Journal of Universal Computer Science, vol. 18, no. 8 (2012), 1022-1031
submitted: 16/9/11, accepted: 14/12/11, appeared: 28/4/12 © *J.UCS*

Beating Social Pulse: Understanding Information Propagation via Online Social Tagging Systems[1]

Xuan Hau Pham
(Department of Computer Engineering
Yeungnam University, Gyeongsan, Korea
pxhauqbu@gmail.com)

Jason J. Jung[2]
(Department of Computer Engineering
Yeungnam University, Gyeongsan, Korea
j2jung@{gmail.com, ynu.ac.kr})

Dosam Hwang
(Department of Computer Engineering
Yeungnam University, Gyeongsan, Korea
dosamhwang@gmail.com)

Abstract: Social media (e.g., Twitter and FaceBook) have been one of the most popular online communication channels to share information among users. It means the users can give (and have) cognitive influences to (and from) the others. Thus, it is important for many online collaborative applications to understand how the information can be propagated via such social media. In this paper, we focus on a social tagging system where users can easily exchange resources as well as their tags with other users. Given a certain tag from a temporal folksonomy, the social pulse can be established by counting the number of users (or resources). Particularly, we can discover meaningful relationship between tags by computing inducibility. To conduct experimentation, a tag search system has been implemented to collect a dataset from Flickr.

Key Words: Social pulse; Social tagging; Information propagation; Inducibility.
Category: H.1.1, H.3.5, I.2.11

1 Introduction

Information (or knowledge) can be propagated and disseminated by interactions among people through various communication media. Depending on many internalization processes (e.g., enriching background knowledge and psychological modification), the people can somehow understand the information, and they will exploit it for their tasks in various contexts. Thus, it is important to discover meaningful patterns from this social phenomena of information propagation via certain media [Cha et al. 2009]. Various patterns of information prop-

[1] This paper is significantly revised from an earlier version presented at the 4th Asian Conference on Intelligent Information and Database Systems (ACIIDS 2012).
[2] Corresponding author

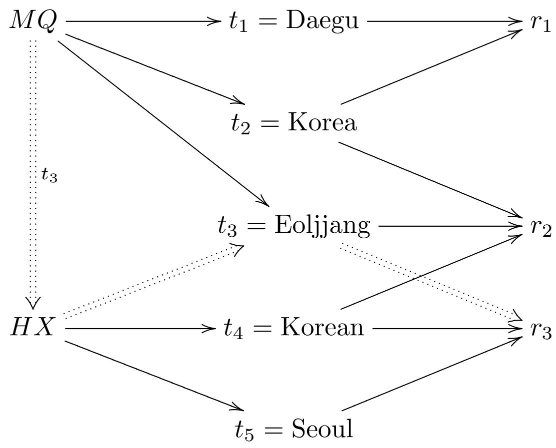

Figure 1: A simple example on a social tagging system with two online users (i.e., MQ and HX); They have tagged three resources r_1, r_2, and r_3 with five tags t_1, t_2, t_3, t_4 and t_5.

agation have been studies in many scientific areas, e.g., information science, cognitive science, neuroscience, and so on.

Especially, online social media have been playing an important role of propagating information to many users [Yang and Leskovec 2010]. There have been a number of social network services (SNS), e.g., Twitter and FaceBook, to distribute up-to-date information quickly and widely. Various social media analysis schemes have been studied for understanding information flow [Jung 2012a, Jung 2012c], identifying experts, and topical authorities [Pal and Counts 2011].

In this paper, we focus on online social tagging systems (also, called folksonomies) which is one of the well known online collective intelligence applications. Online users can represent various resources (e.g., bookmarks, musics, and photos) as a set of keywords, and share them with other users. We can also find the information is propagated among the users through this social tagging system. For example, in Fig. 1, user MQ has added t_3 (Eoljjang[3]) which is a *neologism* (a newly generated term among young generations). User HX will have a chance to recognize the new term and learn the meaning of the term.

Since a certain information, e.g., neologisms, is generated within a social tagging system by some events and someones, various propagation patterns can be revealed. Also, we can regard this propagation process as a contextual synchronization among users [Jung 2011a]. Hence, we propose a novel method to extract such propagation patterns within a social tagging system. Thereby, temporal folksonomy and social pulse are formulated to understand the social propaga-

[3] http://en.wikipedia.org/wiki/Eoljjang

tion patterns.

