Cross-Site Prediction on Social Influence for Cold-Start Users in Online Social Networks

Qingyuan Gong, Yang Chen, Xinlei He (Fudan University), Yu Xiao (Aalto University), Pan Hui (University of Helsinki / HKUST), Xin Wang (Fudan University), Xiaoming Fu (University of Göttingen)

gongqingyuan@fudan.edu.cn

Selected by the Journal Track of WWW’22
Online Social Networks (OSNs)

Social Connection
Information Diffusion
Immersive Experience
People Are Using Cross-Site Social Services

• Multiple OSNs provide various social services
• Users register accounts on multiple OSN and utilize the services at the same time
• Behaviors of users on different OSNs are different because of the corresponding services, but still correlated
Importance of Social Influence


People tend to follow the main stream to make decision

Users who generate high influence are not always the very popular users

The more people show the behavior, the larger impact it will be

Echos of information will attract more attention
Existing Metrics for Social Influence

- Number of followers
- PageRank value of the nodes in social network
- Total number of "Retweets/Likes" received by UGCs (User-Generated Contents)
- H-index value of the number of "Retweets/Likes" received by UGCs

**Definition of High Influential Users**

\[
\text{The user } u \text{ is} \begin{cases} 
\text{High influential user} & (I_u \geq t) \\
\text{Ordinary user} & (I_u < t) 
\end{cases}
\]
“Cold-start” Users in OSNs

- Cold-start users often exist in emerging OSNs
- The Medium account belonging to the same Twitter user only shows the basic information, generating fewer activities

<table>
<thead>
<tr>
<th>Medium</th>
<th>Twitter</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Profile:</strong></td>
<td><strong>Profile:</strong></td>
</tr>
<tr>
<td>Name: AlwaysPeace</td>
<td>Name: Me andering</td>
</tr>
<tr>
<td>Bio: Positive energy brings positive change.</td>
<td>Bio: No reconciliation without truth. Living on lands not clearly a part of a specific nation.</td>
</tr>
<tr>
<td>#Followers: 62</td>
<td>#Followers: 614</td>
</tr>
<tr>
<td>#Followings: 158</td>
<td>#Followings: 1613</td>
</tr>
</tbody>
</table>

**Medium Story:** Null

https://medium.com/@zz2aa

**Twitter**

https://twitter.com/zz2aa
Correlation Analysis on Social Influence across OSNs

194,850 pairs of Medium and Twitter accounts belonging to one user

Users' social influence on Twitter and on Medium cannot be considered as equivalent, but are positively correlated at a weak level

<table>
<thead>
<tr>
<th>Social influence metrics on Medium</th>
<th>Social influence metric on Twitter</th>
<th>Spearman's Rank Correlation Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of “Likes” received</td>
<td>Number of retweets received</td>
<td>0.231</td>
</tr>
<tr>
<td>H-index value of the number of “Likes” received</td>
<td></td>
<td>0.232</td>
</tr>
</tbody>
</table>
Comparison on Typical Twitter Behavior between Influential and Ordinary Users on Medium

- **Number of followers**
  - Less than 100 for ordinary users
  - More than one thousand for 20% high influential users

- **Number of “Likes” received by original tweets**
  - Less than 5000 for ordinary users
  - More than one thousand for 20% high influential users

- **Number of retweets received by original tweets**
  - Less than 1 for ordinary users
  - Between [1,100] for 40% high influential users
Cross-Site Prediction on Social Influence for Cold-Start Users

Feature Extraction on Dominant OSN

- **Descriptive Feature Extraction**
  - Username:...
  - Location: ...
  - Number of Following/Followers: ...
  - Indexes on activities: ...

- **Dynamic Feature Extraction**
  - Activity Seq. Construction
  - Activity element
  - Activity Seq. Analysis

Prediction of Influential users on Emerging OSN

- Feature set
  - Descriptive features
  - Dynamic features
  - Decision Maker
  - Influential user
  - Ordinary user

User’s Twitter Homepage
Cross-Site Prediction on Social Influence for Cold-Start Users

- Labels showing whether this user is influential user on Medium
- Profile and UGCs on Twitter
- LSTM network models tweet publishing sequence and produces the tweet sequence features
- Four categories of features extracted from account information
- Decision maker aggregates the five parts of features
Evaluation Results

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Social influence metrics on Medium</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Total number of “Likes” received</td>
<td>0.848</td>
</tr>
<tr>
<td>2</td>
<td>H-index value of the number of “Likes” received</td>
<td>0.852</td>
</tr>
</tbody>
</table>

- Metric
  - AUC values measure the probability that the decision maker will rank the influential user higher than the ordinary user

- Results
  - The proposed model achieves high AUC values for two different definitions of influential users on Medium
  - Using the public data on Twitter, the decision make can predict the influential users on Medium
Conclusions

• Social influence of users are correlated on multiple OSNs
• Whether “cold-start” users on an emerging OSN will become influential users can be predicted using the UGCs on dominate OSNs
• Modeling users' sequential behaviors is effective to predict users' properties
Future Work

• We plan to investigate more emerging OSNs for the cross-site user behavior research

• We will further study the evolvement of a user's social influence on emerging OSNs

• We wish to study different types of user classification problems from the perspective of cross-site user behavior analysis
Thank you for your listening!

gongqingyuan@fudan.edu.cn
https://gongqingyuan.wordpress.com/