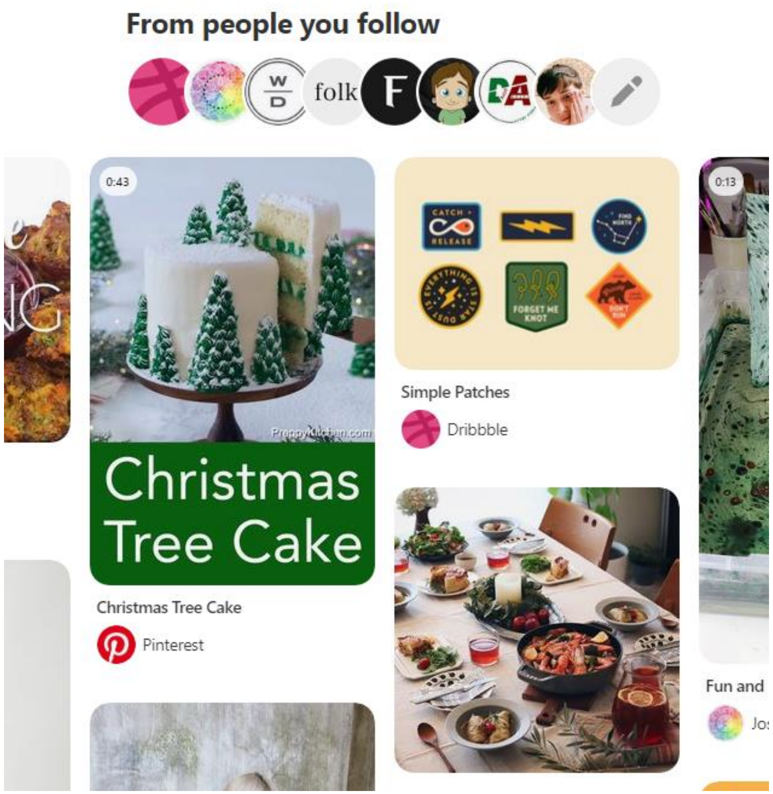
A Model for Social Influence on Topic of Users Contents Generating Behavior

Mirai Igarashi

(Work with Nobuhiko Terui, P.K. Kannan, Kunpeng Zhang)