The outline of this paper is as follows. In Sect. 2, we will describe backgrounds and basic notations. Sect. 3 introduces social pulses for modeling online social tagging systems. Sect. 4 shows an experimental results for evaluating the proposed methods. In Sect. 5, several discussion issues will be addressed. Finally, Sect. 6 draws our conclusions of this work.

2 Backgrounds and notations

Social tagging process among multiple users can generate a specific information sphere, called a folksonomy, where information can be propagated.

Definition 1 (Folksonomy [Jung 2012b]). Given a set of users U, their own tag sets T can be employed to describe their contexts about resources R. Thus, a folksonomy can be represented as

$$\mathcal{F} = \langle U \times R \times T \rangle \tag{1}$$

where R indicates a set of resources tagged by the users.

Definition 2 (Temporal folksonomy). Given a folksonomy, it can be integrated with a temporal information. A temporal folksonomy is represented as

$$\mathcal{F}^{\tau} = \langle U \times R \times T \times \tau \rangle \tag{2}$$

where τ is a timestamp of the corresponding tag.

For example, as shown in in Table 1, suppose that a temporal folksonomy has been built. Once we analyze the timestamp of tags, we can find out that t_3 has been propagated from MQ to HX after a certain time delay.

3 Discovering and understanding social pulses

Once a set of tags of each user is projected from a temporal folksonomy, it can be easily interpreted as a discrete signal which is a sequence of events over time [Jung 2010b]. More importantly, we can set a time window (of which size is δ), since we need to understand how the tags is propagated in the folksonomy. Thus, given a temporal folksonomy $\mathcal{F}_{\tau}$, we can discover a *social pulse*, which is composed of co-occurred tagging patterns within a time window.

Definition 3 (Social pulse). Given a tag $t \in T$, a social pulse is composed of two parts; a time window $w_{\tau_t} = [\tau_t, \tau_t + \delta]$, and a pitch (a height of the pulse). Especially, the pitch can be represented as either the number of users who have

Table 1: An example of a temporal folksonomy

Users	Tags	Resources	Timestamps
MQ	t_2	r_2	2011/02/15 15:12:37 ($\tau_{t_2}^0$)
HX	t_5	r_3	2011/02/16 20:21:53 ($\tau_{t_5}^0$)
MQ	t_2	r_1	2011/02/20 09:12:53 ($\tau_{t_2}^1$)
HX	t_4	r_3	2011/02/22 22:34:41 ($\tau_{t_4}^0$)
HX	t_4	r_2	2011/02/23 13:43:34 ($\tau_{t_4}^1$)
MQ	t_3	r_2	*2011/03/01 21:12:23* ($\tau_{t_3}^0$)
HX	t_3	r_3	*2011/03/01 23:05:31* ($\tau_{t_3}^1$)
MQ	t_1	r_1	2011/03/04 22:54:21 ($\tau_{t_1}^0$)

applied the same tag or the number of resources which have been attached by the tag. Thus, it can be formalized by

$$\mathcal{P}_t = \{\langle w_{\tau_t}, \pi_{\tau_t}\rangle | \tau_t \in \tau\} \quad (3)$$

where the pitch of a time window w_{τ_t} can be computed by

$$\pi_{\tau_t} = |\{u_i | u_i \times t_j \times r_k \times \tau_{t_j} \in \mathcal{F}_\tau, t_j = t, \tau_{t_j} \in [\tau_t + \delta]\}| \quad (4)$$

$$= |\{r_k | u_i \times t_j \times r_k \times \tau_{t_j} \in \mathcal{F}_\tau, t_j = t, \tau_{t_j} \in [\tau_t + \delta]\}| \quad (5)$$

where $u_i \in U$ and $r_k \in R$. These two different measures (Equ. 4 and Equ. 5) are user-based pitch and resource-based pitch, respectively.

From Table 1, the social pulses can be represented like in Table 2. Equ. 4 is used to measuring the pitch. Basically, these social pulses are a sequence of discrete step functions, and they can be easily transformed into a cumulative form.

