An Analysis on Information Cascade for Detecting Influencers in Twitter

ツイッターにおけるインフルエンサー検出ための情報カスケードの解析

Geerajit Rattanaritnont, Masashi Toyoda and Masaru Kitsuregawa

**Graduate School of Information Science and Technology, The University of Tokyo**

Background

Nowadays, Twitter is one of the most popular micro-blogging services. Among its users, we believe that there is a minority of users who influence a large number of people in the network, called "Influencers"or "Influentials".

Knowing these people brings us two main benefits.
1. It will enable viral marketing in online social networks.
2. It will help us find important and insight sources of interested information.

However, not all information is diffused because of user influence. It is sometimes spread widely due to the content itself. In this study, we investigate the evolution patterns of information cascade among users to understand the characteristic of user influence in Twitter.
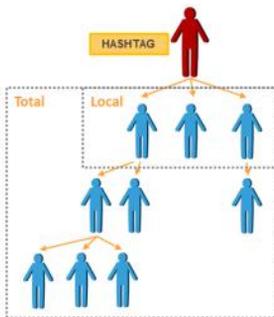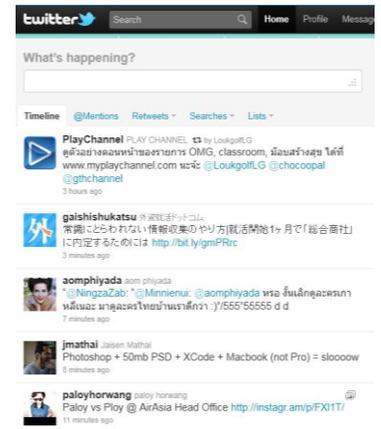
Information Cascade

Almost recent work measured user influence by using quantitative data and twitter graph such as the number of followers, retweets and mentions.

However, studying information cascade between users seems to be more relevant to detecting the influencers. We consider two types of cascade in term of "hashtag" usage.
1. Local cascade means the number of a user's immediate followers that reposted the given hashtag after him.
2. Total cascade refers to overall number of a user's direct and indirect followers who reposted the given hashtag after him.
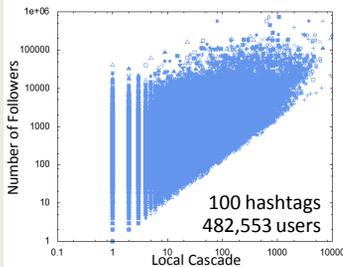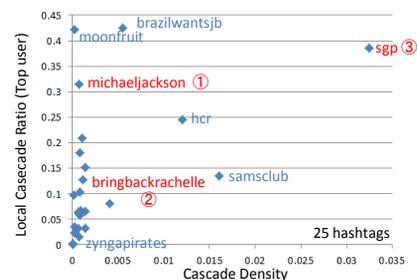
Cascade Density & Ratio

$$\text{Cascade Density} = \frac{\sum_{u \in U(h)} C(u,h)}{\frac{(|U(h)|-1)*|U(h)|}{2}}$$

$$\text{Local Cascade Ratio} = \frac{C(u,h)}{|U(h)|}$$

where
$U(h)$ is a set of users who use hashtag h
$C(u,h)$ is the number of local cascade of user u posting hashtag h

Cascade Patterns

1. **Low density – High ratio**

- Mainly be spread due to outside happenings.
- Users tend to be independent of each other.
- Influencers are elite users.

Ex. #michaeljackson
- Michael Jackson died on June 25, 2009. (Peak day)
- Density = 0.718x10^{-3}
- Ratio = 0.31

2. **Low density – Low ratio**

- Users tend to be independent of each other.
- Influencers are specifically related to the topic.

Ex. #bringbackrachelle
- Rachelle Lefevre, playing Victoria in New Moon, was announced to be replaced in Eclipse on July 28, 2009. (One day before peak)
- Density = 1.151x10^{-3}
- Ratio = 0.13

3. **High density – High ratio**

- Likely to be spread due to internal sources.
- Users tend to form a community.
- Difficult to detect influencers or they do not exist.

Ex. #sgp
- Smart Girl Politics is an organization of American women who subscribe to conservative political beliefs.
- Density = 32.448x10^{-3}
- Ratio = 0.39