

Published in "Physica A: Statistical Mechanics and its Applications 453(): 236–241, 2016"
which should be cited to refer to this work.

Characterizing popularity dynamics of online videos

Zhuo-Ming Ren ^{a,*}, Yu-Qiang Shi ^b, Hao Liao ^{c,a}

^a Department of Physics, University of Fribourg, Chemin du Musée 3, CH-1700, Fribourg, Switzerland

^b School of Manufacturing Science and Engineering, Southwest University of Science and Technology, Mianyang 621010, PR China

^c Guangdong Province Key Laboratory of Popular High Performance Computers, College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, PR China

H I G H L I G H T S

- A temporal analysis of the popularity dynamics in two online video-provided websites.
 - Dynamics of the online video popularity can be characterized by the burst behaviors.
 - The burst behaviors typically occur in the early life span of videos.
 - Lately the online video popularity restricts to the classic preferential mechanism.
-

Online popularity has a major impact on videos, music, news and other contexts in online systems. Characterizing online popularity dynamics is nature to explain the observed properties in terms of the already acquired popularity of each individual. In this paper, we provide a quantitative, large scale, temporal analysis of the popularity dynamics in two online video-provided websites, namely MovieLens and Netflix. The two collected data sets contain over 100 million records and even span a decade. We characterize that the popularity dynamics of online videos evolve over time, and find that the dynamics of the online video popularity can be characterized by the burst behaviors, typically occurring in the early life span of a video, and later restricting to the classic preferential popularity increase mechanism.

1. Introduction

The online individual behaviors for videos, music, news and other contexts are implicated in the online popularity whose breadth surpasses individual awareness [1–3]. The online behaviors can predict the popularity trend, and then the online services such as Netflix, Youtube, Facebook would benefit greatly from an in-depth understanding of the popularity trend to optimize strategies for update and recommendation [4,5]. For example, in Netflix, if the evolution of the video popularity is predictable in advance, the Netflix service can replicate more copies of videos with increasing popularity and reduce replication of videos losing popularity to make space for the hot videos. Meanwhile, the online behavior information produces a big amount of time stamped data, making it possible to study the dynamics of the online popularity and how it evolves over time on a global scale [2,6]. From a theoretical point of view, the massive amount of available data from these online services provides an unprecedented opportunity to understand the online popularity dynamics. Hence, characterizing

* Corresponding author.

E-mail addresses: zhuomingren@gmail.com (Z.-M. Ren), hao.liao@unifr.ch (H. Liao).

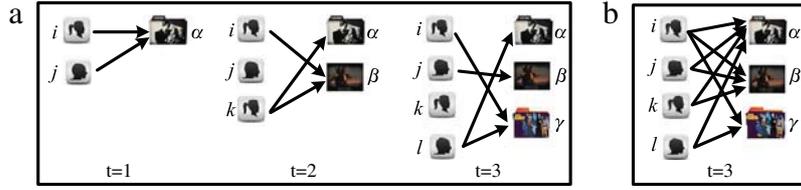


Fig. 1. An example of the relationship between user and video which represents a bipartite network with the temporal information. (a) The temporal bipartite networks. The new user and the new video enter the bipartite network as the time varying. At the time point $t = 2$ and $t = 3$, a new video and a new user are added to the network respectively. The new video is more popular than the old ones, and is selected by the two users, but the old video is selected by one user. (b) shows that historical snapshot until the time $t = 3$. The older video α has been selected four times. The old one β has been selected three times. The new one γ has been selected two times. The example indicates that the older the video is, the more accumulative but the less temporal popular it will be.