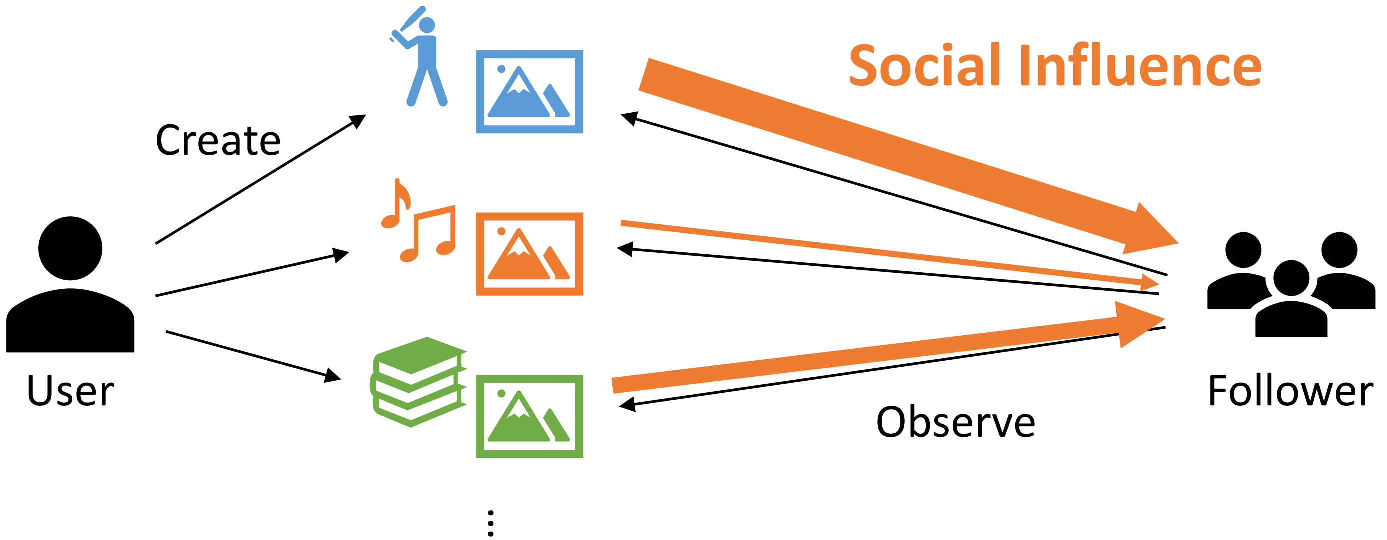
Social influence

Social Media (Pinterest)
-> **User-Generated-Contents**



Media Streaming (Apple Music)
-> **Observed adoption**

Social influence on social media



RQ

Do social influences vary for topics of user-generated-contents?

In this research, we propose a topic model for capturing topic-specific influence

How to measure social influence?

Literature consider SI from **qualitative** perspective

- **Volume**
  - number of word-of-mouth (Godes & Mayzlin, 2004)
  - number of peers who adopted the product (Bollinger & Gillingham, 2012)
  - length of review text (Lu, Wu & Tseng, 2018)
- **Valence**
  - average rating of product reviews (Moe, Trusov & Smith, 2011)
  - emotions of review text (Wu et al., 2015)
- **Variance**
  - variance of rating (Moe, Trusov & Smith, 2011)

How to measure social influence?

Contents perspectives have recently attracted attention

- **Product characteristics**
  - technology vs fashion-related products (Wang, Aribarg & Atchadé, 2013)
  - high and low utilitarian products (Schulze, Schöler & Skiera, 2014)
  - hedonic vs functional products (Park et al., 2018)

Contribution

- focus on **contents generating behavior** on social media
  -> above, they focus on consumers' purchase behavior
- estimate influence and **its dimensions** simultaneously
  -> above, we need to specify dimensions before analysis

Where do social influences occur?

Situation where SI occur is also important

- **Structure of social network** → don't consider (yet)
  - degree distribution (Dover, Goldenberg & Shapira, 2012)
  - cluster coefficient (Choi, Kim & Lee, 2010)
  - centrality (Cho, Wang & Lee, 2012; Susarla, Oh & Tan, 2012)
- **Tie strength** → only friend network
  - weak vs strong ties (Granovetter, 1973)
  - friend vs community network (Zhang & Godes, 2018; Ameri, Honka & Xie, 2019)
- **Multiple network** → only single network
  - networks of economic, social, etc. (Chen, Van der Lans & Phan, 2017)

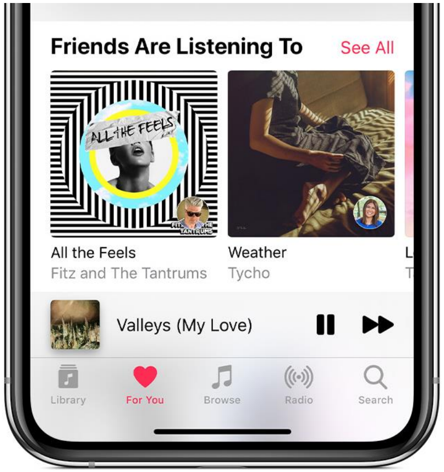
We don't consider other situation and compare them (further research)

