Modeling and Mining Information Popularity Online

Flavio Figueiredo Jussara M. Almeida
Computer Science Department - Universidade Federal de Minas Gerais, Brazil
{flaviov, jussara}@dcc.ufmg.br

1. INTRODUCTION

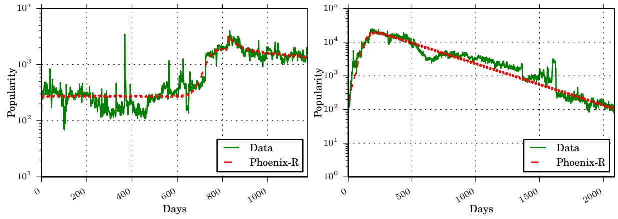
Nowadays, there is an unprecedented amount of user generated content being produced online. This fact is one of the driving forces of what is know as the *social media* phenomenon. Social media shifted how information is produced and propagated. While in traditional media select individuals are responsible for the production, curation and propagation of information, in the social media setting anyone can produce and share information online. One major question in this setting is: *What drives the popularity of information in social media?* This is an interesting question since even with the overload of information that accompanies this mass production of content, some pieces of information manage to attract the attention of millions of users, while the majority remain obscure.

One example of the complexity behind social media popularity is the YouTube channel of *Henri, le Chat-Noir*[1]. The first video of Henri was uploaded in 2007 and remained in obscurity for years. However, in 2012 a user of the Tumblr social network found the video and posted online[2]. Currently, the video and channels has millions of visits from a wide range of different sources (e.g., OSNs, search engines, word-of-mouth and so forth). We can use this single example to motivate our research. Important questions that we raised and approached were as follows:

What is the impact of incoming links on the popularity of online information? There are multiple forms through which users can reach content and, thus, there are multiple driving forces that may impact the popularity of information. Identifying such forces is crucial for designing more cost-effective content dissemination strategies. For instance, should a content creator invest time on perfecting the keywords describing content (for better search rankings) or focus on campaigning videos in OSNs? Our current results show that search engines and social propagation inside a service (say YouTube) are major factors in driving popularity [4].

How does information popularity evolve over time? Here, we aim at answering if there are different patterns which capture the major trends in which information popularity evolves over time. In a birds-eye-view, we found that there exists a combination of two trends governing the popularity of information. One trend consists of contents that tend to remain attractive over time with an always increasing or steady-state popularity [4]. The other, accounts for content that tend to peak in popularity for a short while, with three different popularity decay characteristics after the peak. Examples of both trends are shown in Figure 1.

How do users perceive the quality of popular and unpopular information? Most research in online popularity neglect the users perception of the information being disseminated. We studied if users tend to like very popular content or dislike very unpopular content. Based on a user based study on Mechanical Turk and selected videos from YouTube, we showed that, while users perceptions of content are highly subjective, when the majority of users like a certain YouTube video that video tends to be popular. This was interesting since it shows that there is more than social propagation to popularity. Based on our results [2] we hypothesize that videos like the Henri example would likely be less popular (regardless of OSN propagation) if their content did not appeal to users.



(a) Rock Song (b) User Dancing Video

Figure 1: Different YouTube videos as captured by our model.

Can we model and predict the future popularity of information? Two of our most recent results showed that we can model [5] the popularity of information over time and predict the future popularity [1]. Based on our previous findings that we discussed, we developed the Phoenix-R model which can capture the long term popularity evolution as showed in Figure 1. Also, we combined social network propagation and early view patterns of news media to develop prediction models with the user of machine learning tools [1]. These two results show the applicability of our results to mining tasks such as popularity prediction.

With these questions we summarize some of the work on information popularity online that we are pursuing. Our results show that this a promising and new area of research. Currently, we are working on optimizing how early can we predict popularity [3] and practical applications for models and predictions (e.g., search engine rankings or advertising).

2. REFERENCES

[1] F. Figueiredo, J. Almeida, Y. Matsubara, B. Ribeiro, and C. Faloutsos. Revisit Behavior in Social Media: The Phoenix-R Model and Discoveries. In *Proc. ECML/PKDD*, 2014.

[2] F. Figueiredo, F. Benevenuto, J. Almeida, and K. Gummadi. Does Content Determine Information Popularity in Social Media? A Case Study of YouTube Videos' and their Popularity. In *Proc. CHI.*, 2014.

[3] F. Figueiredo, F. Benevenuto, M. Gonçalves, and J. Almeida. TrendLearner: Early Prediction of Popularity Trends of User Generated Content. *Online at: http://arxiv.org/abs/1402.2351*, 2014.

[4] F. Figueiredo, F. Benevenuto, M. Gonçalves, and J. Almeida. On the Dynamics of Social Media Popularity: A YouTube Case Study. *ACM Transactions on Internet Technology*, To appear.

[5] F. Figueiredo, M. Gonçalves, and J. Almeida. Improving the Effectiveness of Content Popularity Prediction Methods using Time Series Trends. In *ECML/PKDD Discovery Challenge*, 2014.

[1] `http://www.youtube.com/user/HenriLeChatNoir`

[2] `http://knowyourmeme.com/memes/henri-le-chat-noir`