the online popularity dynamics is necessary for explaining the observed features in terms of the already acquired online popularity of each individual and predicting the popularity trend. The behaviors of the online popularity dynamics have been widely studied in the literature focused on videos [7–9], music [10], news [11], and other online social collective dynamics [12,13]. Cha et al. [14] found that a high linear correlation existed between the number of video views on early days and later days in YouTube. Borghol et al. [15] performed an empirical study to measure the popularity of videos, and argued a strong linear rich-get-richer behavior with the number of previous views as the most important factor [16]. Shen et al. [17] employed the reinforcement Poisson mechanism documenting the well-known rich-get-richer phenomenon to model the popularity dynamics. Compared to a stronger presence of the rich-get-richer phenomenon, Vasconcelos et al. [18] observed a lower correlation between the early and late popularities in Foursquare. Chen et al. [19] considered the lifetime perspective of video popularity, and found that the relative popularity of a video was not only dependent on its age, but also the type of the video. Specifically, the online video of popularity is by no means restricted to rich-get-richer behaviors. It can depend on exogenous attributes of the online contexts which make popularity dynamics suitable for applications in many different contexts in the online systems [20]. Furthermore, Ratkiewicz et al. [21] examined the popularity of Wikipedia topics and Web pages and presented a model that combines the classic preferential popularity increase mechanism with the occurrence of random popularity shifted capturing the influence of exogenous factors on online popularity. There also have some works based on the exogenous attributes to result in changes to modeling the popularity dynamics in different online contexts such as user-generated videos [22], the citation of scientific release [23,24], and the activity of scientists [25]. While, few works are centralized when the online contexts emerge suddenly popularly in their life span. In this paper, we conduct an in-depth study on the online popularity dynamics in two online video-provided websites, namely MovieLens [26] and Netflix [27], who both invite users to rate videos. The two big time stamped data sets contain over 100 million rating records and even span a decade, which provide an opportunity to understand the dynamics of online popularity and how they evolve over time on a global scale. We characterize that how video rating behaviors of the online popularity dynamics evolve over time, and find that the dynamics of rating popularity are characterized by the burst behaviors (i.e. far exceed preferential popularity increase). Typically, the presences of rating burst occur in the early life span of a video, and later restrict to the classic preferential popularity increase mechanism.

2. Model and method

Here, we use bipartite networks with temporal information to represent the online rating systems which include the set of users (denoted by $U(t)$), the set of videos (denoted by $O(t)$) and the records between users and videos (denoted by $L(t)$). A link in the bipartite network connecting user i and video α represents a historical record $l_{i\alpha} (\in L)$. If a video has been chosen by more users than others, we could see this video is more popular than others. Thus, the popularity of the video $k_{\alpha}(t)$ is defined as the number of rating records which the video α was received at time t . We give a simple example in Fig. 1 to show how to construct a temporal bipartite network. The new video gets more temporal popularity than the old videos (at time point $t = 3$, the older video α has 1 records, the old video β has 1 records, and the new video γ has 2 records) as displayed in Fig. 1(a). While as shown in Fig. 1(b), the older videos get more accumulative popularity than the newer objects' (until time $t = 3$, the older video α has 4 records, the old video β has 3 records, and the new γ has 2 records).

Preferential attachment [28] focuses on the fact that the rich gets richer, which indicates that the popular videos get more popularity. While, as the time varying, there always emerge new popular videos, and some old popular videos perform less popularly than the new popular videos. The fitness model [29] describes that the ability between the new node and the old node to compete for popularity. Numerous examples [21,23,24] convincingly indicate that in real systems the new one is easy to get more popularity in the early life span. For example, the online rating systems provide services or products (i.e., Netflix, MovieLens for videos, Amazon for books/other products, alibaba, Ebay e-business platform for goods). The users could prefer to pay more attention at the appearance of new videos, new news or new goods in the online rating systems, then the new products immediately get a lot of attention and become popular. The burst is considered as accumulations of a large number of rapidly occurring events during short time intervals which is observed in many real systems such as human dynamics, citation dynamics and other online collective dynamics [30]. To measure the burst of production, Goh et al. [31]

Table 1
Topological features of the real network data sets considered.