Table 2: An example on social pulses with Table 1; The size of time window δ is set to a day.

Tag	Social pulse
t_2	$\{\langle[\tau_{t_2}^0, \tau_{t_2}^0 + \delta], 1\rangle, \langle[\tau_{t_2}^1, \tau_{t_2}^1 + \delta], 0\rangle\}$
t_5	$\{\langle[\tau_{t_5}^0, \tau_{t_5}^0 + \delta], 1\rangle\}$
t_4	$\{\langle[\tau_{t_4}^0, \tau_{t_4}^0 + \delta], 1\rangle, \langle[\tau_{t_4}^1, \tau_{t_4}^1 + \delta], 0\rangle\}$
t_3	$\{\langle[\tau_{t_3}^0, \tau_{t_3}^0 + \delta], 2\rangle, \langle[\tau_{t_3}^1, \tau_{t_3}^1 + \delta], 0\rangle\}$
t_1	$\{\langle[\tau_{t_1}^0, \tau_{t_1}^0 + \delta], 1\rangle\}$

Since the social pulses of tags are established from a given temporal folksonomy, we can extract the following two propagation patterns for understudying the temporal folksonomy; *i*) speed, and *ii*) convergence rate.

3.1 Propagation speed

We can comprehend the *speed* of tag propagation, i.e., how quickly a tag is propagated to other users in a certain time duration. Since a tag t has been firstly used at τ_t^0, we can measure the pitch of the social pulse π_{τ_t}.

Definition 4 (Speed). Given a tag t and its social pulse $\mathcal{P}_t$, the speed of its propagation can be measured by

$$\mathcal{S}_t = \max_{\mathcal{P}_t} \left(\frac{\pi_{\tau_t}}{\delta} \right) \tag{6}$$

where δ is a size of window.

The speed simply indicates the maximum propagation power of the corresponding tag. We can measure the speed, when the social pulse of the tag shows the highest pitch. In other words, if we can construct a cumulative curve of the social pulse, we can easily find the speech at the steepest slope.

The speed of tag propagation can be used for measuring spontaneous and prompt response on the the tag. For example, it is easy to understand most online neologisms as well as fun to use them. For example, in Table 2, tag t_3 shows the highest speed, since it can an easy term to understand.

3.2 Propagation convergence rate

Also, we can measure the convergence rate of the propagated tag, i.e., how quickly the tag is spread to most of users.

Definition 5 (Convergence rate). Given a tag t and its social pulse $\mathcal{P}_t$, the convergence rate of its propagation can be measured by a temporal duration

$$\mathcal{C}_t = |\tau_t^{\Omega} - \tau_t^0| \tag{7}$$

where τ_t^{Ω} is the time ending the social pulse (i.e., $\pi_{\tau_t^{\Omega}} = 0$). It means that there is no more meaningful social pulse after τ_t^{Ω}.

Similar to the propagation speed, convergence rate is also an important indicator for measuring propagation power. Moreover, we assume that convergence rate means how long it take for all users in the folksonomy to realize the tag (or its meaning).

3.3 Relationship between tags

More interesting task is to discover relationships between tags in a given folksonomy. Most of the previous studies for folksonomy analysis focus mainly on co-occurrence patterns between tags. For example, if two tags are applied together to more resources (or by more users), then we can determine that the relationship between these tags are stronger than others.

Different from such co-occurrence pattern-based schemes, in this paper, we are interested in a new measurement between tags, called *inducibility.* We assume that as the tags in a folksonomy can be propagated to other users, they can encourage the users to apply some relevant tags (especially, to create some new tags, e.g., neologism), depending on various lingual practice of the online users [Jung 2012b].