-> ex. topic-specific influence on weak ties

Managerial impact

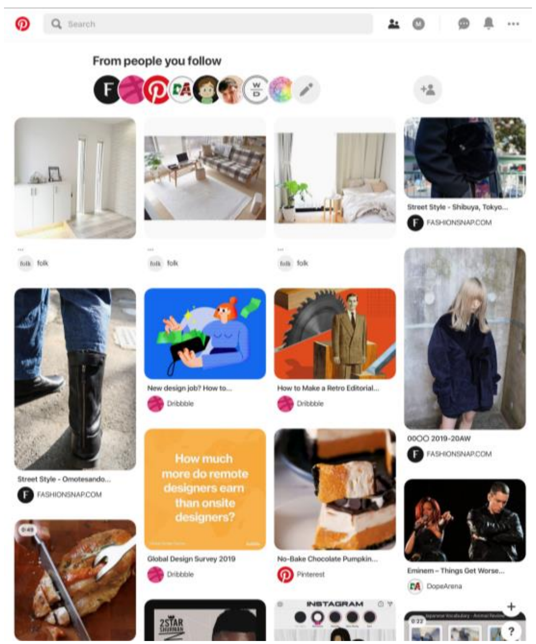
- **Digital advertisement**
  - Current methods consider overall social influence
  - This study can take into account topic-specific influence of ad receivers, and it makes ad delivery strategy efficient

- **Recommendation system**
  - Some media streaming services (e.g. AppleMusic) incorporate recommendation system within limited space
  - If we know topics on which the focal user have strong influence, we can make use of the limited space

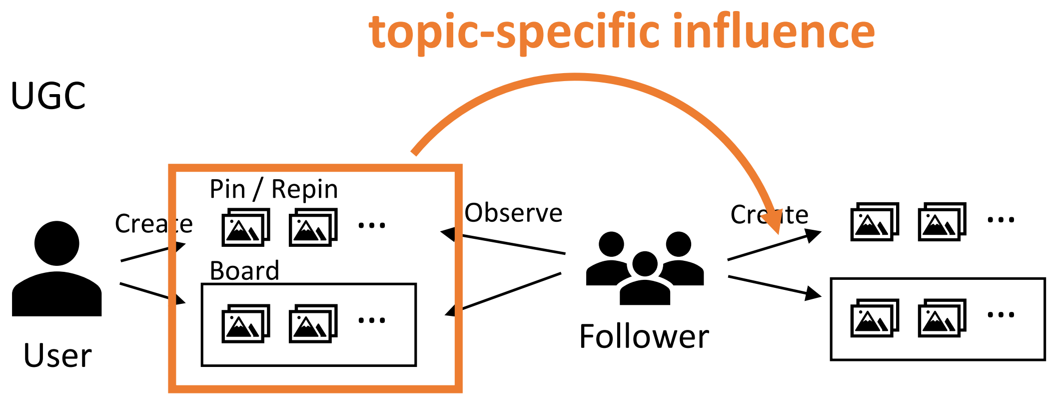


Pinterest

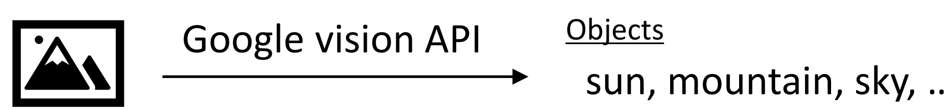
Dataset



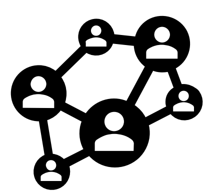
UGC



Objects in the photo



Social network (static)



Following – Follower network

Model specification

d: user t: time period (weekly) n: object index k: topic

$$\text{topic} \quad z_{dtn} \sim Multi\left(\theta'_{dt}\right), \qquad \theta'_{dtk} = \frac{\exp(\theta_{dtk})}{\sum_{k'} \exp(\theta_{dtk'})}$$

$$\text{object} \quad w_{dtn} | z_{dtn} = k \sim Multi\left(\phi_k\right)$$

$$\theta_{dt} = \underbrace{\theta_{dt-1} \cdot \alpha}_{\text{AR term}} + \underbrace{\sum_{f \in F_d} \bar{z}_{ft-1} \cdot \beta_{df}}_{\text{social influence}} + \underbrace{\gamma_d + \delta_t}_{\text{fixed effects}} + \epsilon_{dt}, \qquad \epsilon_{dt} \sim \underbrace{N(0, \Sigma)}_{\text{correlated topic model}}$$