Data sets	Videos	Users	Records	Start date	End date
MovieLens	10,681	69,878	10,000,054	Jan 09, 1995	Jan 05, 2009
Netflix	17,770	480,189	100,480,507	Nov 11, 1999	Dec 31, 2005

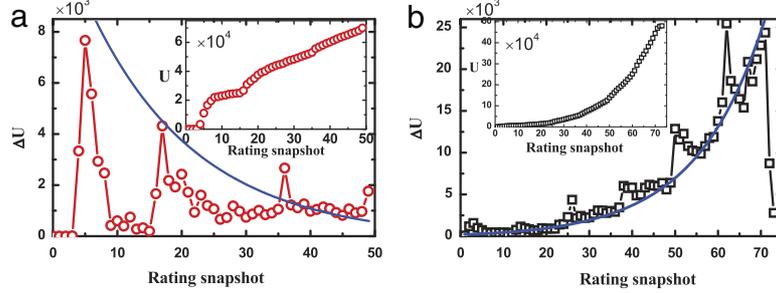


Fig. 2. (Color online) Illustration of the growth of users ΔU and the number of users U which evolve with rating snapshots in the data sets. (a) In MovieLens, we could see that the ΔU fits $10^{4.1-0.026t}$. (b) In Netflix, we could see that the ΔU fits $10^{2.5+0.027t}$.

designed the burstiness coefficient $B = (\delta_\tau - \mu_\tau)/(\delta_\tau + \mu_\tau)$, where δ_τ , and μ_τ are the mean and standard deviation of the time between two consecutive events. Another definition is that the relative rate $\Delta k_i(t+1)/k_i(t)$ is used to analyze the burst behaviors of popularity dynamics in Wikipedia and the Web space [22], the cite networks [24] and the online social networks [20,32]. As we know, according to the preferential attachment (PA), the temporal popularity that the video i owned at time $t+1$ is predicted by $\Delta k_i(t+1)_{(PA)} = \Delta L(t+1)k_i(t)/L(t)$, where $\Delta L(t+1) = \sum_j \Delta k_j(t+1)$ the temporal popularity of all videos is added to the online system at the time t , $L(t) = \sum_j k_j(t)$ the popularity is owned by the total videos until time t . While, in the reality, the video i actually owns the popularity $\Delta k_i(t+1)$ at time $t+1$. In order to gauge quantitatively the gap between the actual and the expected with the time evolving, we consider that the ratio between the actual and the expected with the time evolving which defines as the relevance [23],

$$R_i(t+1) = \frac{\Delta k_i(t+1)}{\Delta k_i(t+1)_{(PA)}} = \frac{\Delta k_i(t+1)L(t)}{k_i(t)\Delta L(t+1)}. \quad (1)$$

This expression is obviously understood that if $R_i(t) = 0$, the video i gets no temporal popularity at time t (i.e. $\Delta k_i(t) = 0$). If $0 < R_i(t) < 1$, the video i gets temporal popularity less than the expected at time t . If $R_i(t) \geq 1$, the video i gets temporal popularity more than (or reach) the expected at time t . Especially, when $R_i(t) \gg 1$, the video i gets temporal popularity much more than the expected at time t . Consequently, we can observe that at time t , the popularity of video i rises much faster than that of the system, and we can claim that the popularity of video i emerges the burst at the time t .

3. Analysis of real data sets

3.1. Materials

We test two representative real networks which are sampled from MovieLens [26] and Netflix [27]. MovieLens is an online video recommendation web site, who invites users to rate videos. Netflix web site has DVD rental service and users can also vote on the videos. The rating record for both MovieLens and Netflix is from one (i.e., worst) to five (i.e., best). Based on the user historical records, we can construct a user-video bipartite network. If a user i selects a video α and rates it, a link between user i and video α would be established as shown in Fig. 1. The relevant topological features of the two different networks are summarized in Table 1. From these data sets, we separate the time-aggregated networks into different rating snapshots with the interval of the time unit which are one month in Netflix, 102 days in MovieLens. MovieLens and Netflix have 50 and 74 rating snapshots respectively.