Definition 6 (Inducibility). Two given tags t_i and t_j, inducibility $\mathcal{I}_{t_i \to t_j}$ can be measured by the minimized summation of temporal delays between the corresponding social pulses (i.e., $\mathcal{P}_{t_i} = \{\langle w_{r_{t_i}}, \pi_{r_{t_i}} \rangle\}$, $\mathcal{P}_{t_j} = \{\langle w_{r_{t_j}}, \pi_{r_{t_j}} \rangle\}$). It is computed by the following equations

$$\mathcal{I}_{t_i \to t_j} = \frac{1}{1 + \min \Delta_w(\mathcal{P}_{t_i}, \mathcal{P}_{t_j})} \quad (8)$$

$$= \frac{|\mathcal{P}_{t_i}|}{1 + \min \Delta_w(\mathcal{P}_{t_i}, \mathcal{P}_{t_j})} \quad (9)$$

$$\Delta_w = \sum_{\pi_{r_{t_i}} \geq \zeta, \pi_{r_{t_j}} \geq \zeta} |w_{r_{t_j}} - w_{r_{t_i}}| \quad (10)$$

where ζ is a threshold to remove the trivial social pulses. In Equ. 9, the denominator can normalize the value by choosing the maximum length of the social pulse.

Once we have social pulses from a given temporal folksonomy, we can compare a pair of social pulses for measuring inducibility between tags. Moreover, we take into account the directionality of the inducibility (i.e., $\mathcal{I}_{t_i \to t_j} \neq \mathcal{I}_{t_j \to t_i}$).

4 Experimental results

In order to evaluate the proposed schemes, we have built a temporal folksonomy $\mathcal{F}^\tau$ from Flickr[4]. Particularly, for sampling tags, we have collected 55 Korean internet neologism (i.e., $|T| = 55$) from wikipedia[5]. With $\delta = 28$ days, Fig. 2 and Fig. 3 show temporal distributions of resources (i.e., photos, since Flickr is a social media for sharing photos) and users, respectively. From both cumulative plots, we can find out that there are various temporal distributions.

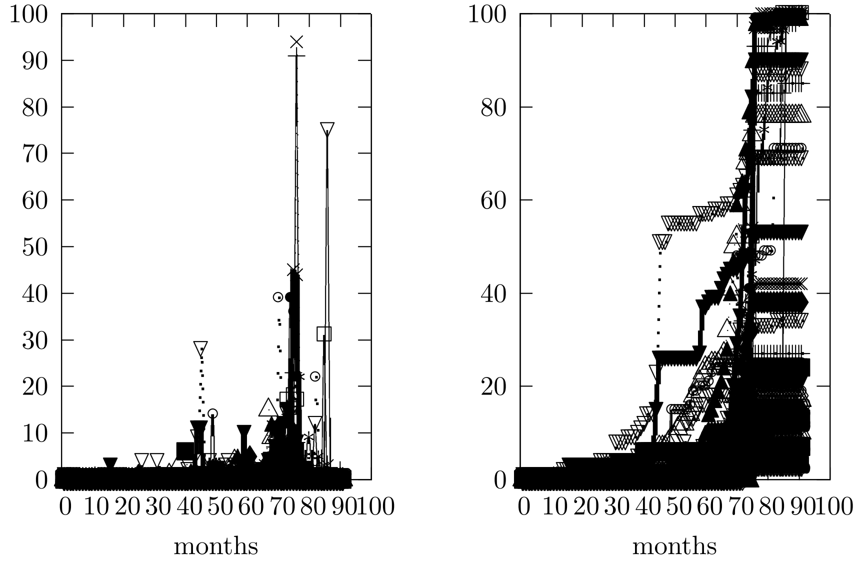

Figure 2: Temporal distributions of tags by counting (a) a number of resources over time, and (b) a cumulative number of resources over time

Moreover, during analyzing the higher pitches, we have realized that some users have exploited a certain tag to many resources at the same time. It has become even more difficult to consider the propagation effects. Thus, we have decided that the number of users (i.e., Fig. 3) is the more important (precise) factor than the number of resources (i.e., Fig. 2) is.

After removing trivial tags, we have measured inducibility by pairing all possible combinations among the selected tags. Table 3 shows directed inducibility between two arbitrary tags in the form of an asymmetric matrix.

5 Discussion

In this paper, we have focused on understanding how information is propagated via social media (particularly, folksonomy). Two heuristics (i.e., Equ. 4 and Equ. 5) have been designed to build the social pulses by simply counting the number of taggings.

Here, we want to list discuss several important issues on understanding the propagation patterns.