Identification problem

$$\theta_{dt} = \theta_{dt-1} \cdot \alpha + \sum_{f \in F_d} \bar{z}_{ft-1} \cdot \beta_{df} + \gamma_d + \delta_t + \epsilon_{dt}, \qquad \epsilon_{dt} \sim N(0, \Sigma)$$

- **Homophily**

  -> Individuals tend to connect with those who are similar to them

  -> There is positively correlation between behaviors of them and peers

Some methods for avoiding homophily problem

- Fixed effect (Nair, Manchanda & Bhatia, 2010)
- Heterogeneity (Hartmann, 2010)
- Co-evolution model (Bhattacharya et al., 2019)

Identification problem

$$\theta_{dt} = \theta_{dt-1} \cdot \alpha + \sum_{f \in F_d} \bar{z}_{ft-1} \cdot \beta_{df} + \gamma_d + \delta_t + \epsilon_{dt}, \qquad \epsilon_{dt} \sim N(0, \Sigma)$$

- **Simultaneity**

-> Agents and peers are affected by each other

-> It causes biased estimation because of correlation between explanatory variables and error term

Some methods for avoiding simultaneity problem

- Instrument variables (Nair, Manchanda & Bhatia, 2010)
- Modeling equilibrium (Hartmann, 2010)
- Field experiment (Wang, Aribarg & Atchadé, 2013)

In this research, we assume **sparsity** for SI coefficients
(Trusov, Bodapati & Bucklin, 2010)

$$\beta_{df} = \beta_d \times \zeta_{df}, \qquad \zeta_{df} = \begin{cases} 1 & \text{with prob. } p_d \\ 0 & \text{with prob. } 1 - p_d \end{cases}$$

Identification problem

$$\theta_{dt} = \theta_{dt-1} \cdot \alpha + \sum_{f \in F_d} \bar{z}_{ft-1} \cdot \beta_{df} + \gamma_d + \delta_t + \epsilon_{dt}, \qquad \epsilon_{dt} \sim N(0, \Sigma)$$

- **Unobserved variables**

  -> ex. common advertising and time events for multiple agents

  Some methods for avoiding simultaneity problem

  - Propensity score matching (Aral, Muchnik & Sundararajan, 2009)
  - Fixed effect (Nair, Manchanda & Bhatia, 2010)

Estimation

No conjugacy

$$z_{dn} \sim Multi\left(softmax\left(\theta_d\right)\right) \quad \Longleftrightarrow \quad \theta_d \sim N(\mu, \Sigma)$$

Gibbs sampling with **Pólya-Gamma data augmentation** (Polman et al., 2013)

$$\begin{aligned}
p(z_{dn}|\theta_d) &= \left[\prod_{n=1}^{N_d}\left(\frac{\exp(\theta_{d1})}{\sum_{k'}\exp(\theta_{dk'})}\right)^{N_{d1}}\cdots\left(\frac{\exp(\theta_{dK})}{\sum_{k'}\exp(\theta_{dk'})}\right)^{N_{dK}}\right] \\
&\propto \frac{\exp(\psi_{dk})^{N_{dk}}}{(1+\exp(\psi_{dk}))^{N_d}}, \qquad \psi_{dk} = \theta_{dk} - \log\sum_{k'\neq k}\exp(\theta_{dk'}) \\
&= 2^{-N_d}\exp\left((N_{dk} - N_d/2)\psi_{dk}\right)\int_0^\infty \exp(-\omega_{dk}\psi_{dk}^2/2)p(\omega_{dk})d\omega_{dk}, \qquad \omega_{dk} \sim PG(N_d, 0)
\end{aligned}$$

In addition to above, we use **Forward-Filtering Backward-Sampling** to take into account time-dependences of θ_{dt}

Conclusion

- Conclusion
  - The proposed model captures social influence **varying for topics of UGC** on social media
  - Unlike existing research, it is **not necessary to determine the dimensions** social influences vary
    -> It can be applied for various digital marketing (ad, recommend)

- Next step
  - Real data analysis
  - Validation of estimated influence (simulation or prediction)