Fig. 2 shows that the growth of users ΔU with the time in MovieLens and Netflix. The growth of users in MovieLens follows $10^{4.1-0.026t}$ in Fig. 2(a) which suggests the growth of users is in the decline trend. While, as shown in Fig. 2(b), the growth of users in Netflix follows $10^{2.5+0.027t}$, and we could see that the growth of users is in the rise trend. Fig. 3 shows that the relative growth of links (i.e. ratings) $\Delta L/L$ decreases in a sharp period, and later, the decay becomes very slow, then maintaining the stationary value.

3.2. The popularity dynamics of online videos

Firstly, we investigate the distribution of $R_i(t) = 0$ between the rating snapshots (i.e. time t) and the release date of the video (t_i). The $R_i(t) = 0$ means $\Delta k_i(t) = 0$ according to Eq. (1) and corresponds to express that the video i wins no

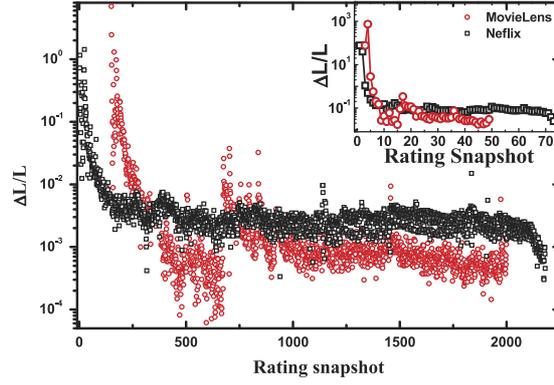


Fig. 3. (Color online) The relative variation of increments $\Delta L/L$ in the time unit in MovieLens and Netflix. We can see that $\Delta L/L$ descends slowly after a sharp fall in the early time. The each rating snapshot contains an interval of the time unit which is one day. Inset plot shows that the number of the rating snapshots are 50 and 74 in MovieLens and Netflix respectively.

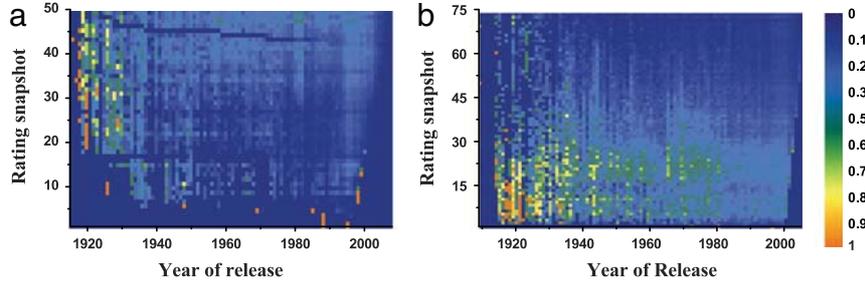


Fig. 4. (Color online) The heat map of the distribution of the $R_i(t) = 0$ between the rating snapshots (time t) and release date of the videos (t_i). (a) In the MovieLens and (b) in the Netflix. The $R_i(t) = 0$ means the $\Delta k_i(t) = 0$ according to Eq. (1) and corresponds to express that the video i wins no new popularity between time $t - 1$ and t . As we can see in (a) and (b), the more early the video release year is, the more percent of the $R_i(t) = 0$ is.

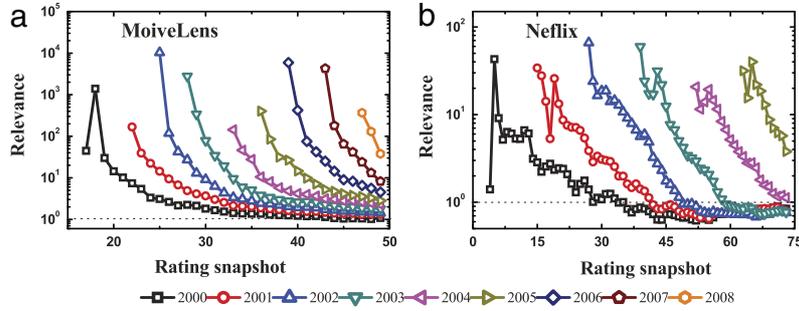


Fig. 5. (Color online) The relevance values of videos which are released in different years evolve with time (i.e. the rating snapshot). (a) The year from 2000 to 2008 in the MovieLens and (b) the year from 2000 to 2005 in the Netflix. Here, X -axis is the rating snapshot. It is notable that $R_i(t) \gg 1$ which takes place in the early life span and manifests that the burst behaviors occur within the early life span of videos.