[4] http://flickr.com/
[5] http://ke.yu.ac.kr/wiki/index.php/OnlineNeologism

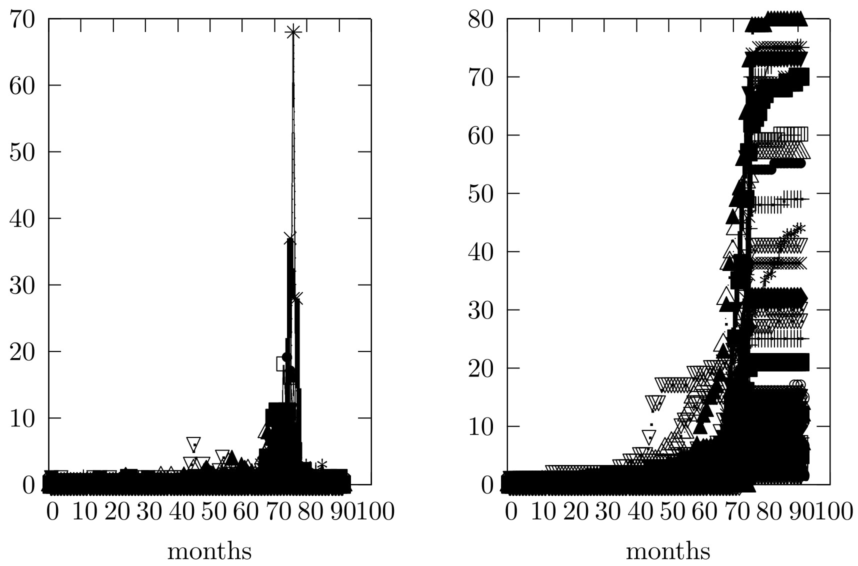

Figure 3: Temporal distributions of tags by counting (a) a number of users over time, and (b) a cumulative number of users over time

5.1 Pitch as propagation speed

Give a social pulse, pitch has been measured in a certain time moment. In this work, we have assumed that the pitch indicates the propagation speed of the corresponding tag. It means that with the higher pitch, there might be a certain event and news that more people have received. Thus, we can understand how an event can be influenced on online users (and social media).

As a research limitation, Flickr API is not providing a tagging timing but a photo uploading timing. We have regarded that both are same. Thus, there might be an error (e.g., temporal difference) if a user has added a new tag to some photos which have been already uploaded, and also modify the tags.

5.2 Discovering various relationships from a folksonomy

By collecting a set of social pulses, we can measure various relationships (e.g., similarity and distance) [Jung 2012b]. In this work, we have proposed a new measurement, called "inducibility," indicating a temporal closeness. Table 3 has been computed by Equ. 8. More importantly, this measurement is directive, meaning that $\mathcal{I}_{t_i \to t_j} \neq \mathcal{I}_{t_j \to t_i}$.

Table 3: Inducibility measurement among the sampled Korean tags by Equ. 8

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}	t_{15}	t_{16}
t_1	-	1	0	1	0	0	1	1	0	0	0	0	0	0	0	0
t_2	0.25	-	0	0.5	0	0.5	0.5	0.5	0	0	0	0.5	0	0	1	1
t_3	0.17	0.33	-	0	0	0	0	0	0	0	0	0	0	0	0	0
t_4	1	0.33	0.5	-	0	0	1	1	0	0	0	0.5	0	0	1	0
t_5	0.5	0.1	0.08	0.5	-	0	0	0	0	0	0	0	0	0	0	0
t_6	1	1	0.33	1	0.5	-	1	1	0	0	0	0	0	0	1	0
t_7	0.5	0.33	0.2	0.5	0.13	0.33	-	1	0	0	0	0.5	0	0	1	1
t_8	1	1	0.33	1	0.5	1	0.33	-	0	0	0	0	0	0.5	1	0
t_9	0.06	0.08	0.09	0.5	0.05	0.33	0.07	0.13	-	0	0	0	0	0	0	0
t_{10}	0.5	0.1	0.08	0.5	1	0.5	0.13	0.5	0.05	-	0	0	0	0	0	0
t_{11}	1	0.11	0.09	1	0.5	1	0.14	1	0.05	0.5	-	0	0	0	0	1
t_{12}	0.5	0.25	0.5	1	1	0.5	0.17	0.5	0.13	1	0.5	-	0	0.08	0	0
t_{13}	0.5	0.2	0.14	0.5	0.17	0.5	0.33	0.5	0.06	0.17	0.2	0.17	-	0	0	0
t_{14}	0.17	0.33	1	0.5	0.08	0.33	0.2	0.5	0.11	0.08	0.09	0.5	0.14	-	0	0
t_{15}	1	0.5	0.5	1	0.5	1	0.25	1	0.5	0.5	1	0.5	0.2	0.5	-	0.2
t_{16}	1	0.5	0.5	1	0.5	1	0.33	1	0.25	0.5	1	0.5	1	0.5	1	-

