus believe that our findings are an underestimation of the occurrence of colocation cascade. We envision users to favor localized access if an application suite existed.

One challenge of further developing our model is that of gathering a dataset that combines three pieces of information simultaneously: online social graph, physical colocation and content access traces. Few such datasets exist due to the difficulty anonymizing them, as experienced with Harvard's Facebook dataset (Lewis et al. 2012).

Related Work

Several works have looked into location information that users self-reported via “check-ins” on online social networks. Location information was then used to infer and study community formation and dynamics (Brown et al. 2012; Allamanis et al. 2012). We believe that sentient colocation follows distinct dynamics and can serve a different suite of applications and perhaps system-level services. More recently, Socievola et al. introduced the concept of multi-layer social network, where they recognize implicit colocation information as a system-level activity rather than a social application in itself (Socievole et al. 2014). Distributed social networks, like Persona and Diaspora (Baden et al. 2009; Bielenberg et al. 2012) put user location at the core of the social network design; they are not close to replacing online social networks. The wealth of user data already on OSNs as well as the exhaustiveness of the social connection directory accumulated by their current users are likely to make users hesitate before quitting an OSN for a distributed version.

Conclusions

We have introduced the concept of colocation cascade and attempted to estimate quantified it via simulation and using a real-life dataset of user encounters and their online social graphs. In the process, we built two models of online social cascade based on empirical studies of two of the largest online social networks: Flickr and Facebook. Our findings show that colocation can be an effective means for sharing socially-propagated content, laying the groundwork for further analysis and application.

References

- Allamanis, M., Scellato, S., & Mascolo, C. (2012, November). Evolution of a location-based online social network: analysis and models. In *Proceedings of the 2012 ACM conference on Internet measurement conference* (pp. 145-158). ACM.
- Baden, R., Bender, A., Spring, N., Bhattacharjee, B., & Starin, D. (2009, August). Persona: an online social network with user-defined privacy. In *ACM SIGCOMM Computer Communication Review* (Vol. 39, No. 4, pp. 135-146). ACM.
- Bielenberg, A., Helm, L., Gentilucci, A., Stefanescu, D., & Zhang, H. (2012, March). The growth of diaspora—a decentralized online social network in the wild. In *Computer Communications Workshops (INFOCOM WKSHPS), 2012 IEEE Conference on* (pp. 13-18). IEEE.
- Bigwood, G., Rehunathan, D., Bateman, M., & Bhatti, S. (2011, June). CRAWDAD project, “A community resource for archiving wireless data at Dartmouth,” <http://crawdad.cs.dartmouth.edu>, 2012.
- Brown, C., Nicosia, V., Scellato, S., Noulas, A., & Mascolo, C. (2012, August). The importance of being placefriends: discovering location-focused online communities. In *Proceedings of the 2012 ACM workshop on Workshop on online social networks* (pp. 31-36). ACM.
- Cha, M., Mislove, A., Adams, B., & Gummadi, K. P. (2008, August). Characterizing social cascades in flickr. In *Proceedings of the first workshop on Online social networks* (pp. 13-18). ACM.
- Costa, P., Mascolo, C., Musolesi, M., & Picco, G. P. (2008). Socially-aware routing for publish-subscribe in delay-tolerant mobile ad hoc networks. *Selected Areas in Communications, IEEE Journal on*, 26(5), 748-760.
- Dow, P. A., Adamic, L. A., & Friggeri, A. (2013, June). The Anatomy of Large Facebook Cascades. In *ICWSM*.
- Hui, P., Crowcroft, J., & Yoneki, E. (2011). Bubble rap: Social-based forwarding in delay-tolerant networks. *Mobile Computing, IEEE Transactions on*, 10(11), 1576-1589.
- Jahanbakhsh, K., Shoja, G. C., & King, V. (2010, February). Social-greedy: a socially-based greedy routing algorithm for delay tolerant networks. In *Proceedings of the Second International Workshop on Mobile Opportunistic Networking* (pp. 159-162). ACM.
- Lewis, K., Kaufman, J., Gonzalez, M., Wimmer, A., & Christakis, N. (2008). Tastes, ties, and time: A new social network dataset using Facebook. *com. Social networks*, 30(4), 330-342.
- Socievole, A., De Rango, F., & Caputo, A. (2014, November). Wireless contacts, Facebook friendships and interests: Analysis of a multi-layer social network in an academic environment. In *Wireless Days (WD), 2014 IFIP* (pp. 1-7). IEEE.
- Sun, E., Rosenn, I., Marlow, C., & Lento, T. M. (2009, May). Gesundheit! Modeling Contagion through Facebook News Feed. In *ICWSM*.
- Yoneki, E., Hui, P., Chan, S., & Crowcroft, J. (2007, October). A socio-aware overlay for publish/subscribe communication in delay tolerant networks. In *Proceedings of the 10th ACM Symposium on Modeling, analysis, and simulation of wireless and mobile systems* (pp. 225-234). ACM.