new popularity between time $t - 1$ and t . That is to say, in a time unit, the video has been rated by no people. As shown in Fig. 4, we limit the analysis to the relative $R_i(t) = 0$ between the release year and the rating snapshots. In MovieLens, the percentage of the relative $R_i(t) = 0$ is close to 1 when the videos are released in the early time, and their rating snapshots outnumber 20. While as given in Fig. 4(b), the percentage of the relative $R_i(t) = 0$ is inclined to decrease when the year of release increases in Netflix. In general, the earlier the video release year is, the more the percent of the $R_i(t) = 0$ is, which indicates the videos are released earlier, their popularity are more likely to keep stably (i.e. $\Delta k_i(t) = 0$).

The relevance values of videos in MovieLens and Netflix which are released in different years evolving time as presented in Fig. 5(a)–(b). We select the videos which are released from year 2000 to 2008 in MovieLens and from year 2000 to 2005 in Netflix. Almost all relevance values of videos experience an exponential decay near the beginning of their life which indicates that the popularity of videos occurs a burst, but then the scale of burst decays very fast even then vanishes after a few time units. After the burst, we could observe that the relevance decays very slowly or even produce a stationary relevance value

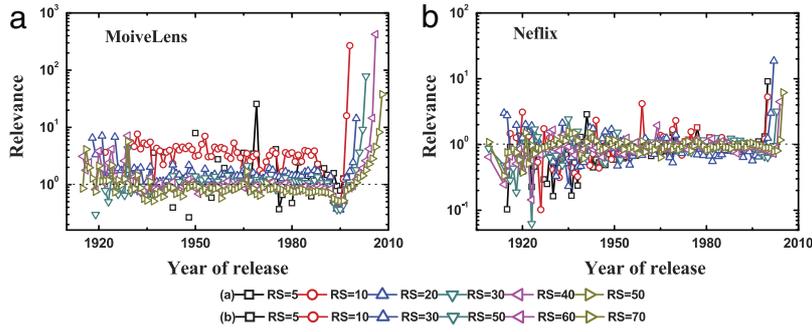


Fig. 6. (Color online) Illustration of the video relevance values $R_i(t)$ in different rating snapshots. (a) In the MovieLens, we select the rating snapshots 5, 10, 20, 30, 40, 50 respectively. (b) In the Netflix, we select the rating snapshots 5, 10, 30, 50, 60, 70 respectively. Here, X-axis is the time t which is the year of the video released. We could observe that when the videos are released early, their relevance values tend to stabilize. While if those videos are in the early life span, the later they are released, the higher their relevance will be. That is to say, the newer the videos are released, the easier and sharper their burst behaviors are.

($R = 1$). The $R = 1$ according to Eq. (1) expresses that the popularity of the video is added by the expected numbers forming the preferential attachment ($\Delta k_i(t+1)_{(PA)} = \Delta L(t+1)k_i(t)/L(t)$).

Further, we also investigate the video relevance values $R_i(t)$ in different rating snapshots as given in Fig. 6. We could see that in MovieLens as shown in Fig. 6(a), the relevance values of videos which are released before 1995 are fluctuated around the stable value, but the larger the rating snapshot is, the lower relevance value is. After the time point $t = 1995$, the relevance values increase exponentially which reveals the bursts, in addition we can find the relevance values increase faster in the early rating snapshots. While in Netflix as shown in Fig. 6(b), the relevance of videos which are released before $t = 2000$ in the Netflix is fluctuated around 1. But, after the time point $t = 2000$, the relevance increases exponentially which similarly indicates that the bursts occur, and the relevance value also rises faster in early rating snapshots. We could conclude that when the videos are released early, their relevance tends to stabilize. While those videos are in the early life span, the later they are born, the higher their relevance will be, specially, as exponential form. That is to say, the newer the videos are, the easier and sharper their burst behaviors are.