Again, due to the drawback by Flickr API, we have some problem on obtaining the correct time and computing the inducibility.

5.3 Potential applications

Since most of collaborative workplaces, e.g., virtual enterprises, have employed online social media, there should be an efficient platform and methodology for monitoring the information propagation within a given workplace [Jung 2011c, Jung 2011b, Jung 2010a]. Thus, we are expecting there will be a number of potential domains that the proposed work can be applied, as follows.

- detecting social events (or measuring propagation patterns)
- knowledge management system in various organizations

6 Conclusion and Future Work

Understanding how the information can be propagated is an important task on many applications using online social media. Also, in terms of shared understanding (e.g., contextual synchronization [Jung 2011a]) among multiple users, it is a critical topic how online users are cognitively responding.

In this paper, we have formulated social pulses by taking into account the number of tagging actions on folksonomies. They are similar to a set of discrete time series dataset where we can extract several main features, e.g., pitches and frequencies. Most importantly, we have proposed an inducibility relationship between two social pulses (in fact, two corresponding tags).

Acknowledgments

This work was supported by the Korea Ministry of Knowledge Economy (MKE) under Grant No.2009-S-034-01.

References

[Cha et al. 2009] Cha, M., Mislove, A., Gummadi, K. P.: A measurement-driven analysis of information propagation in the flickr social network. Proceedings of the 18th International Conference on World Wide Web, WWW 2009, Madrid, Spain, April 20-24; 721–730; ACM, 2009.

[Jung 2010a] Jung, J. J.: Integrating social networks for context fusion in mobile service platforms. Journal of Universal Computer Science, 16(15), 2099–2110.

[Jung 2010b] Jung, J. J.: On sustainability of context-aware services among heterogeneous smart spaces. Journal of Universal Computer Science, 16(13), 1745–1760.

[Jung 2010c] Jung, J. J.: Reusing Ontology Mappings for Query Segmentation and Routing in Semantic Peer-to-Peer Environment, Information Sciences, 180(17), 3248–3257.

[Jung 2011a] Jung, J. J.: Boosting social collaborations based on contextual synchronization: An empirical study. Expert Systems with Applications, 38(5), 4809–4815.

[Jung 2011b] Jung, J. J.: Service chain-based business alliance formation in service-oriented architecture. Expert Systems with Applications, 38(3), 2206–2211.

[Jung 2011c] Jung, J. J.: Ubiquitous conference management system for mobile recommendation services based on mobilizing social networks: A case study of u-conference. Expert Systems with Applications, 38(10), 12786–12790.

[Jung 2012a] Jung, J. J.: Constraint graph-based frequent pattern updating from temporal databases. Expert Systems with Applications, 39(3), 3169–3173.

[Jung 2012b] Jung, J. J.: Discovering community of lingual practice for matching multilingual tags from folksonomies. Computer Journal, 55(3), 337–346.

[Jung 2012c] Jung, J. J.: Evolutionary approach for semantic-based query sampling in large-scale information sources. Information Sciences, 182(1), 30–39.

[Pal and Counts 2011] Pal, A., Counts, S.: Identifying topical authorities in microblogs. Proceedings of the 4th International Conference on Web Search and Web Data Mining (WSDM 2011), Hong Kong, China, February 9-12, 45–54, ACM, 2011.

[Yang and Leskovec 2010] Yang, J., Leskovec, J.: Modeling information diffusion in implicit networks. Proceedings of The 10th IEEE International Conference on Data Mining (ICDM 2010), Sydney, Australia, 14-17 December, 599–608, IEEE Computer Society, 2010.