4. Conclusions and discussions

In summary, we studied the online popularity dynamics on MovieLens and Netflix two online video-provided data sets which contain over 100 million records and even span a decade. The two big time stamped data sets provide a chance to deepen understanding about the dynamics of online popularity evolving over time on a global scale. Our findings indicate that the dynamics of online video popularity are characterized by the burst behaviors (i.e. far exceed preferential popularity increase). Typically, the presences of burst behaviors occur in the early life span of videos, and later restrict to the classic preferential popularity increase mechanism. Although this kind proposed methods we have studied here cannot ultimately resolve the debates, which prompt us to question our intuitive understanding of online social processes, by pointing out possible directions for empirical studies. For example, our results might find practical applications in optimized online services including videos, news, music and other events in the online systems, which can guide the future design of online popularity prediction methods, which in turn, can benefit various other services, including information filtering and recommendation, as well as more cost-effective marketing strategies. In a broader context, our work could be relevant to other fields of online social processes [33], such as online behavior pattern, online marketing, word of mouth spreading or other dynamical processes, which may provide insights in the analysis of these collective behaviors, from social influence to biomedical responses.

Acknowledgments

We acknowledge Prof. Yi-Cheng Zhang and Yi-Xiu Kong for fruitful discussion and suggestions. This work is partially supported by the EU FP7 Grant 611272 (Project GROWTHCOM), the Swiss National Science Foundation (Grant Nos. 200020-156188). Z.-M. Ren is supported by China Scholarship Council. Hao Liao is supported by Science and Technology Innovation Commission of Shenzhen (Nos. JCYJ20150625101524056).

References

- [1] A. Vespignani, Predicting the behavior of techno-social systems, *Science* 325 (5939) (2009) 425.
- [2] R.M. Bond, C.J. Fariss, J.J. Jones, A.D. Kramer, C. Marlow, J.E. Settle, J.H. Fowler, A 61-million-person experiment in social influence and political mobilization, *Nature* 489 (7415) (2012) 295–298.
- [3] L. Lü, M. Medo, C.H. Yeung, Y.-C. Zhang, Z.-K. Zhang, T. Zhou, Recommender systems, *Phys. Rep.* 519 (1) (2012) 1–49.

- [4] G. Szabo, B.A. Huberman, Predicting the popularity of online content, *Commun. ACM* 53 (8) (2010) 80–88.
- [5] J.P. Gleeson, D. Cellai, J.-P. Onnela, M.A. Porter, F. Reed-Tsochas, A simple generative model of collective online behavior, *Proc. Natl. Acad. Sci.* 111 (29) (2014) 10411–10415.
- [6] C. Lynch, Big data: How do your data grow? *Nature* 455 (7209) (2008) 28–29.
- [7] M.J. Salganik, P.S. Dodds, D.J. Watts, Experimental study of inequality and unpredictability in an artificial cultural market, *Science* 311 (5762) (2006) 854–856.
- [8] M.-S. Shang, L. Lü, Y.-C. Zhang, T. Zhou, Empirical analysis of web-based user–object bipartite networks, *Europhys. Lett.* 90 (4) (2010) 48006.
- [9] C.X. Zhang, Z.K. Zhang, C. Liu, An evolving model of online bipartite networks, *Physica A* 392 (23) (2013) 6100–6106.
- [10] R. Crane, D. Sornette, Robust dynamic classes revealed by measuring the response function of a social system, *Proc. Natl. Acad. Sci.* 105 (41) (2008) 15649–15653.
- [11] F. Wu, B.A. Huberman, Novelty and collective attention, *Proc. Natl. Acad. Sci.* 104 (45) (2007) 17599–17601.
- [12] A.-L. Barabási, The origin of bursts and heavy tails in human dynamics, *Nature* 435 (7039) (2005) 207–211.
- [13] L. Hou, X. Pan, Q. Guo, J.-G. Liu, Memory effect of the online user preference, *Sci. Rep.* 4 (2014).
- [14] M. Cha, H. Kwak, P. Rodriguez, Y.-Y. Ahn, S. Moon, I tube, you tube, everybody tubes: Analyzing the world’s largest user generated content video system, in: *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement, IMC’07*, ACM, New York, NY, USA, 2007, pp. 1–14.
- [15] Y. Borghol, S. Ardon, N. Carlsson, D. Eager, A. Mahanti, The untold story of the clones: Content-agnostic factors that impact youtube video popularity, in: *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD’12*, ACM, New York, NY, USA, 2012, pp. 1186–1194.
- [16] K. Lerman, R. Ghosh, Information contagion: An empirical study of the spread of news on digg and twitter social networks, *ICWSM 10* (2010) 90–97.
- [17] H.-W. Shen, D. Wang, C. Song, A.-L. Barabási, Modeling and predicting popularity dynamics via reinforced Poisson processes, in: *The Twenty-Eighth AAAI Conference on Artificial Intelligence, 2014, AAAI-14*.
- [18] M. Vasconcelos, J. Almeida, M. Gonçalves, D. Souza, G. Gomes, Popularity dynamics of foursquare micro-reviews, in: *Proceedings of the Second ACM Conference on Online Social Networks, COSN’14*, ACM, New York, NY, USA, 2014, pp. 119–130.
- [19] L. Chen, Y. Zhou, D.M. Chiu, A lifetime model of online video popularity, in: *2014 23rd International Conference on Computer Communication and Networks, ICCCN, IEEE, 2014*, pp. 1–8.
- [20] S. Fortunato, A. Flammini, F. Menczer, Scale-free network growth by ranking, *Phys. Rev. Lett.* 96 (21) (2006) 218701.
- [21] J. Ratkiewicz, S. Fortunato, A. Flammini, F. Menczer, A. Vespignani, Characterizing and modeling the dynamics of online popularity, *Phys. Rev. Lett.* 105 (15) (2010) 158701.
- [22] Y. Borghol, S. Mitra, S. Ardon, N. Carlsson, D. Eager, A. Mahanti, Characterizing and modelling popularity of user-generated videos, *Perform. Eval.* 68 (11) (2011) 1037–1055.
- [23] M. Medo, G. Cimini, S. Gualdi, Temporal effects in the growth of networks, *Phys. Rev. Lett.* 107 (23) (2011) 238701.
- [24] Y.-H. Eom, S. Fortunato, Characterizing and modeling citation dynamics, *PLoS One* 6 (9) (2011) e24926.
- [25] M. Medo, Statistical validation of high-dimensional models of growing networks, *Phys. Rev. E* 89 (3) (2014) 032801.
- [26] <http://movielens.org> and <http://grouplens.org/>.
- [27] J. Bennett, S. Lanning, The netflix prize, in: *Proceedings of KDD Cup and Workshop, Vol. 2007, 2007*, p. 35.
- [28] A.-L. Barabási, R. Albert, Emergence of scaling in random networks, *Science* 286 (5439) (1999) 509–512.
- [29] G. Bianconi, A.-L. Barabási, Competition and multiscaling in evolving networks, *Europhys. Lett.* 54 (4) (2001) 436.
- [30] G. García-Pérez, M. Boguná, M. Serrano, Regulation of burstiness by network-driven activation, *arXiv preprint arXiv:1410.3816*.
- [31] K.-I. Goh, A.-L. Barabási, Burstiness and memory in complex systems, *Europhys. Lett.* 81 (4) (2008) 48002.
- [32] J.-G. Liu, Z.-M. Ren, Q. Guo, D.-B. Chen, Evolution characteristics of the network core in the facebook, *PLoS One* 9 (8) (2014) e104028.
- [33] J. Saramaki, P. Holme, Exploring temporal networks with greedy walks, *Eur. Phys. J. B* 88 (12) (2015) 1–8.